NBER WORKING PAPER SERIES

EFFECT OF THE JAMAICA EARLY CHILDHOOD STIMULATION INTERVENTION ON LABOR MARKET OUTCOMES AT AGE 31

Paul Gertler
James J. Heckman
Rodrigo Pinto
Susan M. Chang
Sally Grantham-McGregor
Christel Vermeersch
Susan Walker
Amika Wright

Working Paper 29292
http://www.nber.org/papers/w29292

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2021

The authors acknowledge financial support from the Strategic Impact Evaluation Fund at the World Bank and from the Eunice Kennedy Shriver National Institute of Child Health and Human Development of the National Institutes of Health under award number R37HD065072. The views expressed herein are those of the authors alone and do not necessarily reflect the views of the National Bureau of Economic Research, the World Bank, or the National Institutes of Health. The authors declare that they have no financial or material interests in the results reported in this paper.

Effect of the Jamaica Early Childhood Stimulation Intervention on Labor Market Outcomes at Age 31
Paul Gertler, James J. Heckman, Rodrigo Pinto, Susan M. Chang, Sally Grantham-McGregor, Christel Vermeersch, Susan Walker, and Amika Wright
NBER Working Paper No. 29292
September 2021
JEL No. C31,I21,J13

ABSTRACT

We report the labor market effects of the Jamaica Early Childhood Stimulation intervention at age 31. The study is a small-sample randomized early childhood education stimulation intervention targeting stunted children living in the poor neighborhoods of Kingston, Jamaica. Implemented in 1987-1989, treatment consisted of a two-year home-based intervention designed to improve nutrition and the quality of mother-child interactions to foster cognitive, language and psycho-social skills. The original sample is 127 stunted children between 9 and 24 months old. Our study is able to track and interview 75%of the original sample 30 years after the intervention, both still living in Jamaica and migrated abroad. We find large and statistically significant effects on income and schooling; the treatment group had 43% higher hourly wages and 37% higher earnings than the control group. This is a substantial increase over the treatment effect estimated for age 22 where we observed a 25% increase in earnings. The Jamaican Study is a rare case of a long-term follow up for an early childhood development (ECD) intervention implemented in a less-developed country. Our results confirm large economic returns to an early childhood intervention that targeted disadvantaged families living in poverty in the poor neighborhoods of Jamaica. The Jamaican intervention is being replicated around the world. Our analysis provides justification for expanding ECD interventions targeting disadvantaged children living in poor countries around the world.

Paul Gertler
Haas School of Business
University of California, Berkeley
Berkeley, CA 94720
and NBER
gertler@haas.berkeley.edu

James J. Heckman
Center for the Economics of
Human Development
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and IZA
and also NBER
jjh@uchicago.edu

Rodrigo Pinto
Department of Economics
8283 Bunche Hall
Los Angeles, CA 90095
rodrig@econ.ucla.edu

Susan M. Chang
Caribbean Institute for Health Research
The University of the West Indies
Kingston, Jamaica
susan.changlopez@uwimona.edu.jm

Sally Grantham-McGregor
University College London
Gower Street
London - WC1E 6BT
sallymcgregor@yahoo.com

Christel Vermeersch
The World Bank
1818 H Street, NW
Washington DC
cvermeersch@worldbank.org

Susan Walker
Caribbean Institute for Health Research
The University of the West Indies
Kingston, Jamaica
Susan.Walker@uwimona.edu.jm

Amika Wright
Caribbean Institute for Health Research
The University of the West Indies
Kingston, Jamaica
Amika.wright@uwimona.edu.jm

Abstract

We report the labor market effects of the Jamaica Early Childhood Stimulation intervention at age 31. The study is a small-sample randomized early childhood education stimulation intervention targeting stunted children living in the poor neighborhoods of Kingston, Jamaica. Implemented in 1987-1989, treatment consisted of a two-year home-based intervention designed to improve nutrition and the quality of mother-child interactions to foster cognitive, language and psycho-social skills. The original sample is 127 stunted children between 9 and 24 months old. Our study is able to track and interview 75% of the original sample 30 years after the intervention, both still living in Jamaica and migrated abroad. We find large and statistically significant effects on income and schooling; the treatment group had 43% higher hourly wages and 37% higher earnings than the control group. This is a substantial increase over the treatment effect estimated for age 22 where we observed a 25% increase in earnings.

KEYWORDS: Early Childhood Development, Jamaican Study, Labor Market Outcomes.
JEL CODES: C31, I21, J13

Contents

1 Introduction **4**

2 The Jamaica Early Childhood Stimulation Intervention Study **5**

3 The New Survey at Age 31 **7**

4 Methods **8**

5 Results **10**

6 Conclusion **11**

Appendices **1**

A Block Permutation Inference **1**

B Inference For Multiple Outcomes **4**

C Attrition **5**

D Baseline Balance and Migration **10**

E Augmented Inverse Propensity Weighting **15**

F Outliers **19**

G Nonstunted Comparison Sample and Catch-up Analysis **24**

1 Introduction

Poor children under 5 years living in low-income countries are vulnerable to developmental risk due to poor nutrition and inadequate stimulation (Engle et al., 2011; Walker et al., 2011a). This paper reports the effects of an early childhood home visiting program, Jamaica Early Childhood Stimulation intervention, on schooling and labor market outcomes at age 31. The Jamaica program was a two-year home-based intervention designed to supplement nutrition and improve the quality of mother-child interactions intended to foster cognitive, language, and psycho-social skills (Grantham-McGregor et al., 1991). The intervention targeted stunted disadvantaged children living in the poor neighborhoods of Kingston. It is a very influential program that has been emulated around the world (Grantham-McGregor and Smith, 2016; Grantham-McGregor et al., 1991; Network, 2014; Tanner et al., 2015; Walker et al., 2011a).[1]

Conducted in 1987-1989, the program was evaluated by a randomized trial that targeted stunted children between 9 and 24 months. A follow-up survey tracked and interviewed 75% of the original sample some 30 years later. The survey interviewed participants still living in Jamaica as well as those who migrated. We estimate treatment effects on schooling and labor market outcomes using permutation-based statistical inference suitable for the small sample size of the study. We implement block-permutation tests specific to the randomization protocol implemented at the onset of the intervention. We also address a range of issues including the possibility of non-random attrition, multiple outcome hypothesis testing, and the presence of outliers.

We find large and statistically significant effects on income and schooling, but not on employment. The treatment group has 43% higher wages and 37% higher earnings than the control group. This is a substantial increase over the treatment effect at age 22 where we observed a 25% increase in earnings (Gertler et al., 2014). Our work aligns with that of Walker et al. (2021), who evaluate the effects of Jamaica Early Childhood Stimulation intervention on psychological measures at the same age we study. They find substantial and sustained benefits of the intervention on cognitive and non-cognitive skills that other studies have shown to be rewarded in the labor market (Heckman et al., 2019; Heckman and Kautz, 2012; Heckman et al., 2013).

Our study contributes to the literature that demonstrates that investing in skill formation at early stages in life has long-lasting economic returns later in life (Campbell et al., 2013, 2014; Carneiro and Ginja, 2014; Conti et al., 2016; Gertler et al., 2014; Heckman et al., 2010b). These interventions have been shown to be cost-effective and yield lifetime gains across several domains including education, earnings, behavior, and health (Elango et al., 2016; García et al., 2021a; García et al., 2017, 2018; García et al., 2019; Heckman, 2006, 2007; Kautz et al., 2014).

The Jamaica Study is unique in having both long-term follow up on the labor market benefits of a solely home-based early childhood intervention and evidence on its effectiveness in a less developed country. The Perry preschool program evaluated through age 54 combines a home

[1] See later studies by Andrew et al. (2019, 2018); Attanasio et al. (2020); Doyle (2020); Gertler et al. (2014); Grantham-McGregor et al. (2020, 1991); Hamadani et al. (2019); Heckman et al. (2021); Rubio-Codina et al. (2019); Smith et al. (2018).

visiting component with a center-based program and has been evaluated through age 55 (see García et al., 2021a and García et al., 2021b). Most other long term evidence is from US-based studies of center-based care.

2 The Jamaica Early Childhood Stimulation Intervention Study

The study enrolled 129 stunted children age 9-24 months identified by a survey of disadvantaged neighborhoods of Kingston, Jamaica.[2] The study used stunting, a condition that can be accurately and easily observed, to identify socially and biologically disadvantaged children. Stunting stems mostly from malnutrition during gestation and the first two years of life, often combined with chronic or repeated infection, and is strongly associated with poor cognitive development (Walker et al., 2007). Stunting is defined as having height more than two standard deviations below median of the National Center for Health Statistics (NCHS) well-nourished reference population standards (Hamill et al., 1979), the most commonly used reference at the time.

The original sample was stratified by age (above and below 16 months) and gender. Within each stratum, children were randomly assigned to one of four groups: (1) psychosocial stimulation (N=32); (2) nutritional supplementation (N=32); (3) both psychosocial stimulation and nutritional supplementation (N=32); (4) a control group that received neither (N=33). All children were given access to free health care regardless of the group to which they were assigned.

Two of the initial 129 children originally assigned to the stimulation arm of the intervention did not complete the intervention. They were dropped from the study before the first followup due to failure to complete the intervention so that the actual sample consisted of 127 children.[3]

Stimulation Intervention

The stimulation intervention is applied to groups 1 and 3. It consisted of weekly one-hour home visits in which a community public health aide engaged mothers to interact with their children. All health aides had some level of secondary education. They had been previously trained in health and nutrition and received a four week training in child development, teaching techniques, and toy making in addition to basic training in nutrition and primary health (Walker et al., 1990). The intervention lasted for two years.

The curriculum was designed to develop child cognitive, language, and socioemotional skills. Activities included mediating the environment through labeling, describing objects, and actions in the environment, responding to the child's vocalizations and actions, playing educational games, and using picture books and songs that facilitated language acquisition. In the intervention before 18 months included Piagetian concepts such as use of a tool and object permanence (Uzgiris and Hunt, 1975). After 18 months concepts such as size, shape, quantity, color, and classification based

[2]See Walker et al. (1990) and Walker et al. (1991) for a more complete description of the intervention.

[3]One mother decided not to participate shortly after enrolment and another moved to another city and could not be followed.

on Palmer (1971) were included. Particular emphasis was placed on the use of praise and giving positive feedback to both the mother and the child.

A major focus of the weekly visits was on improving the quality of the interaction between mother and child. At every visit the use of homemade toys was demonstrated. The toys were left for the mother and child to use until the next visit when they were replaced with different ones. Mothers were encouraged to continue the activities between visits. The intervention was innovative not only for its focus on structured activities aimed at the individual child's level of development to promote cognitive, language, and socioemotional development but also for its emphasis on supporting the mothers to promote their child's development.

Supplementation (Nutritional) Intervention

The nutritional intervention was applied to groups 2 and 3. It consisted of a weekly supply of nutritional supplements that aimed to compensate for nourishment deficiencies that may have caused stunting. The supplements consisted of one kilogram of milk-based formula containing 66% of daily-recommended energy (calories), and 100% of daily-recommended protein and micronutrients (Walker et al., 1992). The child's family also received 0.9 kilograms of cornmeal and skimmed milk powder to prevent the sharing of the nutrition formula among family members. Despite this, sharing was common and uptake of the supplement decreased significantly during the intervention (Walker et al., 1991). The nutrition intervention lasted 2 years and ran concurrently with the stimulation intervention.

Previous Studies

The 127 participants who completed the program were surveyed at baseline and at the end of the second year of the intervention. Subsequent surveys occurred at ages 7, 11, 17, 22, and 31. The previous literature has shown large and persistent causal effects of the stimulation treatment on cognition. At the end of the 2-year intervention, the developmental levels of children who received stimulation (groups 1 and 3) were significantly above those who did not (groups 2 and 4) (Grantham-McGregor et al., 1991). Significant long-term benefits were sustained through age 31 (Walker et al., 2005, 2011b, 2021). Moreover, stimulation treatment had positive and long-lasting impacts on psychosocial skills, and schooling attainment. It reduced participation in violent crimes at age 22 (Walker et al., 2005, 2021).

The nutrition intervention did not share the same strong and lasting effects of the stimulation arms. There are no significant long-term effects of nutrition on any measured outcome after the end of the 2-year trial (Walker et al., 2011a, 2005). This is in contrast with a study in Guatemala which nutritional supplementation did affect both long-term health status and earnings (Hoddinott et al., 2008; Maluccio et al., 2009).

The Guatemala and Jamaican experiments differ in how the nutrition intervention was conducted. The Guatemala Study offered nutrition supplements to pregnant women and from birth for 7 years, prior to the onset of stunting during the first 1,000 days thought to be critical for stunting, most of the Jamaican children were older than 12 months and were already stunted. The late onset of the Jamaican intervention likely explains the lack of long-term nutritional effects. Other reasons are the smaller size of the Jamaican supplement and the fact that it was shared by family members, whereas in Guatemala, supplements were given directly to the child at the center (Hoddinott et al., 2008; Walker et al., 1992, 1990).

