Effect of the Jamaica Early Childhood Stimulation Intervention on Labor Market Outcomes at Age 31 *

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

September 30, 2021

*We thank David Contreras, Carene Lindsay, Stacey-Ann Aiken, and Michelle Williams for excellent research assistance, and Azeem Shaik for helpful comments. This paper also has benefited from comments by seminar participants at University of Chicago, University of California at Berkeley, University of California at Los Angeles, and Tsinghua University in China. The authors acknowledge financial support from the Strategic Impact Evaluation Fund at the World Bank. The authors declare that they have no financial or materials interests in the results reported in this paper. Research reported was supported by 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 are solely those of the authors and do not represent the official views of the National Institutes of Health.

Abstract

We report the labor market effects of the Jamaica Early Childhood Stimulation intervention at age 31. The study is a small-sample randomized experiment targeting 127 stunted children between 9 and 24 months old living in poor neighborhoods of Kingston, Jamaica, during 1987-1989. The treatment consisted of a two-year home-based educational intervention focusing on mother-child interactions. It was designed to improve nutrition, cognition, language and psycho-social skills. We tracked 75% of the original sample 30 years after the intervention. We find large and statistically significant effects on labor market outcomes such as income and schooling.

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

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., 2011). This paper reports the effects of an early childhood home visiting program, the 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., 2011).[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 the 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., 2021; 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., 2021 and García et al., 2021). 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 ages 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. 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 kilogram 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, 2011, 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., 2011, 2005). This is in contrast with a study in Guatemala in 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 psychosocial stimulation (groups 2 and 4) can be pooled. Statistical evidence suggests to pool the psychological 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 G 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 our appendix for more details.).

Attrition is well-balanced across treatment groups for the 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 our 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 who 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 online 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 B 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 of 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 B of our 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 E 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 C 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 B of our 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 of our 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

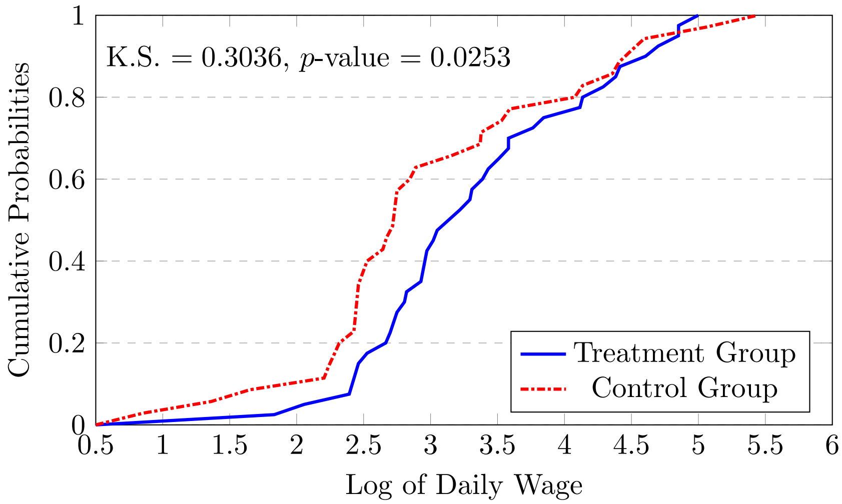
This paper evaluates the long-term economic impacts of the Jamaica Early Childhood Stimulation intervention program, an early childhood intervention for socially and biologically disadvantaged children living in a poor country. The study consists of a randomized control trial that enrolled 127 stunted 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.

