Effect of the Jamaica Early Childhood Stimulation Intervention early childhood stimulation on schooling and labor market outcomes at age 31: Supporting Online Material

Paul Gertler
Haas School of Business
University of California, Berkeley
gertler@berkeley.edu

James Heckman
Department of Economics
University of Chicago
Center for the Economics
of Human Development
jjh@uchicago.edu

Rodrigo Pinto
Department of Economics
University of California, UCLA
rodrig@econ.ucla.edu

Susan M. Chang-Lopez
The University of The West Indies
susan.changlopez@uwimona.edu.jm

Sally Grantham-McGregor
University College London
sallymcgregor@yahoo.com

Christel Vermeersch
The World Bank
cvermeersch@worldbank.org

Susan Walker
The University of The West Indies
susan.walker@uwimona.edu.jm

Amika S. Wright
The University of the West Indies
amika.wright@uwimona.edu.jm

October 7, 2021

Contents

1 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. (2010, 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

2 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).

3 Attrition

Our analysis sample consists of 95 of the original 127 participants. Thirty-two (25%) participants were not interviewed at age 31. Table A.1 describes the attrition patterns and shows that they are surprisingly balanced across multiple surveys rounds and arms of the study. Table A.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 A.3 shows that baseline variables are also balanced across observed participants and the attrition group at age 30.

Table A.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 A.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 A.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 A.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 A.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.

4 Baseline Balance and Migration

The first and second panels in Table A.5 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 A.5 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 A.5 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 A.5: - Unconditional Block Permutation Inference on Baseline Variables by Treatment Status at Age 31

Variables	Sample #C	Sample #T	Control Mean	Treatment Effects	Effect Size	Asymptotic (two-sided) t-stat	Asymptotic (two-sided) Single p-val	Permutation (two-sided) Single p-val	Permutation (two-sided) 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 A.5 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 A.6 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 A.7 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 A.6: - Conditional Block Permutation Inference on Baseline Variables by Migration Status at Age 31

Variables	Sample		Non-migrants	Migrant	Effect	Asymptotic (two-sided)		Permutation (two-sided)	
	#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 A.7: Treatment Effects on Mortality and Migration

Variables	Sample #C	Sample #T	Control Mean	Treatment Effects	Effect Size	Asymptotic (two-sided) t-stat	Asymptotic (two-sided) Single p-val	Permutation (two-sided) Single p-val	Permutation (two-sided) 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.

5 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 A.8–A.9. 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 A.2–A.3 which suggest that non-random attrition is not of primary concern in the Jamaican Intervention.

Table A.8: AIPW Estimates of Treatment Effects on Labor Outcomes (All Data)

	Sample		**Control**	**Treatment**	**Effect**	**Asymptotic** (one-sided)		**Permutation** (one-sided)	
Variables	**#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 A.9: AIPW Estimates of Treatment Effects on Labor Outcomes for Non-migrants and by Gender

	Reverse Variable	Non-migrants (Sample N_C =39, N_T =40) Treat. Effect	Effect Size	Permutation p-val	Permutation Stepdown	Males (Sample N_C =26, N_T =25) Treat. Effect	Effect Size	Permutation p-val	Permutation Stepdown	Females (Sample N_C =21, N_T =23) 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.

6 Outliers

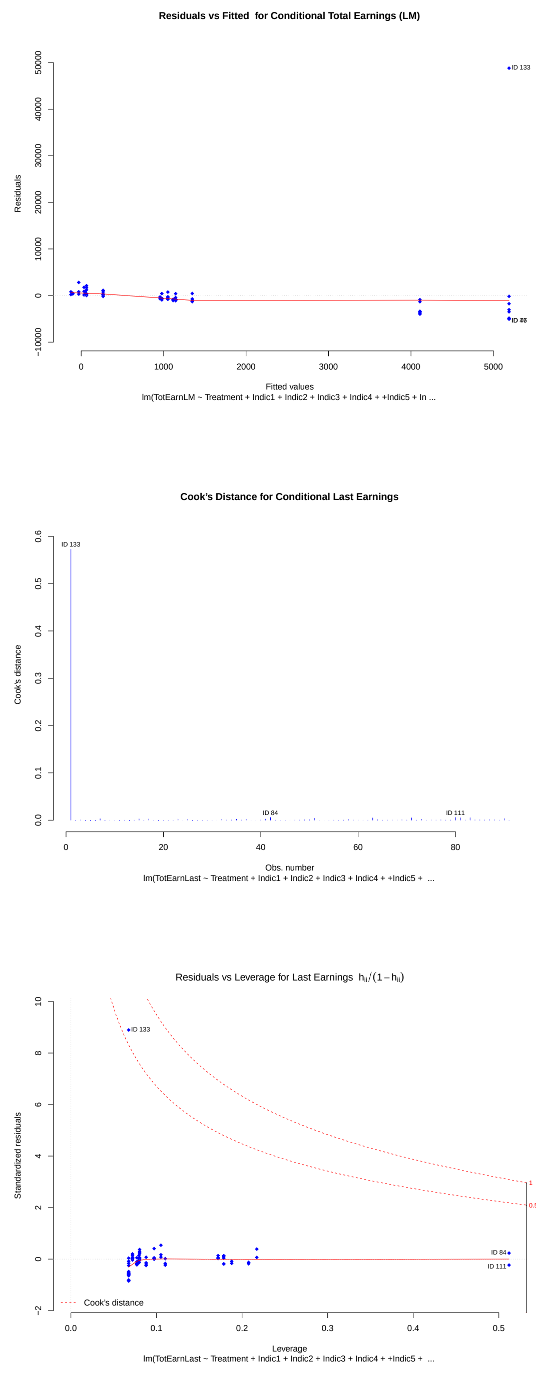
The statistical literature offers a range of methods to detect data points that can be labeled as outliers. Three commonly used methods are the analysis of the residual plot, the use of Cook's Distance and the Leverage Index (Rousseeuw and Leroy (1996)).

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 A.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 A.10 presents inference on the income treatment effects for the full data set and the subsets of non-migrants, males and females. Table A.11 presents the results for schooling outcomes while Table A.12 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 A.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 A.10: 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 A.11: 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 A.12: Treatment Effects on Employment Using Rank-sum Statistic

	Sample		**Control**	**Treatment**	**Effect**	**Rank-sum Statistics**		**Permutation** (one-sided)	
Variables	**#C**	**#T**	**Mean**	**Effects**	**Size**	**Z-value**	**p-val**	**Single p-val**	**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.

7 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 A.13–A.15 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 A.13 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 A.14 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 A.15 focuses on males only. The data has to much variance for such small sizes to produce conclusive results.

Table A.13: 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 A.14: 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 A.15: 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

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

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

Conover, W. J. and D. S. Salsburg (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.

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

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

Heckman, J. J., S. H. Moon, R. Pinto, P. A. Savelyev, and A. Q. Yavitz (2010, July). 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., R. Pinto, and P. A. Savelyev (2013, October). Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *American Economic Review 103*(6), 2052–2086.

Robins, J. M., A. Rotnitzky, and L. P. Zhao (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 M. Wolf (2005). Stepwise multiple testing as formalized data snooping. *Econometrica 73*(4), 1237–1282.

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