Gertler et al. (2014) investigate the effect of the stimulation intervention on labor market outcomes at age 22. They find that the treatment group earned 25% more than the control group, but were no more likely to be employed. They also find that there were no statistically significant or quantitatively important differences in estimated treatment effects between the stimulation and stimulation-nutrition arms on any long-term outcome at age 22. Supplementation had no statistically significant impact on any of the outcomes at age 22. They test and do not reject the hypothesis that the outcomes for the groups that received psychosocial stimulation, groups 1 and 3 are not different and can be pooled. They also test the hypotheses that the groups that did not receive psychological stimulation (groups 2 and 4) can be pooled. Statistical evidence suggests to pool the psychosocial stimulation groups. In light of this evidence, Gertler et al. (2014) combine the two stimulation arms into a single treatment group (N=64) and combine the nutritional supplementation-only group with the pure control group into a single control group (N=65). We do the same in this paper.

The study enrolled an additional sample of 84 nonstunted children living in the same area of the stunted participants. The characteristics of the nonstunted group are described in Gertler et al. (2014). These children are not as disadvantaged as the stunted participants. They have better family backgrounds and socioeconomic outcomes. Nonstunted children were surveyed at age 31. Appendix 7 compares the nonstunted group with the stunted children from both control and the treatment groups. Following Gertler et al. (2014), we examine if treatment enables stunted treatment group members to catch up with nonstunted ones.

3 The New Survey at Age 31

We analyze the most recent survey of the Jamaica Early Childhood Stimulation intervention sample taken when participants were approximately 31 years old. There was an attempt to find all of the 127 initial study participants regardless of location. Researchers contacted relatives to gather information on participants who were not found in Jamaica. The survey follows migrants living in the US, Canada, and UK. Found were 95 (75%) of the original 127 participants at age 31. The attrition rate increased from 17% for the 22 year old follow-up to 25% for this survey. Of the 32 original participants lost to follow up, 11 died, 6 refused to be interviewed, 12 could not be found, and 3 were incarcerated or in hospital (see Table A.1 of the Appendix for more details.).

Attrition is well-balanced across treatment groups for baseline variables. The statistical analysis of attrition at age 31 is presented in Tables A.1–A.3 of our online appendix. The attrition rate is not statistically different across any of the four randomization arms (Table A.3). The means of the baseline variables are not significantly different between the observed and missing participants (Table A.3). Moreover, the treatment status is not a statistically significant predictor of the overall probability of attrition (Table A.3). Table A.4 of the Appendix compares the baseline variables of the missing participants at age 22 with those who attrite at age 30. The baseline characteristics of the additional participants that are missing at age 30 are not statistically different from those who were missing at age 22.

Baseline variables remain balanced across treatment and control groups for the age 31 survey. Table A.5 of our appendix shows that the means of baseline variables are not statistically different for treated and control groups after controlling for the randomization protocol.

The distribution of migrants is balanced across treated and control groups. There are 8 migrants in each group. Table A.5 of our Appendix shows that migration is not statistically significantly different between treatments and controls for the full sample. Baseline variables are balanced between migrants and non-migrants using the full data set. Table A.6 of our Appendix shows that none of the mean differences of baseline variables between migrants and non-migrants is statistically significant. Table A.7 shows a gender-specific migration pattern. Treated females are more likely to migrate than control females. The opposite occurs for males. These results motivate us to present results for the overall sample and also for three sub-samples: males, females and non-migrants.

4 Methods

We examine the impact of the stimulation treatment on labor market outcomes – wages, earned income, and employment – and on schooling, a mechanism for improved economic outcomes. Recall that we follow the previous literature on the Jamaica Early Childhood Stimulation intervention that pools the stimulation-only arm with the stimulation/nutritional supplement arm into a single stimulation treatment group, and pools the control arm with the nutritional supplement arm into a single control group. We evaluate the causal effect of the stimulation treatment conditioned on the baseline variables used for stratification in the randomization protocol (age and gender) and control for the imbalance of pre-program variables. We estimate treatment effects for the whole sample and separately by gender. Section 2 of the online appendix describes the method in detail.

The small sample size raises the issue of the relevance of classical statistical inferential methods based on asymptotic theory. Instead, we primarily employ a non-parametric block-permutation test that does not rely on the asymptotic distribution of the test statistic and is valid in small samples (see, e.g., Heckman et al., 2010a). The test nonparametrically partitions the sample within the blocks used for stratification of treatment assignment. Permutation testing is then performed within each partition block. Details of the procedure are described in Section 2 of the Appendix.

We address the problem arising from cherry picking individual hypotheses ("p hacking") by using multiple outcome hypothesis tests that jointly test the statistical significance of outcomes that share similar interpretation. We implement a stepdown procedure that controls for the family-wise error rate, namely, the probability of rejecting at least one true null hypothesis among a group (Romano and Wolf, 2005).

We also test the significance of treatment effects across multiple outcomes using two methods. The first method is based on a nonparametric index function that aggregates multiple outcomes into a single statistic. We use the rank-average of each participant across multiple outcomes and test the no-treatment hypothesis using differences-in-the-mean of participants' rank-average using our permutation-based inference procedure.[4] We also compute the p-value for the non-bipartite test of Rosenbaum (2005). This is a nonparametric and distribution-free test across multiple outcomes. The test matches participants according to the minimal Mahalanobis distance between outcome ranks. Under the null hypothesis of no treatment effects, we expect a random match between treated and control participants. If treatment affects outcomes, participants are more likely to be matched within their treatment group. The non-bipartite p-value evaluates the likelihood of the matching generated by the observed outcomes.

We investigate the potential bias generated by non-random attrition in Section 5 of our online appendix. We show that the distribution of variables across attriters is surprisingly balanced across randomization arms and that the attrition rate is not statistically different across randomization arms. We also investigate whether the distribution of the treatment indicator and baseline variables are statistically different by attrition status. We do not reject the null hypothesis that the means of the baseline variables are the same for attriters and those who are observed.

Our analyses suggest that non-random attrition is not a major concern. Nevertheless, we correct for potential attrition bias in a robust fashion by using the Augmented Inverse Propensity Weighting (AIPW) model (See Tables A.8–A.9 of the online appendix). The AIPW model is based on an IPW approach that recovers the original distribution of treatment status with no attrition by reweighting the data using baseline variables. The AIPW estimator improves on the standard IPW by exploiting the predictive information on baseline variables to forecast outcomes (Glynn and Quinn, 2010; Huber, 2012; Robins et al., 1994). See section 2 of the online appendix for a detailed description of this approach and results. The AIPW estimates are almost identical to those presented here, providing additional assurance that our estimates do not suffer from attrition bias.

A total of 16 out of 95 participants are migrants who live in the US, UK and Canada. The labor markets of foreign countries differ greatly from the Jamaican market. Wages and earnings from these countries can be substantially larger than those in Jamaica and may therefore introduce outliers that could heavily influence treatment effect estimates, especially with our small sample size. We formally test for the presence of outliers using Cook's Distance and Influence/Leverage Indexes (Rousseeuw and Leroy, 1996). All the tests point to a single outlier in the earnings data, whose value is 35 times larger than the sample average. We exclude this outlier in our analysis of treatment effects, but not from the rank-sum analyses. We found no outliers in wage data.

Finally, we address the fact that wage and earnings data are highly skewed. This matter is of particular concern for small sample permutation tests as a few extreme data points might determine

[4] For details, see Section 3 of our online appendix.

the overall distribution of the test statistics. The literature on linear regression suggests that analysts should limit the skewness of outcomes to ± 2 (Gravetter and Wallnau, 2014; Trochim and Donnelly, 2006). Unfortunately, the skewness of wage and earnings are 2.17 and 2.23 respectively.

We address the problem in two ways. First, we use a log-transformation of the data, which reduces the skewness of wages from 2.17 to 0.32 and of earnings from 2.23 to -0.09. Treatment effects are then interpreted as an estimate of the elasticity of wages or earnings with respect to treatment assignment. Our second solution is to use the generalized Rank-sum statistic to do inference on causal effects (Boos and Stefanski, 2012; Conover and Salsburg, 1988). Rank-sum tests employ a nonparametric statistic based on the cumulative distribution of the data instead of the actual outcome values. The test is robust to the presence of outliers and data skewness. For earnings we also include the outlier in the rank-sum statistics.

5 Results

Figure 1 compares the cumulative distribution functions for the log of wages and earnings for treatment and control groups.[5] The cumulative distributions of the treated stochastically dominate the control distributions except at extremely high values of the outcomes. Kolmogorov-Smirnov tests confirm that the cumulative distributions for treatments and controls are significantly different from one another for both outcomes. These results suggest that both wages and earnings are bigger in the treatment group than in the control for the vast majority of the range of values and that the differences in means are not driven by extreme values.

Table 1 reports the treatment effect estimates for wages and earning for the full sample, non-migrants, males and females. The effects for the combined sample of males and females are reported in the top panel. The estimated effect sizes for wages and earnings are 43% and 37% respectively. The rank-mean statistic consists of an index function that employs the average participant rank cross the outcomes (See Section 2 of the Appendix for more details). The estimated rank-mean effect size for the full sample is 45%. The estimates are statistically significant regardless of the measure used. We find larger effect sizes when we restrict the sample to non-migrants as displayed in the second panel of Table 1. The result suggests that the wages and earnings results are not overly influenced by the migrant data. The last two panels display the treatment effects by gender. Causal effects are much larger for females than males consistent with their elevated levels of schooling attendance.

Schooling is the most plausible mediator for the wages and earnings results. Average treatment effects on schooling are reported in Table 2. The table presents estimates for the pooled sample and estimates by gender. It shows that the average increase on schooling for treated participants is three-fourths of a year. The treatment increases college enrollment by 14 percentage points and increases the likelihood of acquiring a higher education diploma by 26 percentage points. The

[5] Both wages and earnings are measured in US dollars. They were converted to US dollars from local currency using the exchange rate at the time of the survey.

average increase in the rank-mean statistic is 45%. Similar to the earnings results, the treatment effect on schooling is substantially higher for females than for males. This is in line with the causal effects on wages and earnings which are stronger among females.

Employment is another plausible mediator. However, we find no effect of treatment on labor force participation for the pooled sample or separately for males and females (Table 3). We find only a weak effect on employment that requires highly skilled labor. The evidence suggests that skill enhancement (via schooling or otherwise) is responsible for the estimated wage and earnings effects.

Tables A.13–A.15 in the Appendix report the degree to which the intervention enabled the stunted treatment group to catch up to the nonstunted comparison group. Overall, we find that treated participants catch up with nonstunted participants on schooling outcomes, but there is a gender difference on effects for income. Treated females catch up with non-stunted females on income, but treated males do not.

6 Conclusion

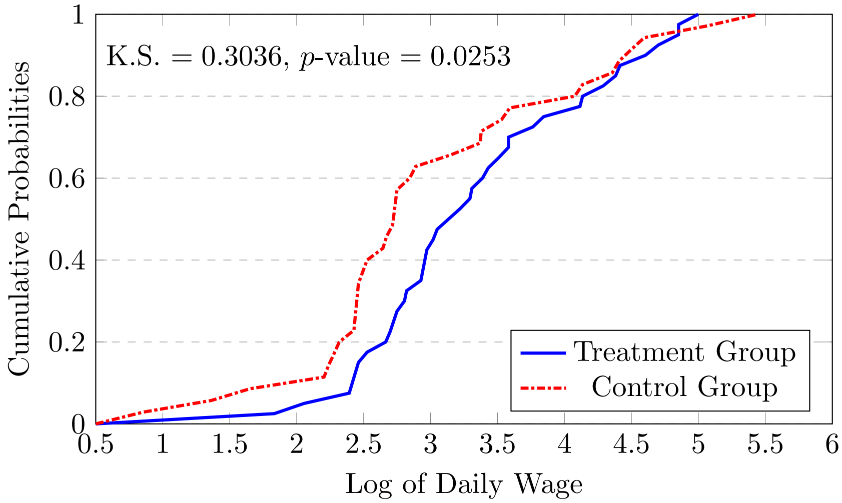
This paper evaluates the long-term economic impacts of the Jamaica Early Childhood Stimulation intervention program, an early childhood intervention for socially and biologically disadvantaged children living in a poor country. The study consists of a randomized control trial that enrolled 127 stunted children between 9 and 24 months old living in Kingston, Jamaica, during the years 1987-89. Treated participants received a two-year home-based intervention designed to improve the quality of mother-child interactions so as to develop cognitive, language, and psycho-social skills. We investigate labor market and educational outcomes surveyed at age 31.

The study tracked and interviewed 75% of the original sample some 30 years after the intervention living both in Jamaica and abroad. We find large and statistically significant effects on income and schooling, but not on employment at age 31. Specifically, the treatment group had 43% higher wages and 37% higher earnings than the control group. Moreover, the treatment effect is larger for females than males. This is a substantial increase over treatment effects at age 22 where we observed a 25% increase in earnings.