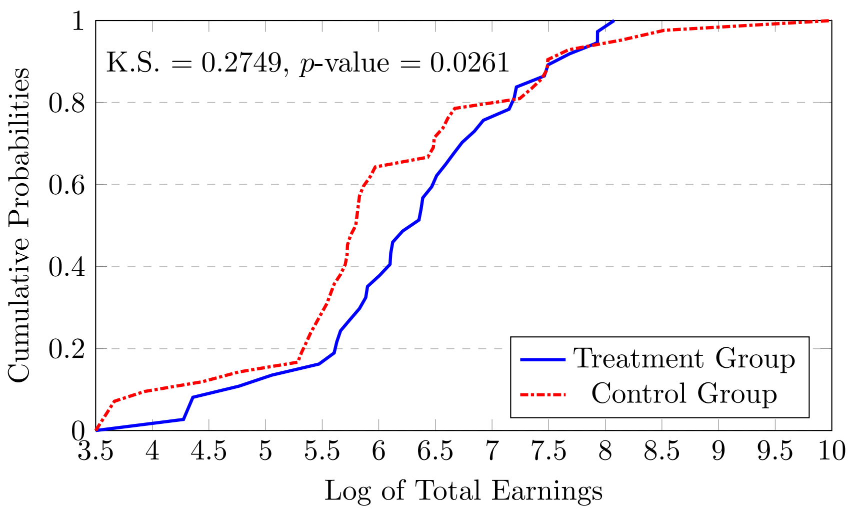
The Jamaican Study is a rare case of a long-term follow up for an early childhood development (ECD) intervention implemented in a less-developed country. Our results confirm large economic returns to an early childhood intervention that targeted disadvantaged families living in poverty in the poor neighborhoods 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 #C	Sample #T	Control Mean	Treatment Effects	Effect Size	Asymptotic (one-sided) t-stat	Asymptotic (one-sided) Single p-val	Permutation (one-sided) Single p-val	Permutation (one-sided) 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 #C	Sample #T	Control Mean	Treatment Effects	Effect Size	Asymptotic (one-sided) t-stat	Asymptotic (one-sided) Single p-val	Permutation (one-sided) Single p-val	Permutation (one-sided) 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).

References

Andrew, A., O. Attanasio, B. Augsburg, M. Day, S. Grantham-McGregor, C. Meghir, F. Mehrin, S. Pahwa, and M. Rubio-Codina (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., O. Attanasio, E. Fitzsimons, S. Grantham-McGregor, C. Meghir, and M. Rubio-Codina (2018, 04). 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., S. Cattan, E. Fitzsimons, C. Meghir, and M. Rubio-Codina (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 L. A. Stefanski (2012). *Essential Statistical Inference: Theory and Methods.* Springer.

Campbell, F. A., G. Conti, J. J. Heckman, S. H. Moon, and R. Pinto (2013, March). 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., 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.

Carneiro, P. and R. Ginja (2014, November). 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 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.

Conti, G., J. J. Heckman, and R. Pinto (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., , J. L. García, J. J. Heckman, and A. Hojman (2016). Early childhood education. In R. A. Moffitt (Ed.), *Economics of Means-Tested Transfer Programs in the United States*, Volume 2, Chapter 4, pp. 235–297. Chicago: University of Chicago Press.

Engle, P. L., L. C. H. Fernald, H. Alderman, J. Behrman, C. O'Gara, A. Yousafzai, M. Cabral de Mello, M. Hidrobo, N. Ulkuer, I. Ertem, and S. Iltus (2011, October). 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., F. Bennhoff, D. Ermini Leaf, and J. J. Heckman (2021). The dynastic benefits of early childhood education. Under Review.

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

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

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

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

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.

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

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

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

Hamadani, J. D., S. F. Mehrin, F. Tofail, M. I. Hasan, S. N. Huda, H. Baker-Henningham, D. Ridout, and S. Grantham-McGregor (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., T. A. Drizd, C. L. Johnson, R. B. Reed, R. A. F., and W. M. Moore (1979). Physical growth: National center for health statistics percentiles. *The American Journal of Clinical Nutrition 32*(3), 607–629.

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

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

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

Heckman, J. J. and T. Kautz (2012, August). 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., B. Liu, M. Lu, and J. Zhou (2021). Treatment effects and the measurement of skills in a prototypical home visiting program. Under Review.

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

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.

Hoddinott, J., J. A. Maluccio, J. R. Behrman, R. Flores, and R. Martorell (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., J. J. Heckman, R. Diris, B. ter Weel, and L. Borghans (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., J. Hoddinott, J. R. Behrman, R. Martorell, A. R. Quisumbing, and A. D. Stein (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.* Stony Brook, NY: State University of New York at Stony Brook.

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

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.

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

Rubio-Codina, M., M. Dormal, and M. C. Araujo (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., H. Baker-Henningham, A. Brentani, R. Mugweni, and S. P. Walker (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., T. Candland, and S. O. Whitney (2015). Later impacts of early childhood interventions: A systematic review. *World Bank Independent Evaluation Group, Working Paper 2015/3.*.

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

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

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

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

Walker, S. P., S. M. Chang, A. S. Wright, R. Pinto, J. J. Heckman, and S. M. Grantham-McGregor (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., C. A. Powell, S. M. Grantham-McGregor, J. H. Himes, and S. M. Chang (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., T. D. Wachs, J. M. Gardner, B. Lozoff, G. A. Wasserman, E. Pollitt, J. A. Carter, and The International Child Development Steering Group (2007, January). Child development: Risk factors for adverse outcomes in developing countries. *Lancet 369*(9556), 145–157.