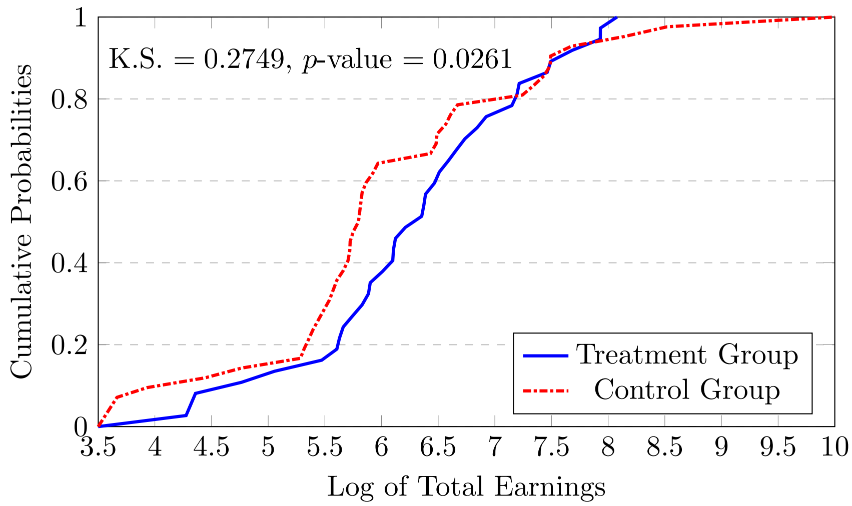
The Jamaican Study is a rare case of a long-term follow up for an early childhood development (ECD) intervention implemented in a less-developed country. Our results confirm large economic returns to an early childhood intervention that targeted disadvantaged families living in poverty in the poor neighborhoods of Jamaica. The Jamaican intervention is being replicated around the world (Grantham-McGregor and Smith, 2016). Our analysis provides justification for expanding ECD interventions targeting disadvantaged children living in poor countries around the world (Richter et al., 2018).

Figure 1: Cumulative Density Functions for Wage and Earnings by Treatment Status

Panel A: CDF of the Log of Daily Wage by Treatment Status



Panel B: CDF of Log of Total Earnings by Treatment Status



Panel A of this figure presents the estimated cumulative density function (CDF) of the Log of Daily Wages by treatment status. Panel B displays the CDF of the Log of Earnings by treatment status. Both wage and earnings variables are measures in US dollars. Treatment group combines the stimulation arms of the intervention while the control group comprise the participants that did not experience the stimulation treatment. We perform the two-sample Kolmogorov-Smirnov (KS) one-tailed test for equality of distributions against the hypothesis that the distribution of the treated group dominates the distribution of the control group. The KS statistic for Log of Daily Wage (Panel A) is given by 0.3036 and its associated p-value is 0.0253. The KS statistic for Log of Last Earnings (Panel B) is given by 0.2749 and its associated p-value is 0.0261.

Table 1: Treatment Effects on Wages and Earnings

Variables	Sample		Control	Percentage	Effect	Asymptotic (one-sided)		Permutation (one-sided)	
	# C	# T	Mean	Change	Size	t-stat	Single p-val	Single p-val	Stepdown
Full Sample									
Log of Daily Wage	35	40	2.98	0.41	0.43	1.82	**0.04**	**0.04**	**0.07**
Log of Monthly Total Earnings	44	46	5.92	0.39	0.37	1.68	**0.05**	**0.05**	**0.05**
Rank Mean	44	46	0.41	0.13	0.45	1.98	**0.03**	**0.03**	–
Non-bibartite p-value	0.549								
Non-migrant Sample									
Log of Daily Wage	29	33	2.66	0.50	0.68	2.37	**0.01**	**0.02**	**0.03**
Log of Monthly Total Earnings	37	38	5.66	0.39	0.42	1.64	**0.05**	**0.06**	**0.06**
Rank Mean	37	38	0.41	0.14	0.50	1.90	**0.03**	**0.04**	–
Non-bibartite p-value	0.270								
Female Sample									
Log of Daily Wage	13	18	2.67	0.71	1.12	2.19	**0.02**	**0.02**	**0.04**
Log of Monthly Total Earnings	20	22	5.53	0.71	0.70	2.05	**0.02**	**0.03**	**0.03**
Rank Mean	20	22	0.38	0.18	0.71	1.92	**0.03**	**0.04**	–
Non-bibartite p-value	**0.090**								
Male Sample									
Log of Daily Wage	22	22	3.19	0.17	0.16	0.56	0.29	0.29	0.37
Log of Monthly Total Earnings	24	24	6.26	0.09	0.08	0.29	0.39	0.39	0.39
Rank Mean	24	24	0.46	0.04	0.13	0.44	0.33	0.33	–
Non-bibartite p-value	**0.032**								

The columns of this table presents the following information. **Col.1**: variable of interest; **Col.2**: control sample size; **Col.3**: treatment sample size; **Col.4**: control mean; **Col.5**: estimated treatment effect of the log of the variables, which can be interpreted as the percentage increase of the outcome mean between treated and control group; **Col.6**: effect size according to the pooled standard deviation (Hedge's g); **Col.7**: t-statistic associated with the treatment effect; **Col.8**: asymptotic p-value for the single hypothesis testing of no treatment effect; **Col.9**: the single hypothesis one-sided mid-p-value based on 15.000 permutations draws; The inference is based on a t-statistic that uses pre-pivoted treatment effect estimate and on a block permutation scheme; **Col.10**: the multiple hypothesis testing (stepdown) for p-values in column 9. Estimates are based on a block permutation inference conditional on main variables at the onset of the intervention. The row termed Rank Mean presents the inference across the block of outcomes. It employs the summary index generated by the average rank of each participant across the outcome variables. Estimates are based on a block permutation inference conditional on main variables at the onset of the intervention. The last row presents the Non-bipartite p-value of Rosenbaum (2005). We suppress an outlier in the total earnings data whose value is 35 times higher than the average earnings.

Table 2: Treatment Effects on Schooling

Variables	Sample		Control	Treatment	Effect	**Asymptotic** (one-sided)		**Permutation** (one-sided)	
	#C	#T	Mean	Effects	Size	t-stat	Single p-val	Single p-val	Stepdown
Full Sample									
Years of education	47	48	10.69	0.77	0.46	2.02	**0.02**	**0.02**	**0.05**
Any college education ?	47	48	0.11	0.14	0.43	1.80	**0.04**	**0.06**	**0.10**
Higher education diploma ?	47	48	0.09	0.26	0.42	1.68	**0.05**	**0.07**	**0.07**
Rank Mean	47	48	0.47	0.07	0.45	1.88	**0.03**	**0.03**	-
Non-bibartite p-value	0.529								
Female Sample									
Years of education	21	23	10.55	1.33	1.18	2.41	**0.01**	**0.01**	**0.01**
Any college education ?	21	23	−0.00	0.31	2.18	3.08	**0.00**	**0.00**	**0.00**
Higher education diploma ?	21	23	−0.03	0.49	1.86	2.15	**0.02**	**0.01**	**0.01**
Rank Mean	21	23	0.44	0.12	1.32	2.53	**0.01**	**0.01**	-
Non-bibartite p-value	**0.081**								
Male Sample									
Years of education	26	25	10.82	0.25	0.13	0.48	0.32	0.30	0.46
Any college education ?	26	25	0.20	−0.01	−0.02	−0.07	0.53	0.63	0.63
Higher education diploma ?	26	25	0.20	0.03	0.04	0.17	0.43	0.51	0.63
Rank Mean	26	25	0.49	0.02	0.10	0.36	0.36	0.36	-
Non-bibartite p-value	0.133								

The columns of this table presents the following information. **Col.1**: variable of interest; **Col.2**: Control sample size; **Col.3**: Treated sample size; **Col.4**: control mean; **Col.5**: estimated treatment effect; **Col.6**: effect size according to the pooled standard deviation (Hedge's g); **Col.7**: t-statistic associated with the treatment effect; **Col.8**: asymptotic one-sided p-value for the single hypothesis testing of no treatment effect; **Col.9**: the single hypothesis one-sided mid-p-value based on 15.000 permutations draws; The inference is based on a t-statistic that uses pre-pivoted treatment effect estimate and on a block permutation scheme; **Col.10**: the multiple hypothesis testing (stepdown) for p-values in column 9. The row termed Rank Mean presents the inference across the block of outcomes. It employs the summary index generated by the average rank of each participant across the outcome variables. Estimates are based on a block permutation inference conditional on main variables at the onset of the intervention. The last row presents the Non-bipartite p-value of Rosenbaum (2005).

Table 3: Treatment Effects on Employment

Variables	Sample		Control	Treatment	Effect	Asymptotic (one-sided)		Permutation (one-sided)	
	#C	#T	Mean	Effects	Size	t-stat	Single p-val	Single p-val	Stepdown
Full Sample									
Working on a high skill job	47	48	0.18	0.14	0.35	1.45	**0.075**	**0.074**	0.176
Working for someone	47	48	0.66	−0.00	−0.01	−0.03	0.513	0.511	0.658
Working OR Self-employed	47	48	0.90	−0.12	−0.43	−1.57	0.940	0.937	0.937
Rank Mean	47	48	0.50	0.00	0.02	0.06	0.476	0.474	-
Non-bibartite p-value	0.644								
Female Sample									
Working on a high skill job	21	23	0.10	0.16	0.53	1.30	0.101	**0.097**	0.235
Working for someone	21	23	0.56	−0.03	−0.06	−0.17	0.568	0.535	0.698
Working OR Self-employed	21	23	0.88	−0.16	−0.54	−1.35	0.908	0.901	0.901
Rank Mean	21	23	0.50	−0.01	−0.04	−0.10	0.538	0.512	-
Non-bibartite p-value	0.629								
Male Sample									
Working on a high skill job	26	25	0.25	0.13	0.29	0.90	0.186	0.182	0.373
Working for someone	26	25	0.75	0.03	0.06	0.19	0.424	0.429	0.560
Working OR Self-employed	26	25	0.93	−0.09	−0.33	−0.88	0.809	0.771	0.771
Rank Mean	26	25	0.49	0.01	0.08	0.22	0.412	0.417	-
Non-bibartite p-value	0.692								

The columns of this table presents the following information. **Col.1**: variable of interest; **Col.2**: Control sample size; **Col.3**: Treated sample size; **Col.4**: control mean; **Col.5**: estimated treatment effect; **Col.6**: effect size according to the pooled standard deviation (Hedge's g); **Col.7**: t-statistic associated with the treatment effect; **Col.8**: asymptotic one-sided p-value for the single hypothesis testing of no treatment effect; **Col.9**: the single hypothesis one-sided mid-p-value based on 15.000 permutations draws; The inference is based on a t-statistic that uses pre-pivoted treatment effect estimate and on a block permutation scheme; **Col.10**: the multiple hypothesis testing (stepdown) for p-values in column 9. The row termed Rank Mean presents the inference across the block of outcomes. It employs the summary index generated by the average rank of each participant across the outcome variables. Estimates are based on a block permutation inference conditional on main variables at the onset of the intervention. The last row presents the Non-bipartite p-value of Rosenbaum (2005).

Appendices

Appendix A Block Permutation Inference

We estimate treatment effects using linear regression controlling for the variables used to stratify treatment assignment. We use permutation based methods to test the null hypothesis that the treatment effects are zero. Permutation-based inference is often termed data-dependent because the computed p-values are conditional on the observed data. These tests are also called *distribution-free* since they do not rely on parametric assumptions about the distribution from which the data have been sampled.

In practice, permutation tests compare a test statistic computed on the original (not permuted) data with a distribution of test statistics constructed from all possible samples of those data. Under the assumption that the null hypothesis is true, the treatment becomes exchangeable and the distribution of the test statistic can be obtained by permuting the treatment indicator. The measure of evidence against the Randomization Hypothesis, the p-value, is computed as the fraction of resampled data which yields a test statistic greater than that yielded by the original data. We refer to Campbell et al. (2014); Heckman et al. (2010a, 2013) for additional information on permutation tests.

Permutations are made within blocks defined by the randomization protocol and potentially baseline variables that are not balanced. A large number of permutation blocks reduces the number of participants that share the same values of baseline variables and may render some permutation blocks invalid if contain only treatments or only controls. Effectively, we lose those observations as the treatment status does not vary within this block.

To avoid this problem, we apply a parsimonious selection of conditioning covariates. Each the cell of analysis has both treated and control participants. Permutation blocks are defined based on four variables: (1) mother's education, (2) supplementation treatment assignment, (3) child's age at study enrollment, and (4) gender. Child's age and gender are based on the randomization protocol and Mother's education is not balanced at baseline. Since we only estimate the treatment effect of stimulation using the combined stimulation and stimulation plus supplementation arms, we include supplementation assignment so that we are comparing the effect of stimulation with those who did not conditional on their supplementation status.

The procedure we use to define the permutation blocks is:

1. First, partition participants according to their maternal education.
2. Second, partition the participants whose mother had low education achievement into those who had supplementation or not.
3. Third, partition each of the last two groups according to whether the child is older than 16 months at enrollment and gender.

This procedure generates a partition of the sample into eight blocks. Each of the blocks contains both treated and control participants.

The following table displays the results of the partitioning and demonstrates that each block includes both treatments and controls.

Observation Number	Treatment Status	Mother Education	Supplementation Intervention	Male Indicator	Child Age (> 16 mo.)	Permutation Blocks
172	0	0	0	0	0	1
40	1	0	0	0	0	1
34	1	0	0	0	0	1
76	0	0	0	0	0	1
13	0	0	0	0	0	1
151	1	0	0	0	1	2
112	1	0	0	0	1	2
106	1	0	0	0	1	2
145	0	0	0	0	1	2
39	1	0	0	0	1	2
74	1	0	0	0	1	2
162	1	0	0	0	1	2
113	0	0	0	0	1	2
150	1	0	0	0	1	2
59	0	0	0	0	1	2
90	1	0	0	0	1	2
157	0	0	0	0	1	2
12	0	0	0	0	1	2
33	1	0	1	0	0	3
123	1	0	1	0	0	3
57	1	0	1	0	0	3
37	0	0	1	0	0	3
140	1	0	1	0	0	3
14	0	0	1	0	0	3

Observation Number	Treatment Status	Mother Education	Supplementation Intervention	Male Indicator	Child Age (> 16 mo.)	Permutation Blocks
46	1	0	1	0	1	4
136	0	0	1	0	1	4
138	0	0	1	0	1	4
1	0	0	1	0	1	4
118	0	0	1	0	1	4
153	0	0	1	0	1	4
114	1	0	1	0	1	4
89	0	0	1	0	1	4
116	1	0	1	0	1	4
38	1	0	1	0	1	4
159	1	0	1	0	1	4
75	0	0	1	0	1	4
160	1	0	1	0	1	4
70	1	0	1	0	1	4
5	0	0	1	0	1	4
139	1	0	0	1	0	5
98	1	0	0	1	0	5
92	0	0	0	1	0	5
83	0	0	0	1	0	5
86	1	0	0	1	0	5
10	1	0	0	1	0	5
181	0	0	0	1	0	5
104	0	0	0	1	0	5
177	1	0	0	1	1	5
77	0	0	0	1	1	5
154	0	0	0	1	1	5
29	0	0	0	1	1	5
134	1	0	0	1	1	5
133	0	0	0	1	1	5
45	1	0	0	1	1	5
22	1	0	0	1	1	5
73	1	0	0	1	1	5
3	0	0	0	1	1	5
36	1	0	1	1	0	6
25	0	0	1	1	0	6
109	0	0	1	1	0	6
178	1	0	1	1	0	6
47	0	0	1	1	0	6
84	1	0	1	1	0	6
142	1	0	1	1	1	7
161	1	0	1	1	1	7
87	0	0	1	1	1	7
99	1	0	1	1	1	7
88	0	0	1	1	1	7
69	1	0	1	1	1	7
15	1	0	1	1	1	7
91	1	0	1	1	1	7
152	0	0	1	1	1	7
124	1	0	1	1	1	7
30	1	0	1	1	1	7
60	1	0	1	1	1	7
149	0	0	1	1	1	7
43	0	0	1	1	1	7
42	1	0	1	1	1	7
103	1	0	1	1	1	7
8	0	0	1	1	1	7

Observation Number	Treatment Status	Mother Education	Supplementation Intervention	Male Indicator	Child Age (> 16 mo.)	Permutation Blocks
11	1	1	0	0	0	8
129	0	1	0	0	1	8
18	0	1	0	0	1	8
44	1	1	1	0	0	8
100	0	1	1	0	1	8
49	0	1	0	1	0	8
111	0	1	0	1	1	8
27	0	1	0	1	1	8
167	1	1	0	1	1	8
53	0	1	0	1	1	8
94	1	1	0	1	1	8
2	0	1	1	1	0	8
135	0	1	1	1	1	8
101	0	1	1	1	1	8
163	0	1	1	1	1	8

Appendix B Inference For Multiple Outcomes

We take two approaches to adjusting p-values for multiple outcomes. First, we use the Romano-Wolfe stepdown procedure described in the main text. Second, we use the rank-sum average, a non-parametric summary statistic that aggregates multiple outcome measures. We first transform the outcomes into the rank of each participant for each outcome. We then compute the mean of the rank of each participant across outcomes. We then use the difference in means of participant rank-average as a test statistic.

Formally, let $\mathcal{I}$ be the set indexing participants of the Jamaican intervention. Let $D = (D_i; i \in \mathcal{I})$ be the vector of treatment assignments, such that D_i takes value 1 if participant i is assigned to treatment and 0 otherwise. Let $\mathcal{K} = \{1, \ldots, K\}$ be an index set for a selection of outcomes sought to be tested, such that $Y_k = (Y_{i,k}; i \in \mathcal{I})$ denotes the vector of k-th outcome associated with index $k \in \mathcal{K}$. Let Y_k be the dimension of outcome vector Y_k. In this notation, we can compute the rank of the participants within outcome k by:

$$\forall i \in \mathcal{I},\ R_{i,k} = \frac{\sum_{j \in \mathcal{I}} \mathbf{1}[Y_{i,k} \geq Y_{j,k}]}{|Y_k|}.$$

when "$\perp\!\!\!\perp$" denotes cardinality of the set of Y_k values.

Let the average rank of participant $i \in \mathcal{I}$ across outcomes in $\mathcal{K}$ be:

$$\forall i \in \mathcal{I},\ R_i = \frac{\sum_{k \in \mathcal{K}} R_{i,k}}{|\mathcal{K}|}.$$

The vector of the rank average across outcomes in $\mathcal{K}$ for all participants in $\mathcal{I}$, that is, $R = (R_i; i \in \mathcal{I})$, can be used as a combined measure across outcomes. The associated test statistic comparing treatment and control is the standard difference in means across treatment groups, namely:

$$\Delta R = \frac{\sum_{i \in \mathcal{I}} D_i R_i}{\sum_{i \in \mathcal{I}} D_i} - \frac{\sum_{i \in \mathcal{I}} (1 - D_i) R_i}{\sum_{i \in \mathcal{I}} (1 - D_i)}.$$

We use permutation methods to obtain the sampling distribution.

The key difference between the stepdown procedure and the rank-mean test is that the rank-mean employs a summary statistic while the stepdown uses an algorithm. The rank-mean test does not control for FWER while the stepdown does. Average rank statistics cannot be included in the stepdown procedure because doing so violates the subset pivotality condition required to implement the stepdown procedure (Romano and Wolf, 2005).

Appendix C Attrition

Our analysis sample consists of 95 of the original 127 participants. Thirty-two (25%) participants were not interviewed at age 31. Table C.1 describes the attrition patterns and shows that they are surprisingly balanced across multiple surveys rounds and arms of the study. Table C.2 confirms that we cannot reject that the hypothesis that the attrition rate is the same across arms. Attrition is also balanced in terms of baseline variables across study arms. Table C.3 shows that baseline variables are also balanced across observed participants and the attrition group at age 30.

Table C.1: Attrition Profile Across Surveys of the Jamaican Study

	Treatment Arms				
	Control	**Supplement**	**Stimulation**	**Both Treats.**	sum
Onset	33	32	32	32	129
Did not Complete			2		127
7 y.o. follow-up	32	31	29	30	122
11 y.o. follow-up	31	30	27	28	116
17 y.o. follow-up	27	28	21	27	103
22 y.o. follow-up	26	26	24	29	105
30 y.o. follow-up	23	24	22	26	95
Died	4	2	2	1	9
Refused	0	1	1	1	3
Previous Attrition	3	3	3	1	10
30 y.o. Attrition	3	2	2	3	10

This table describes the attrition profile of the four randomization arms of the Jamaican intervention across five surveys at ages 7,11,17,22, and 31 years old.

Table C.2: Contingency Table Attrition at Age 31

	Treatment Arms				
	Control	**Supplement**	**Stimulation**	**Both Treats.**	Total
Attrition	10	8	8	6	32
Frequency	30.30	25.00	26.67	18.75	25.20
Observed	23	24	22	26	95
Frequency	69.70	75.00	73.33	81.25	74.80
Total	33	32	32	30	127

	Contingency Table Test	
	Pearson χ^2 Statistic	1.1972
	Degree of Freedom	3
	P-value	0.754

Table C.3: - Inference on Baseline Variables by Attrition at Age 31

	Reverse	**Sample**		**Not Observed**		**Observed**		**Difference**	**Asymptotic** (two-sided)	
	Variable	**#N**	**#O**	**Mean**	**Std.Dev.**	**Mean**	**Std.Dev.**	**Means**	***t*-stat**	**Single *p*-val**
Child characteristics at Enrollment										
Treatment Status	No	32	95	0.44	0.50	0.51	0.50	0.07	0.66	0.511
sex	No	32	95	1.34	0.48	1.46	0.50	0.12	1.18	0.242
enrollment weight-for-height z-score	No	32	95	−1.09	0.66	−1.04	0.72	0.06	0.39	0.700
enrollment height for age z-score	No	32	95	−3.05	0.66	−2.92	0.60	0.13	1.02	0.309
Griffith quotient at age 2	No	32	95	97.97	8.47	98.05	9.07	0.08	0.05	0.963
enrollment housing	No	32	95	7.44	1.34	7.37	1.64	−0.07	−0.21	0.830
child age on enrollment	No	32	95	17.55	3.56	19.07	4.20	1.52	1.84	**0.068**
Parental Characteristics on Enrollment										
enrollment HOME	No	32	95	16.31	4.37	16.43	4.42	0.12	0.13	0.895
enrollment maternal ppvt	No	32	95	85.00	19.97	85.44	19.75	0.44	0.11	0.913

The columns of this table presents the following information. **Col.1**: variable of interest; **Col.2**: indicates if the variable is reverse, that is multiplied by -1 in order to report desired positive treatment effects ; **Col.3**: Not Observed sample; **Col.4**: Observed sample; **Col.5**: control mean; **Col.6**: standard deviation for the control participants; **Col.7**: treatment mean; **Col.8**: standard deviation for the treated participants; **Col.9**: estimated difference-in-means; **Col.10**: t-statistic associated with the treatment effect. **Col.11**: asymptotic two-sided *p*-value for the single hypothesis of no difference between the observed and attrition groups.

Table C.4 compares the baseline variables of the missing participants at age 22 with those who attrite at age 30. The data shows that the mean baseline characteristics of the additional participants that are missing at age 30 are not statistically different than those of the missing participants at pervious surveys.

Table C.4: - Baseline Characteristics of Missing Participants at age 22 versus the Additional Attriters at Age 31

	Reverse	**Sample**		**Missing at 22**		**Additional Attriters at 30**		**Treat.**	**Effect**	**Asymptotic** (two-sided)	
	Variable	**#M**	**#A***	**Mean**	**Std.Dev.**	**Mean**	**Std.Dev.**	**Effects**	**Size**	***t*-stat**	**Single *p*-val**
Child characteristics on Enrolment											
Treatment Status	No	22	11	0.41	0.50	0.55	0.52	0.14	0.27	0.72	0.474
Gender (1 for Males, 2 for Females)	No	22	11	1.32	0.48	1.36	0.50	0.05	0.10	0.25	0.802
enrolment weight-for-height z-score	No	22	11	−1.18	0.58	−0.90	0.77	0.28	0.48	1.18	0.248
enrolment height for age z-score	No	22	11	−3.06	0.52	−2.95	0.92	0.11	0.21	0.44	0.665
Griffiths quotient at age 2	No	22	11	97.18	7.73	98.55	10.36	1.36	0.18	0.43	0.673
enrolment housing	No	22	11	7.50	1.22	7.45	1.63	−0.05	−0.04	−0.09	0.929
child age on enrolment	No	22	11	17.72	3.48	17.85	4.31	0.13	0.04	0.09	0.928
Parental Characteristics on Enrolment											
enrolment HOME	No	22	11	15.50	4.24	18.18	4.09	2.68	0.63	1.73	**0.093**
enrolment maternal ppvt	No	22	11	82.86	17.79	88.36	23.66	5.50	0.31	0.75	0.459

The columns of this table presents the following information. **Col.1**: variable of interest; **Col.2**: indicates if the variable is reverse, that is multiplied by -1 in order to report desired positive treatment effects ; **Col.3**: Sample of missing participants at age 22; **Col.4**: Sample of additional attrition at age 30; There are 11 participants that were present at age 22 survey but are missing at age 30. One participant that were missing at age 22 became available at age 30. **Col.5**: control mean; **Col.6**: standard deviation for the control participants; **Col.7**: treatment mean; **Col.8**: standard deviation for the treated participants; **Col.9**: estimated treatment effect; **Col.10**: Hedges g effect size according to Rosenthal and Rosnow (1991) and Becker (2000). **Col.11**: t-statistic associated with the treatment effect. **Col.12**: asymptotic one-sided p-value for the single hypothesis testing of no difference between the treatment and control means. Estimates in this table are not conditioned on baseline variables.

Appendix D Baseline Balance and Migration

The first and second panels in Table D.1 presents the difference of baseline variables between treated and control groups at age 31 conditioned on permutation blocks. None of the baseline variables are statistically significant. The last panel of Table D.1 shows that migration decision at age 31 is not statistically different between treated and control groups. The age that participants were surveyed is not statistically different between treated and control groups either.

We also present the unconditional inference of baseline variables at age 31 to clarify the role of conditioning. Table D.1 presents the unconditional inference whether baseline variables are statistically different from treatment and control groups. Overall, the unconditional analysis of baseline variables shows a very balanced sample between treated and control participants.

We only observe significant differences in 2 out of 15 variables: Mother Education (the indicator of secondary exams completion) and the Z-score (weight for height) of development. These imbalances were already present in the full baseline sample of 127, which suggests that they were the result of sampling variation in the original randomization rather than differential sample attrition. These imbalances are more likely to reduce the treatment effects as children in the control group have mothers with slightly higher education. Lastly, conditioning on the permutation blocks eliminates these discrepancy between treatment arms.

Table D.1: - Unconditional Block Permutation Inference on Baseline Variables by Treatment Status at Age 31

Variables	Sample		Control	Treatment	Effect	Asymptotic (two-sided)		Permutation (two-sided)	
	#C	#T	Mean	Effects	Size	t-stat	Single p-val	Single p-val	Stepdown
Child characteristics on Enrolment									
Enrollment Age in Months	47	48	19.13	−0.12	−0.03	−0.14	0.893	0.893	0.893
Gender	47	48	1.45	0.03	0.06	0.31	0.755	0.750	0.937
Birth Weight	47	47	2.95	−0.14	−0.31	−1.44	0.154	0.150	0.592
Z-score (Weight for Height) at Onset	47	48	−0.89	−0.30	−0.42	−2.05	**0.043**	**0.046**	0.269
Height-for-age	47	48	−2.87	−0.11	−0.17	−0.86	0.390	0.393	0.857
Enrollment DQ	47	48	97.06	1.96	0.21	1.05	0.295	0.298	0.801
Housing Score at Onset	47	48	7.51	−0.28	−0.15	−0.83	0.407	0.415	0.796
Rank Mean	47	48	0.52	−0.03	−0.29	−1.46	0.149	0.154	-
Parental Characteristics on Enrolment									
HOME Score at Onset	47	48	17.06	−1.25	−0.25	−1.39	0.169	0.171	0.523
Mothers PPVT at Onset	47	48	85.36	0.16	0.01	0.04	0.969	0.972	0.972
Young Mother Indicator	47	48	0.23	0.02	0.04	0.18	0.858	0.853	0.980
Mother Working Indicator	47	48	0.17	0.08	0.21	0.95	0.345	0.358	0.732
Mother Education Indicator	47	48	0.23	−0.15	−0.35	−2.04	**0.044**	**0.048**	0.214
Rank Mean	47	48	0.51	−0.02	−0.15	−0.79	0.431	0.433	-
Follow-up Characteristics									
Age at the 30-year Survey	47	48	31.79	0.01	0.02	0.07	0.942	0.941	0.997
Migrant Indicator at age 30	47	48	0.17	−0.00	−0.01	−0.05	0.964	0.955	0.955
Housing (factor score)	47	47	−0.08	0.12	0.11	0.59	0.557	0.560	0.910
Rank Mean	47	48	0.50	0.01	0.04	0.18	0.859	0.858	-

The columns of this table presents the following information. **Col.1**: variable of interest; **Col.2**: Control sample size; **Col.3**: Treated sample size; **Col.4**: control mean; **Col.5**: estimated treatment effect; **Col.6**: Hedges g effect size according to Rosenthal and Rosnow (1991) and Becker (2000). **Col.7**: t-statistic associated with the treatment effect. **Col.8**: asymptotic two-sided p-value for the single hypothesis testing of no treatment effect. **Col.9**: the single hypothesis two-sided mid-p-value based on 15.000 permutations draws. Test statistic uses the pre-pivoted treatment effect estimate and the permutation scheme is either a naïve or block permutation. **Col.10**: the multiple hypothesis testing (stepdown) for p-values in column 10. The last variable of each group of outcome consists of the average rank of each participant across the outcomes. Estimates in this table are not conditioned on baseline variables.

Table D.1 has shown that the difference on migration by treatment status is not statistically significant for the joint data set. It is useful to examine if baseline variables differ between migrants and non-migrants. Table D.2 investigates the difference of baseline variables by migration status. It displays the permutation test on migrant status and show that none of the baseline variables are statistically different between migrants and non-migrants. Table D.3 shows no treatment effect on mortality nor migration for the full sample. We do observe treatment effects on migration conditioned on gender. Namely, treated females are more likely to migrate than the controls. On the other hand, treated males are less likely to migrate when compared to the control ones.

Table D.2: - Conditional Block Permutation Inference on Baseline Variables by Migration Status at Age 31

	Sample		Non-migrants	Migrant	Effect	Asymptotic (two-sided)		Permutation (two-sided)	
Variables	#N	#M	Mean	Effects	Size	t-stat	Single p-val	Single p-val	Stepdown
Child characteristics on Enrollment									
Enrollment Age in Months	79	16	19.21	−0.80	−0.28	−0.83	0.410	0.467	0.845
Gender	79	16	1.46	0.03	0.16	0.51	0.614	0.722	0.722
Birth Weight	79	15	2.90	−0.12	−0.27	−0.81	0.418	0.459	0.912
Z-score (Weight for Height) at Onset	79	16	−1.08	0.23	0.32	1.07	0.286	0.259	0.830
Height-for-age	79	16	−2.95	0.12	0.22	0.71	0.478	0.458	0.953
Enrollment DQ	79	16	97.49	3.32	0.43	1.36	0.178	0.185	0.733
Housing Score at Onset	79	16	7.42	−0.31	−0.19	−0.64	0.526	0.549	0.794
Rank Mean	79	16	0.50	0.02	0.21	0.68	0.496	0.514	-
Parental Characteristics on Enrolment									
HOME Score at Onset	79	16	16.52	−0.54	−0.13	−0.42	0.675	0.678	0.678
Mothers PPVT at Onset	79	16	85.93	−2.90	−0.16	−0.51	0.612	0.612	0.853
Young Mother Indicator	79	16	0.26	−0.12	−0.30	−0.97	0.333	0.379	0.907
Mother Working Indicator	79	16	0.22	−0.06	−0.15	−0.48	0.636	0.585	0.932
Rank Mean	79	16	0.51	−0.03	−0.38	−1.25	0.215	0.235	-

The columns of this table presents the following information. **Col.1**: variable of interest; **Col.2**: Not sample size; **Col.3**: Migrant sample size; **Col.4**: control mean; **Col.5**: estimated treatment effect; **Col.6**: Hedges g effect size according to Rosenthal and Rosnow (1991) and Becker (2000). **Col.7**: t-statistic associated with the mean difference between migrants and non-migrants. **Col.8**: asymptotic two-sided p-value for the single hypothesis testing of the null hypothesis that the difference in means btween migrants and non-migrants is zero. **Col.9**: the single hypothesis two-sided mid-p-value based on 15.000 permutations draws. Test statistic uses the pre-pivoted treatment effect estimate and the permutation scheme is either a naïve or block permutation. **Col.10**: the multiple hypothesis testing (stepdown) for p-values in column 10. The last variable of each group of outcome consists of the average rank of each participant across the outcomes. Estimates in this table are conditioned on main baseline variables used in the randomization protocol.

Table D.3: Treatment Effects on Mortality and Migration

	Sample		Control	Treatment	Effect	**Asymptotic** (two-sided)		**Permutation** (two-sided)	
Variables	**#C**	**#T**	**Mean**	**Effects**	**Size**	***t*-stat**	**Single *p*-val**	**Single *p*-val**	**Stepdown**
Full Sample									
Migrant at 30 y.o.	47	48	0.16	0.02	0.06	0.27	0.790	0.793	0.793
Dead at Age 30	65	62	0.11	−0.04	−0.13	−0.76	0.451	0.340	0.565
Rank Mean	65	62	0.52	−0.01	−0.09	−0.50	0.618	0.539	-
Female Sample									
Migrant at 30 y.o.	21	23	−0.02	0.26	2.01	2.98	**0.005**	**0.005**	**0.006**
Dead at Age 30	27	28	0.04	0.00	0.01	0.03	0.973	0.688	0.688
Rank Mean	27	28	0.48	0.05	0.59	1.72	**0.091**	0.122	-
Male Sample									
Migrant at 30 y.o.	26	25	0.32	−0.20	−0.50	−1.78	**0.082**	**0.085**	0.162
Dead at Age 30	38	34	0.16	−0.07	−0.21	−0.91	0.368	0.297	0.297
Rank Mean	38	34	0.56	−0.07	−0.41	−1.70	**0.095**	0.112	-

The columns of this table presents the following information. **Col.1**: variable of interest; **Col.2**: Control sample size; **Col.3**: Treated sample size; **Col.4**: control mean; **Col.5**: estimated treatment effect; **Col.6**: effect size according to the pooled standard deviation (Hedge's g); **Col.7**: t-statistic associated with the treatment effect; **Col.8**: asymptotic two-sided p-value for the single hypothesis testing of no treatment effect; **Col.9**: the single hypothesis two-sided mid-p-value based on 15.000 permutations draws; The inference is based on a t-statistic that uses pre-pivoted treatment effect estimate and on a block permutation scheme; **Col.10**: the multiple hypothesis testing (stepdown) for p-values in column 9. Estimates are based on a block permutation inference conditional on main variables at the onset of the intervention. The last row of each block of variables presents the inference on a summary index generated by the average rank of each participant across the outcome variables.

Appendix E Augmented Inverse Propensity Weighting

We correct for any potential attrition bias using statistical models that adjust missing data using observed covariates. Specifically, we retrieve statistics for the full outcome distribution through reweighing non-missing observations according to their likelihood of compliance. To do so, we use Augmented Inverse Propensity Weighting (AIPW) (Glynn and Quinn (2010); Robins et al. (1994)). The AIPW is extends standard Inverse Propensity Weighting (IPW). AIPW improves upon the standard IPW by exploiting the predictive information of conditioning variables at baseline about the outcome variable.

The AIPW formula for ATE can be described as following:

$$\widehat{ATE}_{AIPW} = \frac{(Y_i \cdot \mathit{1}[D_i = 1] \cdot \mathit{1}[A_i = 1] - (\mathit{1}[D_i = 1] \cdot \mathit{1}[A_i = 1] - \xi_{i,1})\hat{y}_{i,1}) \cdot \omega_{i,1}}{N_1} \tag{1}$$

$$- \sum_{i=1}^{N} \frac{\Big(Y_i \cdot \mathit{1}[D_i = 0] \cdot \mathit{1}[A_i = 1] - (\mathit{1}[D_i = 0] \cdot \mathit{1}[A_i = 1] - \hat{\xi}_{i,0})\hat{y}_{i,0}\Big) \cdot \omega_{i,0}}{N_0} \tag{2}$$

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

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

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

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

$$\xi_{i,d} = \Pr(D = d | A_i = 1, X_i, Z_i) \qquad d \in \{0, 1\}$$

$$y_{i,d} = E(Y_i | A_i = 1, D_i = d, X_i, Z_i) \qquad d \in \{0, 1\}$$

where N is the total sample size and $\hat{p}_{i,d}, \hat{\pi}_{i,d}, \hat{\xi}_{i,d}, \hat{y}_{i,d}$ are estimates for $p_{i,d}, \pi_{i,d}, \xi_{i,d}, y_{i,d}$ respectively. Probabilities are estimated using the logit regression while the outcome expectation is estimated using the OLS regression.

For the Jamaican intervention, probabilities $\hat{p}_{i,d}$ are estimated using Logit regression based on a selection of covariates that predict attrition. Our selection of AIPW covariates is based on a method that minimizes information criteria of the Logit estimation. Our selection of covariates for is also age and gender specific.

The AIPW covariate selection is as follows:

1. We first do inference on pre-program variables to select the ones that are statistically not balanced between the attrited and non-attrited groups for each age, gender and treatment status;

2. Our selection is based on a series of Logit regressions for each set statistically significant pre-program variables for each gender and age. The dependent variable of each Logit regression is the age-specific attrition indicator. Covariates of each Logit regression are the treatment status indicator and a subset the statistically significant pre-program variables evaluated in

item 1.

3. The small sample size of our data limits the number of possible covariates in the Logit regressions in item 2. An excessive number of covariates generates the exact forecast of the values of the attrition indicator. To solve this problem we limit the number of covariates to 3,4, 5 and 6 variables.

4. For each fixed number of covariates, we run a Logit regression of the attrition indicator on the treatment status and all possible combinations of the pre-program variables defined in item 1.

5. We then select the covariates associated with the lowest value of the Akaike Information Criteria among all combinations of Logit regressions.

6. Finally, out of the Logit regressions that generate the lowest value of information criteria, we select the maximum number of covariates that do not generate perfect forecast of the attrition indicator.

Three main baseline variables often reach the minimum information criteria. Those are age, height-to-weight and gender.

The empirical results that employ the method are displayed in Tables E.1–E.2. The results presented in these tables are close related to the results presented in the main paper. The AIPW estimates does not qualitatively change the empirical results of the main paper. This fact corroborates the empirical analysis of Tables C.2–C.3 which suggest that non-random attrition is not of primary concern in the Jamaican Intervention.

Table E.1: AIPW Estimates of Treatment Effects on Labor Outcomes (All Data)

Variables	Sample		Control	Treatment	Effect	Asymptotic (one-sided)		Permutation (one-sided)	
	#C	#T	Mean	Effects	Size	t-stat	Single p-val	Single p-val	Stepdown
Log of Wage & Earnings (no outlier)									
Log Daily Wage	35	40	3.02	0.33	0.32	1.54	**0.06**	**0.01**	**0.02**
Log Total Earnings (no outlier)	44	46	5.96	0.32	0.28	1.42	**0.08**	**0.04**	**0.04**
Rank Mean	44	46	0.42	0.11	0.37	1.72	**0.04**	**0.02**	-
Schooling									
Years of Education	47	48	10.77	0.63	0.36	1.72	**0.04**	**0.02**	**0.06**
Any college education?	47	48	0.13	0.10	0.30	1.31	**0.10**	**0.03**	**0.06**
Years of college education	47	48	0.34	0.31	0.28	1.21	0.11	0.12	0.12
Higher education diploma ?	47	48	0.13	0.18	0.30	1.22	0.11	**0.04**	**0.06**
Rank Mean	47	48	0.47	0.05	0.35	1.47	**0.07**	**0.03**	-
Employment									
Working for someone	47	48	0.68	−0.04	−0.08	−0.38	0.65	0.57	0.87
Ever Worked for Someone	47	48	0.74	0.11	0.24	1.28	0.10	0.13	0.36
Currently Self-employed	47	48	0.43	−0.13	−0.29	−1.43	0.92	0.81	0.94
Working OR Self-employed	47	48	0.91	−0.14	−0.54	−2.06	0.98	0.89	0.89
Rank Mean	47	48	0.51	−0.03	−0.22	−1.01	0.84	0.74	-

This table evaluates treatment effects using the AIPW method. The columns of this table present the following information. **Col.1**: variable of interest; **Col.2**: Control sample size; **Col.3**: Treated sample size; **Col.4**: control mean; **Col.5**: estimated treatment effect; **Col.6**: effect size according to the pooled standard deviation (Hedge's g); **Col.7**: t-statistic associated with the treatment effect; **Col.8**: asymptotic one-sided p-value for the single hypothesis testing of no treatment effect; **Col.9**: the single hypothesis one-sided mid-p-value based on 15.000 permutations draws; Test statistic uses the pre-pivoted treatment effect estimate and a block permutation scheme; **Col.10**: the multiple hypothesis testing (stepdown) for p-values in column 9; The last variable of each group of outcome consists of the average rank of each participant across the outcomes. Estimates are based on a block permutation inference conditional on main variables at the onset of the intervention.

Table E.2: AIPW Estimates of Treatment Effects on Labor Outcomes for Non-migrants and by Gender

	Reverse	**Non-migrants (Sample N_C =39, N_T =40)**				**Males (Sample N_C =26, N_T =25)**				**Females (Sample N_C =21, N_T =23)**			
	Variable	**Treat. Effect**	**Effect Size**	**Permutation p-val**	**Permutation Stepdown**	**Treat. Effect**	**Effect Size**	**Permutation p-val**	**Permutation Stepdown**	**Treat. Effect**	**Effect Size**	**Permutation p-val**	**Permutation Stepdown**
Log of Wage & Earnings (no outlier)													
Log Daily Wage	No	0.46	0.67	**0.00**	**0.01**	0.16	0.14	0.31	0.40	0.61	1.24	**0.00**	**0.00**
Log Total Earnings (no outlier)	No	0.37	0.42	**0.04**	**0.04**	0.06	0.05	0.36	0.36	0.63	0.68	**0.01**	**0.01**
Rank Mean	No	0.14	0.48	**0.03**	–	0.04	0.12	0.32	–	0.16	0.65	**0.04**	–
Schooling													
Years of Education	No	0.94	0.76	**0.00**	**0.00**	−0.04	−0.02	0.28	0.48	1.39	1.50	**0.02**	**0.04**
Any college education?	No	0.15	0.55	**0.01**	**0.01**	−0.07	−0.17	0.47	0.56	0.30	∞	**0.00**	**0.01**
Years of college education	No	0.65	1.87	**0.00**	**0.00**	−0.30	−0.20	0.68	0.68	1.00	∞	**0.02**	**0.03**
Higher education diploma ?	No	0.35	∞	**0.00**	**0.00**	−0.03	−0.04	0.37	0.53	0.43	∞	**0.05**	**0.05**
Rank Mean	No	0.08	0.80	**0.01**	–	−0.02	−0.08	0.34	–	0.12	2.55	**0.01**	–
Employment													
Working for someone	No	−0.07	−0.15	0.76	0.87	−0.01	−0.02	0.49	0.62	−0.05	−0.10	0.45	0.80
Ever Worked for Someone	No	0.11	0.24	0.21	0.54	0.07	0.21	0.22	0.57	0.16	0.34	0.11	0.32
Currently Self-employed	No	−0.09	−0.18	0.64	0.92	0.01	0.03	0.46	0.78	−0.31	−0.64	0.98	0.98
Working OR Self-employed	No	−0.17	−0.64	0.95	0.95	−0.08	−0.32	0.76	0.76	−0.21	−0.77	0.88	0.97
Rank Mean	No	−0.03	−0.25	0.70	–	−0.00	−0.01	0.42	–	−0.05	−0.41	0.87	–

This table evaluates treatment effects using the AIPW method. The columns of this table present the following information. **Col.1**: variable of interest; **Col.2**: indicates if the variable is reverse, that is multiplied by -1; **Col.3**: estimated treatment effect for nonmigrants; **Col.4**: Hedges g effect size for non-migrants. **Col.5**: the single hypothesis one-sided mid-p-value based on 15.000 permutations draws. Test statistic uses the pre-pivoted treatment effect estimate and a block permutation scheme. **Col.6**: the multiple hypothesis testing (stepdown) for p-values for non-migrants. **Col.7**: treatment effect for males; **Col.8**: effect size for males; **Col.9**: mid-p-value for males; **Col.10**: stepdown p-values for males. **Col.11**: treatment effect for females; **Col.12**: effect size for females; **Col.13**: mid-p-value for females; **Col.14**: stepdown p-values for females. Estimates are based on a block permutation inference conditional on main variables at the onset of the intervention.

Appendix F Outliers

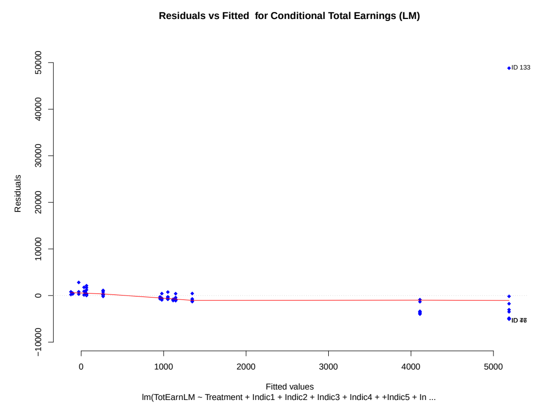
The statistical literature offers a range of methods to detect data points that can be labeled as outliers. Three commonly used methods are the analysis of the residual plot, the use of Cook's Distance and the Leverage Index (Rousseeuw and Leroy (1996)).

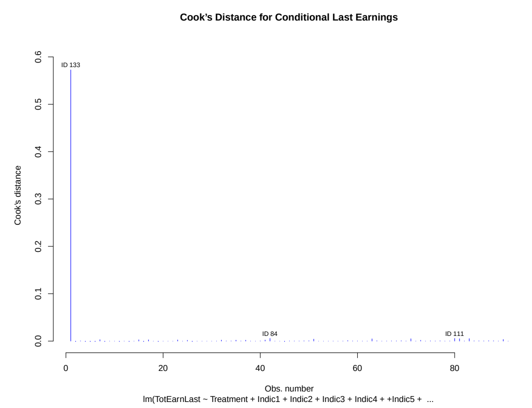
All 3 outlier detection methods identify a single outlier in the earnings data and none in the wage data. This observation is a male migrant in the control group, whose income is 35 times bigger than the sample average. Figure F.1 presents three graphs associated with the residual plot, the use of Cook's Distance and the Leverage Index.

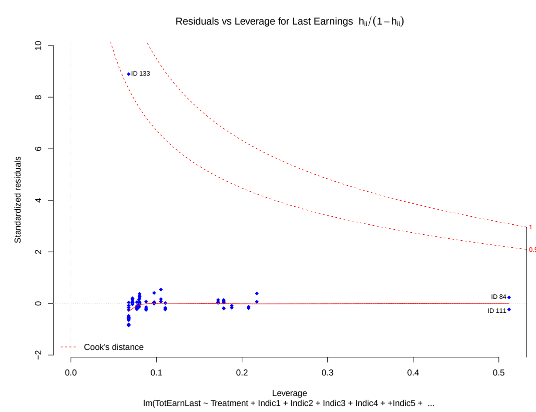
We take 2 approaches to addressing potential bias from this outlier. First, we to drop the data point and employ the t-statistic associated with conditional difference in means between treatment and control groups as the test statistic. Second, we use a test statistic that is less affected by the presence of outliers. Specifically, the Mann-Whitney-Wilcoxon (rank-sum) test statistic is robust to the presence of outliers. It is based on the rank of the data instead of its numerical value (see, for instance, Boos and Stefanski (2012) for a description of the rank-sum test). Table F.1 presents inference on the income treatment effects for the full data set and the subsets of non-migrants, males and females. Table F.2 presents the results for schooling outcomes while Table F.3 investigates the employment outcomes. The inference results on these tables not only corroborate our main results but also yield sharper inference results.

Finally, we also employ the generalized Rank-sum tests of Conover and Salsburg (1988). The method offers a series of statistics that can be interpreted as a smooth transition between the t-statistic associated with the conditional difference-in-means between treatment and control groups and rank-sum statistic. The inference tables comply qualitatively with the empirical results of the main and are available under request. As expected, the inferences using the generalized Rank-sum tests can be placed between those generated by the difference-in-means statistic and the rank-sum test.

Figure F.1: Outlier Analysis for Earnings Outcome: Residual Plot, Cook's Distance, Leverage Index.







This figure presents three graphs that are useful to evaluate the presence of outliers in the earnings data. The first graph plots the residuals form a standard regression that evaluates a linear regression model of earnings data conditional on baseline variables. The second graph displays the Cook's Distance statistics for each of the data points. The third graph employs the Leverage index Analysis for the selection of outlier data points. The conclusion of each of the analysis is the same. All methods show that a single data point is consistently labeled as an outlier for all the statistical methods. This data point consists of the earnings of a migrant control male whose income is 35 times higher than the sample mean of the earnings data.

Table F.1: Treatment Effects on Wage and Earnings Using Rank-sum Statistic

Variables	Sample		Control	Treatment	Effect	Rank-sum Statistics		Permutation (one-sided)	
	#C	#T	Mean	Effects	Size	Z-value	p-val	Single p-val	Stepdown
Full Sample									
Log of Daily Wage	35	40	2.98	0.40	0.43	1.91	**0.03**	**0.02**	**0.03**
Log of Monthly Total Earnings	44	46	5.92	0.39	0.36	1.66	**0.05**	**0.03**	**0.03**
Rank Mean	44	46	0.44	0.12	0.44	1.74	**0.04**	**0.02**	**0.02**
Non-migrant Sample									
Log of Daily Wage	29	33	2.66	0.50	0.68	2.54	**0.01**	**0.01**	**0.02**
Log of Monthly Total Earnings	37	38	5.66	0.39	0.42	1.88	**0.03**	**0.05**	**0.05**
Rank Mean	37	38	0.41	0.14	0.49	1.97	**0.02**	**0.04**	**0.04**
Female Sample)									
Log of Daily Wage	13	18	2.67	0.71	1.13	1.91	**0.03**	**0.02**	**0.04**
Log of Monthly Total Earnings	20	22	5.54	0.70	0.69	1.75	**0.04**	**0.03**	**0.03**
Rank Mean	20	22	0.38	0.18	0.69	1.78	**0.04**	**0.03**	**0.03**
Male Sample									
Log of Daily Wage	22	22	3.19	0.17	0.16	0.60	0.27	0.29	0.36
Log of Monthly Total Earnings	24	24	6.26	0.09	0.08	0.40	0.34	0.34	0.34
Rank Mean	24	24	0.46	0.04	0.13	0.46	0.32	0.31	0.31

The columns of this table present the following information. **Col.1**: variable of interest; **Col.2**: Control sample size; **Col.3**: Treated sample size; **Col.4**: control mean; **Col.5**: estimated treatment effects of the log-transformed variables, which can be interpreted as the percentage increase of the outcome mean between treated and control group; **Col.6**: effect size according to the pooled standard deviation (Hedge's g); **Col.7**: Z-value is an approximate test statistic for the rank-sum inference; **Col.8**: asymptotic one-sided p-value based on the rank-sum statistics; **Col.9**: the single hypothesis one-sided mid-p-value permutation test generated upon 15.000 permutations draws; The adopted test statistic is the pre-pivoted rank-sum statistic based a block permutation scheme; **Col.10**: the multiple hypothesis testing (stepdown) for p-values in column 9. Estimates are based on a block permutation inference conditional on main variables at the onset of the intervention. We suppress an outlier in the earnings data whose value is 35 times higher than the average earnings.

Table F.2: Treatment Effects on Schooling Using Rank-sum Statistic

Variables	Sample		Control	Treatment	Effect	Rank-sum Statistics		Permutation (one-sided)	
	#C	#T	Mean	Effects	Size	Z-value	p-val	Single p-val	Stepdown
Full Sample									
Years of Education	47	48	10.69	0.77	0.46	1.04	0.15	**0.08**	**0.08**
Any college education?	47	48	0.11	0.14	0.43	1.28	0.10	**0.04**	**0.09**
Years of college education	47	48	0.30	0.38	0.32	1.36	**0.09**	**0.05**	**0.09**
Higher education diploma ?	47	48	0.09	0.26	0.42	1.31	**0.10**	**0.03**	**0.07**
Rank Mean	47	48	0.47	0.07	0.44	1.14	0.13	**0.05**	**0.05**
Non-migrant Sample									
Years of Education	39	40	10.43	1.14	0.87	1.41	**0.08**	**0.06**	**0.06**
Any college education?	39	40	0.06	0.18	0.68	1.81	**0.03**	**0.02**	**0.05**
Years of college education	39	40	0.01	0.77	1.52	2.30	**0.01**	**0.00**	**0.01**
Higher education diploma ?	39	40	−0.05	0.45	2.08	2.49	**0.01**	**0.00**	**0.01**
Rank Mean	39	40	0.45	0.10	0.90	1.43	**0.08**	**0.05**	**0.05**
Female Sample									
Years of Education	21	23	10.55	1.33	1.18	1.60	**0.05**	**0.06**	**0.06**
Any college education?	21	23	−0.00	0.31	2.18	2.71	**0.00**	**0.00**	**0.01**
Years of college education	21	23	0.01	0.98	2.46	2.47	**0.01**	**0.01**	**0.03**
Higher education diploma ?	21	23	−0.03	0.49	1.86	1.96	**0.03**	**0.03**	**0.08**
Rank Mean	21	23	0.44	0.12	1.56	1.78	**0.04**	**0.03**	**0.03**
Male Sample									
Years of Education	26	25	10.82	0.25	0.13	−0.08	0.53	0.32	0.45
Any college education?	26	25	0.20	−0.01	−0.02	−0.64	0.74	0.53	0.53
Years of college education	26	25	0.56	−0.18	−0.12	−0.43	0.66	0.56	0.57
Higher education diploma ?	26	25	0.20	0.03	0.04	0.05	0.48	0.30	0.50
Rank Mean	26	25	0.49	0.01	0.08	−0.06	0.52	0.29	0.29

The columns of this table present the following information. **Col.1**: variable of interest; **Col.2**: Control sample size; **Col.3**: Treated sample size; **Col.4**: control mean; **Col.5**: estimated treatment effect; **Col.6**: effect size according to the pooled standard deviation (Hedge's g); **Col.7**: Z-value is an approximate test statistic for the rank-sum inference; **Col.8**: asymptotic one-sided p-value based on the rank-sum statistics; **Col.9**: the single hypothesis one-sided mid-p-value permutation test generated upon 15.000 permutations draws; Test statistic uses the pre-pivoted rank-sum statistic based on a block permutation scheme; **Col.10**: the multiple hypothesis testing (stepdown) for p-values in column 9. Estimates are based on a block permutation inference conditional on main variables at the onset of the intervention. We suppress an outlier in the earnings data whose value is 35 times higher than the average earnings.

Table F.3: Treatment Effects on Employment Using Rank-sum Statistic

Variables	Sample #C	Sample #T	Control Mean	Treatment Effects	Effect Size	Rank-sum Statistics Z-value	Rank-sum Statistics p-val	Permutation (one-sided) Single p-val	Permutation (one-sided) Stepdown
Full Sample									
Working for someone	47	48	0.66	−0.00	−0.01	−0.36	0.64	0.52	0.87
Non-migrant Sample									
Working for someone	39	40	0.68	−0.05	−0.12	−0.63	0.74	0.67	0.93
Female Sample									
Working for someone	21	23	0.56	−0.03	−0.06	−0.34	0.63	0.56	0.91
Male Sample									
Working for someone	26	25	0.75	0.03	0.06	−0.09	0.54	0.43	0.72

The columns of this table present the following information. **Col.1**: variable of interest; **Col.2**: Control sample size; **Col.3**: Treated sample size; **Col.4**: control mean; **Col.5**: estimated treatment effect; **Col.6**: effect size according to the pooled standard deviation (Hedge's g); **Col.7**: Z-value is an approximate test statistic for the rank-sum inference. **Col.8**: asymptotic one-sided p-value based on the rank-sum statistics. **Col.9**: the single hypothesis one-sided mid-p-value permutation test generated upon 15.000 permutations draws. Test statistic uses the pre-pivoted rank-sum statistic based on a block permutation scheme. **Col.10**: the multiple hypothesis testing (stepdown) for p-values in column 7. Estimates are based on a block permutation inference conditional on main variables at the onset of the intervention.

Appendix G Nonstunted Comparison Sample and Catch-up Analysis

The Jamaican Study enrolled an additional sample of nonstunted children for comparison purposes. The initial comparison sample consists of 32 children nonstunted children living in the same area of the stunted children. The participants of the comparison sample were selected to match the age (plus or minus 3 months) and sex of the stunted participants. At age 7, the nonstunted sample was supplemented with another 52 children who had been identified in the initial survey as being nonstunted and fulfilled all other inclusion criteria. The number of participants in the nonstunted comparison group totals 84 children. Nonstunted participants did not receive any intervention, but did receive the same free health care as those in the stunted experimental group. The group has been followed from age 7 onwards. The properties of these nonstunted participants are described in Gertler et al. (2014).

We assess the degree to which the intervention enabled the stunted treatment group to catch-up to the nonstunted comparison group. We compare the income and schooling outcomes. The comparison between the nonstunted comparison group and the stunted control group is useful to measure how disadvantaged are the control participants with respect to nonstunted participants. The comparison between the nonstunted comparison group and the stunted treatment group is useful to examine if the treatment is able to make treated participants to catch-up with the nonstunted participants.

At age 31, we found and interviewed 64 children out of the 84 children of the original sample of nonstunted participants. The interviewed sample of nonstunted participants is almost identical to the one examined at age-22 survey. Specifically, at age 22, 65 out of the 84 non-stunted participants were interviewed. At age 31, only one additional participant was not interviewed.

Tables G.1–G.3 examine the hypothesis whether the treated could make stunted individuals catch-up with nonstunted ones. We employ the same methods utilized in the tables of the main paper.

The first panel of Table G.1 compares the earnings outcomes for nonstunted participants with stunted control groups. The second panel compares the non-stunted versus stunted treatment group. The difference on earning between nonstunted and either control or treatment groups remains statistically significant, but the effect sizes associated with the treated group are smallers than those of the control group. The third panel focus on the schooling outcomes and shows that the difference on schooling outcomes between nonstunted and controls are statistically significant. Panel four shows that difference on schooling outcomes between nonstunted and treated participants are not statistically significant. We conclude that the treated group catch-ups with the nonstunted participants in schooling outcomes.

Table G.2 focuses on females only. The first and third panels compares earnings and schooling outcomes for nonstunted females versus stunted control females respectively. We see that the difference between nonstunted females and stunted control females is statistically significant. The second and fourth panels compares earnings and schooling outcomes for nonstunted females versus stunted treated females respectively. We observed that difference of conditional means are not statistically significant. We conclude that females catch-up on income and schooling outcomes. Table G.3 focuses on males only. The data has to much variance for such small sizes to produce conclusive results.

Table G.1: Comparison between Non-stunted versus Control and Non-stunted versus Treated for All Data

Variables	Sample #S	Sample #N	Control Mean	Non-stunted Mean Difference	Effect Size	Asymptotic (two-sided) t-stat	Asymptotic (two-sided) Single p-val	Permutation (two-sided) Single p-val	Permutation (two-sided) Stepdown
Log of Wage & Earnings, Non-stunted versus Control, All Data									
Log Daily Wage	35	49	2.84	0.94	0.85	2.75	**0.01**	**0.01**	**0.01**
Log Total Earnings (no outlier)	44	61	5.68	1.10	0.94	3.41	**0.00**	**0.00**	**0.00**
Rank Mean	44	61	0.33	0.26	0.95	3.45	**0.00**	**0.00**	-
Log of Wage & Earnings, Non-stunted versus Treated, All Data									
Log Daily Wage	40	49	3.19	0.59	0.59	1.88	**0.06**	**0.06**	**0.06**
Log Total Earnings (no outlier)	46	61	6.02	0.75	0.69	2.44	**0.02**	**0.01**	**0.02**
Rank Mean	46	61	0.37	0.19	0.68	2.43	**0.02**	**0.02**	-
Schooling, Non-stunted versus Control, All Data									
Years of Education	47	63	10.51	1.21	0.63	2.35	**0.02**	**0.02**	**0.05**
Any college education?	47	64	0.10	0.19	0.51	1.83	**0.07**	**0.07**	0.12
Higher education diploma ?	47	64	0.07	0.23	0.34	1.23	0.22	0.22	0.22
Rank Mean	47	64	0.45	0.09	0.56	2.05	**0.04**	**0.04**	-
Schooling Outcomes, Non-stunted versus Treated, All Data									
Years of Education	48	63	11.18	0.51	0.26	0.97	0.33	0.33	0.58
Any college education?	48	64	0.21	0.07	0.16	0.58	0.56	0.57	0.78
Higher education diploma ?	48	64	0.28	−0.01	−0.01	−0.03	0.98	0.98	0.98
Rank Mean	48	64	0.48	0.03	0.17	0.62	0.53	0.54	-

The columns of this table presents the following information. **Col.1**: variable of interest; **Col.2**: Stunted control or stunted treated sample sizes; **Col.3**: Non-stunted sample size; **Col.4**: control and treated conditional mean; **Col.5**: estimated conditional mean difference; **Col.6**: Cohen-d effect size with the pooled standard deviation according to Rosenthal and Rosnow (1991) and Becker (2000). **Col.7**: t-statistic associated with the conditional mean difference. **Col.8**: asymptotic two-sided p-value for the single hypothesis testing of no treatment effect. **Col.9**: the single hypothesis two-sided mid-p-value based on 15.000 permutations draws. Test statistic uses the pre-pivoted treatment effect estimate and the permutation scheme is either a naïve or block permutation. **Col.10**: the multiple hypothesis testing (stepdown) for p-values in column 10. The last variable of each group of outcome consists of the average rank of each participant across the outcomes. Estimates are based on a block permutation inference conditional on main variables at the onset of the intervention.

Table G.2: Comparison between Non-stunted versus Control and Non-stunted versus Treated for Females

Variables	Sample #S	Sample #N	Control Mean	Non-stunted Mean Difference	Effect Size	Asymptotic (two-sided) t-stat	Asymptotic (two-sided) Single p-val	Permutation (two-sided) Single p-val	Permutation (two-sided) Stepdown
Log of Wage & Earnings, Non-stunted versus Control, Females									
Log Daily Wage	13	24	2.48	1.22	1.34	2.32	**0.03**	**0.02**	**0.02**
Log Total Earnings (no outlier)	20	29	5.34	1.46	1.27	3.19	**0.00**	**0.00**	**0.01**
Rank Mean	20	29	0.31	0.30	1.14	2.76	**0.01**	**0.01**	-
Log of Wage & Earnings, Non-stunted versus Treated, Females									
Log Daily Wage	18	24	3.27	0.35	0.34	0.67	0.51	0.49	0.49
Log Total Earnings (no outlier)	22	29	6.08	0.65	0.58	1.42	0.16	0.15	0.22
Rank Mean	22	29	0.40	0.14	0.52	1.26	0.21	0.19	-
Schooling, Non-stunted versus Control, Females									
Years of Education	21	31	10.27	1.83	1.19	2.80	**0.01**	**0.00**	**0.01**
Any college education?	21	31	−0.03	0.37	1.18	2.59	**0.01**	**0.01**	**0.01**
Higher education diploma ?	21	31	−0.09	0.44	0.83	1.80	**0.08**	0.17	0.17
Rank Mean	21	31	0.40	0.17	1.25	2.83	**0.01**	**0.01**	-
Schooling, Non-stunted versus Treated, Females									
Years of Education	23	31	11.70	0.40	0.18	0.49	0.63	0.63	0.90
Any college education?	23	31	0.32	−0.00	−0.00	−0.01	1.00	1.00	1.00
Higher education diploma ?	23	31	0.41	−0.11	−0.12	−0.31	0.75	0.73	0.91
Rank Mean	23	31	0.49	0.02	0.10	0.27	0.79	0.79	-

The columns of this table presents the following information. **Col.1**: variable of interest; **Col.2**: Stunted control or stunted treated sample sizes; **Col.3**: Non-stunted sample size; **Col.4**: control and treated conditional mean; **Col.5**: estimated conditional mean difference; **Col.6**: Cohen-d effect size with the pooled standard deviation according to Rosenthal and Rosnow (1991) and Becker (2000). **Col.7**: t-statistic associated with the conditional mean difference. **Col.8**: asymptotic two-sided p-value for the single hypothesis testing of no treatment effect. **Col.9**: the single hypothesis two-sided mid-p-value based on 15.000 permutations draws. Test statistic uses the pre-pivoted treatment effect estimate and the permutation scheme is either a naïve or block permutation. **Col.10**: the multiple hypothesis testing (stepdown) for p-values in column 10. The last variable of each group of outcome consists of the average rank of each participant across the outcomes. Estimates are based on a block permutation inference conditional on main variables at the onset of the intervention.

Table G.3: Comparison between Non-stunted versus Control and Non-stunted versus Treated for Males

Variables	Sample		Control	Non-stunted	Effect	Asymptotic (two-sided)		Permutation (two-sided)	
	#S	#N	Mean	Mean Difference	Size	t-stat	Single p-val	Single p-val	Stepdown
Log of Wage & Earnings, Non-stunted versus Control, Males									
Log Daily Wage	22	25	3.10	0.69	0.57	1.43	0.16	0.18	0.18
Log Total Earnings (no outlier)	24	32	6.01	0.72	0.59	1.51	0.14	0.15	0.21
Rank Mean	24	32	0.38	0.17	0.60	1.55	0.13	0.14	-
Log of Wage & Earnings, Non-stunted versus Treated, Males									
Log Daily Wage	22	25	3.21	0.63	0.66	1.67	0.10	**0.09**	**0.09**
Log Total Earnings (no outlier)	24	32	6.03	0.73	0.71	1.78	**0.08**	**0.07**	0.11
Rank Mean	24	32	0.36	0.21	0.74	1.92	**0.06**	**0.05**	-
Schooling, Non-stunted versus Control, Males									
Years of Education	26	32	10.70	0.64	0.30	0.79	0.43	0.45	0.74
Any college education?	26	33	0.18	0.07	0.17	0.46	0.65	0.60	0.82
Higher education diploma ?	26	33	0.22	0.00	0.01	0.01	0.99	1.00	1.00
Rank Mean	26	33	0.48	0.03	0.16	0.43	0.67	0.68	-
Schooling, Non-stunted versus Treated, Males									
Years of Education	25	32	10.69	0.61	0.35	0.90	0.37	0.38	0.64
Any college education?	25	33	0.13	0.11	0.29	0.76	0.45	0.43	0.61
Higher education diploma ?	25	33	0.17	0.07	0.11	0.28	0.78	0.76	0.76
Rank Mean	25	33	0.48	0.03	0.19	0.51	0.61	0.62	-

The columns of this table presents the following information. **Col.1**: variable of interest; **Col.2**: Stunted control or stunted treated sample sizes; **Col.3**: Non-stunted sample size; **Col.4**: control and treated conditional mean; **Col.5**: estimated conditional mean difference; **Col.6**: Cohen-d effect size with the pooled standard deviation according to Rosenthal and Rosnow (1991) and Becker (2000). **Col.7**: t-statistic associated with the conditional mean difference. **Col.8**: asymptotic two-sided p-value for the single hypothesis testing of no treatment effect. **Col.9**: the single hypothesis two-sided mid-p-value based on 15.000 permutations draws. Test statistic uses the pre-pivoted treatment effect estimate and the permutation scheme is either a naïve or block permutation. **Col.10**: the multiple hypothesis testing (stepdown) for p-values in column 10. The last variable of each group of outcome consists of the average rank of each participant across the outcomes. Estimates are based on a block permutation inference conditional on main variables at the onset of the intervention.

References

Andrew, A., Attanasio, O., Augsburg, B., Day, M., Grantham-McGregor, S., Meghir, C., Mehrin, F., Pahwa, S., and Rubio-Codina, M. (2019). Effects of a scalable home-visiting intervention on child development in slums of urban india: evidence from a randomised controlled trial. *Journal of Child Psychology and Psychiatry*.

Andrew, A., Attanasio, O., Fitzsimons, E., Grantham-McGregor, S., Meghir, C., and Rubio-Codina, M. (2018). Impacts 2 years after a scalable early childhood development intervention to increase psychosocial stimulation in the home: A follow-up of a cluster randomised controlled trial in Colombia. *PLOS Medicine*, 15(4):1–19.

Attanasio, O., Cattan, S., Fitzsimons, E., Meghir, C., and Rubio-Codina, M. (2020). Estimating the production function for human capital: results from a randomized controlled trial in colombia. *American Economic Review*, 110(1):48–85.

Boos, D. D. and Stefanski, L. A. (2012). *Essential Statistical Inference: Theory and Methods*. Springer.

Campbell, F. A., Conti, G., Heckman, J. J., Moon, S. H., and Pinto, R. (2013). The effects of early intervention on human development and social outcomes: Provisional evidence from ABC and CARE. University of Chicago, Department of Economics.

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*, 343(6178):1478–1485.

Carneiro, P. and Ginja, R. (2014). Long-term impacts of compensatory preschool on health and behavior: Evidence from Head Start. *American Economic Journal: Economic Policy*, 6(4):135–173.

Conover, W. J. and Salsburg, D. S. (1988). Locally most powerful tests for detecting treatment effects when only a subset of patients can be expected to "respond" to treatment. *Biometrics*, 44:189–196.

Conti, G., Heckman, J. J., and Pinto, R. (2016). The effects of two influential early childhood interventions on health and healthy behaviours. Economic Journal, 126(596):F28–F65.

Doyle, O. (2020). The first 2,000 days and child skills. *Journal of Political Economy*, 128(6).

Elango, S., , García, J. L., Heckman, J. J., and Hojman, A. (2016). Early childhood education. In Moffitt, R. A., editor, *Economics of Means-Tested Transfer Programs in the United States*, volume 2, chapter 4, pages 235–297. University of Chicago Press, Chicago.

Engle, P. L., Fernald, L. C. H., Alderman, H., Behrman, J., O'Gara, C., Yousafzai, A., Cabral de Mello, M., Hidrobo, M., Ulkuer, N., Ertem, I., and Iltus, S. (2011). Strategies for reducing inequalities and improving developmental outcomes for young children in low-income and middle-income countries. *The Lancet*, 378(9799):1339–1353.

García, J. L., Bennhoff, F., Ermini Leaf, D., and Heckman, J. J. (2021a). The dynastic benefits of early childhood education. Under Review.

García, J. L., Heckman, J. J., Leaf, D. E., and Prados, M. J. (2017). The life-cycle benefits of an influential early childhood program. Under Revision, *Journal of Political Economy.*

García, J. L., Heckman, J. J., Leaf, D. E., and Prados, M. J. (2018). Quantifying the life-cycle benefits of a prototypical early childhood program. Forthcoming at the *Journal of Political Economy*, 2020.

García, J. L., Heckman, J. J., and Ronda, V. (2021b). Early childhood education promotes the social mobility of disadvantaged african americans. Under Review.

García, J. L., Heckman, J. J., and Ziff, A. L. (2019). Early childhood education and crime. *Infant Mental Health Journal*, 40(1).

Gertler, P., Heckman, J. J., Pinto, R., Zanolini, A., Vermeersch, C., Walker, S., Chang, S., and Grantham-McGregor, S. M. (2014). Labor market returns to an early childhood stimulation intervention in Jamaica. *Science*, 344(6187):998–1001.

Glynn, A. N. and Quinn, K. M. (2010). An introduction to the augmented inverse propensity weighted estimator. *Political Analysis*, 18(1):36–56.

Grantham-McGregor, S., Adya, A., Attanasio, O., Augsburg, B., Behrman, J., Caeyers, B., Day, M., Jervis, P., Kochar, R., Makkar, P., et al. (2020). Group sessions or home visits for early childhood development in india: a cluster rct. *Pediatrics*, 146(6).

Grantham-McGregor, S. and Smith, J. A. (2016). Extending the jamaican early childhood development intervention. *Journal of Applied Research on Children: Informing Policy for Children at Risk*, 7(2).

Grantham-McGregor, S. M., Powell, C. A., Walker, S. P., and Himes, J. H. (1991). Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: The Jamaican study. *The Lancet*, 338(8758):1–5.

Gravetter, F. and Wallnau, L. (2014). *Essentials of statistics for the behavioral sciences (8th ed.).* Wadsworth., Belmont, CA.

Hamadani, J. D., Mehrin, S. F., Tofail, F., Hasan, M. I., Huda, S. N., Baker-Henningham, H., Ridout, D., and Grantham-McGregor, S. (2019). Integrating an early childhood development programme into bangladeshi primary health-care services: an open-label, cluster-randomised controlled trial. *The Lancet Global Health*, 7(3):e366 – e375.

Hamill, P. V., Drizd, T. A., Johnson, C. L., Reed, R. B., F., R. A., and Moore, W. M. (1979). Physical growth: National center for health statistics percentiles. *The American Journal of Clinical Nutrition*, 32(3):607–629.

Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312(5782):1900–1902.

Heckman, J. J. (2007). The economics, technology and neuroscience of human capability formation. *Proceedings of the National Academy of Sciences*, 104(3):13250–13255.

Heckman, J. J., Jagelka, T., and Kautz, T. (2019). Some contributions of economics to the study of personality.

Heckman, J. J. and Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, 19(4: European Association of Labour Economists 23rd Annual Conference, Paphos, Cyprus, 22–24th September 2011):451–464.

Heckman, J. J., Liu, B., Lu, M., and Zhou, J. (2021). Treatment effects and the measurement of skills in a prototypical home visiting program. Under Review.

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

Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., and Yavitz, A. Q. (2010b). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*, 94(1–2):114–128.

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*, 103(6):2052–2086.

Hoddinott, J., Maluccio, J. A., Behrman, J. R., Flores, R., and Martorell, R. (2008). Effect of a nutrition intervention during early childhood on economic productivity in Guatemalan adults. *The Lancet*, 371(9610):411–416.

Huber, M. (2012). Identifying causal mechanisms in experiments (primarily) based on inverse probability weighting. Discussion Paper 2012-13, University of St. Gallen, School of Economics and Political Science, Department of Economics.

Kautz, T., Heckman, J. J., Diris, R., ter Weel, B., and Borghans, L. (2014). Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. Technical report, Organisation for Economic Co-operation and Development, Paris. Available at https://www.oecd.org/edu/ceri/Fostering-and-Measuring-Skills-Improving-Cognitive-and-Non-Cognitive-Skills-to-Promote-Lifetime-Success.pdf.

Maluccio, J. A., Hoddinott, J., Behrman, J. R., Martorell, R., Quisumbing, A. R., and Stein, A. D. (2009). The impact of improving nutrition during early childhood on education among Guatemalan adults. *Economic Journal*, 119(537):734–763.

Network, W. B. H. D. (2014). Jamaica: Can disadvantage kids ever catch up with better-off peers. *Evidence to Policy series.*

Palmer, F. H. (1971). *Concept training curriculum for children ages two to five.* State University of New York at Stony Brook, Stony Brook, NY.

Richter, L. M., Desmond, C., Behrman, J., Britto, P., Daelmans, B., Devercelli, A. E., Gertler, P., Hoddinott, J., Fawzi, W., Fink, G., Lombardi, J., Boo, F. L., Lu, C., Lye, S., Nores, M., and Yousafzai, A. (2018). G20's initiative for early child development. *Lancet*, 392:2695–2696.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89(427):846–866.

Romano, J. P. and Wolf, M. (2005). Stepwise multiple testing as formalized data snooping. *Econometrica*, 73(4):1237–1282.

Rosenbaum, P. R. (2005). An exact distribution-free test comparing two multivariate distributions based on adjacency. *Journal of the Royal Statistical Society*, 67(4):515–530.

Rousseeuw, P. and Leroy, A. (1996). *Robust Regression and Outlier Detection (3rd ed.)*. John Wiley & Sons.

Rubio-Codina, M., Dormal, M., and Araujo, M. C. (2019). Observing home-visiting quality at scale with the home visit rating scales and a supervisor checklist in peru. *Infant mental health journal*, 40(3):343–362.

Smith, J. A., Baker-Henningham, H., Brentani, A., Mugweni, R., and Walker, S. P. (2018). Implementation of reach-up early childhood parenting programme: Appropriateness, acceptability and feasibility in Brazil and Zimbabwe. *Annals of the New York Academy of Sciences*, 1419(1):120–140.

Tanner, J. C., Candland, T., and Whitney, S. O. (2015). Later impacts of early childhood interventions: A systematic review. *World Bank Independent Evaluation Group, Working Paper 2015/3.*

Trochim, W. M. and Donnelly, J. P. (2006). *The research methods knowledge base (3rd ed.).* Cincinnati, OH.

Uzgiris, I. C. and Hunt, J. M. (1975). *Assessment in infancy: Ordinal scales of psychological development.* University of Illinois Press., Urbana, IL.

Walker, S., Grantham-McGregor, S., Powell, C., Himes, J., and Simeon, D. (1992). Morbidity and the growth of stunted and nonstunted children, and the effect of supplementation. *American Journal of Clinical Nutrition*, 56(3):504–510.

Walker, S., Powell, C., and Grantham-McGregor, S. (1990). Dietary intakes and activity levels of stunted and non-stunted children in Kingston, Jamaica. Part 1. Dietary intakes. *European Journal of Clinical Nutrition*, 44(7):527–534.

Walker, S., Wachs, T., Grantham-McGregor, S., Black, M., Nelson, C., Huffman, S., Baker-Henningham, H., Chang, S., Hamadani, J., Lozoff, B., Meeks Gardner, J., Powell, C., Rahman, A., and Richter, L. (2011a). Inequality in early childhood: risk and protective factors for early child development. *The Lancet*, 378(9799):1325–1338.

Walker, S. P., Chang, S. M., Powell, C. A., and Grantham-McGregor, S. M. (2005). Effects of early childhood psychosocial stimulation and nutritional supplementation on cognition and education in growth-stunted Jamaican children: Prospective cohort study. *The Lancet*, 366(9499):1804–1807.

Walker, S. P., Chang, S. M., Vera-Hernandez, M., and Grantham-McGregor, S. (2011b). Early childhood stimulation benets adult competence and reduces violent behavior. *Pediatrics*, 127(5):849–857.

Walker, S. P., Chang, S. M., Wright, A. S., Pinto, R., Heckman, J. J., and Grantham-McGregor, S. M. (2021). Cognitive, psychosocial, and behaviour gains at age 31 years from the jamaica early childhood stimulation trial. *Forthcoming, Journal of Child Psychiatry and Psychology.*

Walker, S. P., Powell, C. A., Grantham-McGregor, S. M., Himes, J. H., and Chang, S. M. (1991). Nutritional supplementation, psychosocial stimulation, and mental development of stunted children: the Jamaican study. *American Journal of Clinical Nutrition*, 54(4):642–648.

Walker, S. P., Wachs, T. D., Gardner, J. M., Lozoff, B., Wasserman, G. A., Pollitt, E., Carter, J. A., and The International Child Development Steering Group (2007). Child development: Risk factors for adverse outcomes in developing countries. *Lancet*, 369(9556):145–157.