PREP SCHOOL FOR POOR KIDS:  
THE LONG-RUN IMPACTS OF HEAD START ON HUMAN CAPITAL AND ECONOMIC SELF-SUFFICIENCY

Martha J. Bailey, Shuqiao Sun, and Brenden Timpe

March 27, 2018

[click here for most updated draft]

Abstract
This paper evaluates the long-run effects of Head Start on human capital in large-scale, linked administrative data. Our research design exploits the county-level rollout of Head Start between 1965 and 1980 together with program eligibility captured by state-level school-entry age cutoffs. Using the restricted 2000-2013 Census/ACS linked to the Numident, we find that children induced to participate in Head Start achieved 0.29 more years of education, reflecting a 2.1-percent increase in high school completion, an 8.7-percent increase in college enrollment, and 18.5-percent increase in college completion. Consistent with the program benefitting lower income children, participation in Head Start decreased adult poverty by 12 percent and the receipt of public-program income by 29 percent. Our estimates are smaller in magnitude than those reported in other studies, but nevertheless imply substantial returns to investing in large-scale, publicly funded preschool programs.

JEL Codes: I2, J24, J6

Contact Information
Bailey: Department of Economics, University of Michigan, 611 Tappan Street, Ann Arbor, Michigan 48109; Email: baileymj@umich.edu; Website: www-personal.umich.edu/~baileymj. Sun: Department of Economics, University of Michigan, 611 Tappan Street, Ann Arbor, Michigan 48109; sqsun@umich.edu; Timpe: Department of Economics, University of Michigan, 611 Tappan Street, Ann Arbor, Michigan 48109; btimpe@umich.edu.

Acknowledgements
Data collection for the War on Poverty project was generously supported by the National Institutes of Health (R03-HD066145). Data linkage and analyses for this project were generously supported by the Laura and John Arnold Foundation. The opinions and conclusions expressed herein are solely those of the authors and should not be construed as representing the opinions or policy of any agency of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. We gratefully acknowledge the use of the services and facilities of the Population Studies Center at the University of Michigan (funded by NICHD Center Grant R24 HD041028). We are also grateful for support for the Michigan RDC from the NSF (ITR-0427889). During work on this project, Timpe was partially supported by the NIA (T32AG000221) as a UM Population Studies Center Trainee. We are grateful to Doug Almond, Hilary Hoynes, and Diane Schanzenbach for sharing the Regional Economic Information System (REIS) data for the period of 1959 to 1978 and the data and dofiles to replicate their analysis with the PSID; and Clint Carter for the many hours spent helping us disclose these results. Evan Taylor and Bryan Stuart provided exceptional assistance in translating string names in the SSA’s NUMIDENT file into GNIS codes. Ariel Binder, Dorian Carloni, and Bryan Stuart also contributed substantially to the cleaning the restricted census data.
Convincing evidence on the longer-term impacts of scaled-up pre-k programs on academic outcomes and school progress is sparse, precluding broad conclusions. The evidence that does exist often shows that pre-k-induced improvements in learning are detectable during elementary school, but studies also reveal null or negative longer-term impacts for some programs.

~ Brookings Pre-Kindergarten Task Force of Interdisciplinary Scientists (April 2017)

In 1965, the U.S. began a new experiment in the provision of public pre-school for disadvantaged children. The motivation was simple: “the creation of and assistance to preschool, day care, or nursery centers for 3- to 5-year-olds…will provide an opportunity for a head start by canceling out deficiencies associated with poverty that are instrumental in school failure” (United States Senate Committee on Labor and Public Welfare 1964). This now-famous “prep school for poor kids” was aptly named “Head Start” (Levitan 1969), and its educational, health, and social programming aimed to help millions of disadvantaged children escape poverty.

More than fifty years later, Head Start is one of the most popular of the War on Poverty’s programs, serving around 900,000 children annually at a cost of approximately $7 billion (Haskins and Barnett 2010). Unlike expensive, small-scale “model” programs such as Perry Preschool and Abecederian, Head Start’s architects prioritized wide-spread access, calculating that a massive preschool expansion would maximize its poverty-fighting (and political) benefits. Skepticism about the quality of this large-scale pre-school program coupled with difficulties in evaluation have generated controversy about its short-term benefits for decades (Westinghouse Learning Corporation 1969, Currie 2001, Duncan and Magnuson 2013, Phillips et al. 2017). Convincing evidence regarding the Head Start’s long-term poverty fighting benefits has remained even more elusive, thanks to the lack of program randomization in its early years, small sample sizes of longitudinal surveys, and the difficulty of measuring adults’ access to Head Start decades ago. Whether Head Start achieved its broad goal of increasing the life opportunities for children remains an open question.

This paper uses newly available large-scale administrative data to provide precise estimates of Head Start’s long-term effects on human capital and economic self-sufficiency. Not only do we observe adult outcomes for almost one third of the U.S. population in the long-form 2000 Census and 2001-2013 American Community Surveys (ACS); links to the Social Security Administration’s (SSA) Numident allow
us measure Head Start access with great accuracy using school-entry age cutoffs together with exact date and county of birth. The resulting sample is four orders of magnitude larger than longitudinal surveys with less potential for endogenous measurement error in Head Start access to bias our estimates.

Our research design exploits county-level launch of Head Start programs from 1965 to 1970 at the Office of Economic Opportunity (OEO) (Levine 1970, Bailey 2012, Bailey and Danziger 2013, Bailey and Duquette 2014, Bailey and Goodman-Bacon 2015). This approach mitigates concerns about the correlation of Head Start funding with program performance and OEO funding for other local programs and eliminates well-known problems of measurement error in archival grant data (Barr and Gibbs 2017). An additional strength of our design is that we compare cohorts born in the same county who were age-eligible for the program when it launched (ages 3-5) to cohorts that were age-ineligible (children 6 and older). Similar to a regression kink design (Card et al. 2015), our key identifying assumption is Head Start’s causal effect is the only reason for a change in the relationship between a child’s age at the program’s launch and her outcomes as an adult.

Our results provide new evidence that a large-scale public preschool program increased the human capital and economic self-sufficiency of disadvantaged children. Children participating in Head Start achieved 0.29 more years of education, were 2.1 percent more likely to complete high school, 8.7 percent more likely to enroll in college, and 18.5 percent more likely to complete college. Consistent with the program benefitting children at risk of living in poverty as adults, Head Start decreased adult poverty by 12 percent and reduced the likelihood that affected children received income from public programs as adults by 29 percent. Although we find that Head Start’s effects on human capital are significantly smaller than in previous studies, our estimates show that the program achieved its goal of increasing children’s economic opportunities.

I. THE LAUNCH OF HEAD START IN THE 1960s

In the 1960s, the (then) revolutionary idea that preschool could improve children’s cognitive development suggested an innovative strategy for poverty prevention. Because poor children started school with significantly less educational background, comprehensive preschool could give them a “Head Start,” improving their success in school and addressing a root cause of poverty.
Funded by the OEO, Head Start first began as an 8-week summer program in 1965 and, after a successful first summer, Johnson announced that Head Start would become a full-year program for children ages 3 to 5.\(^1\) The director of the OEO wrote 35,000 letters to public health directors, school superintendents, mayors and social services commissioners to encourage applications. The OEO also made a special effort to generate applications in America’s 300 poorest counties (Ludwig and Miller 2007).

Head Start’s political popularity led to an even faster launch than other War on Poverty programs. By 1966, Head Start had begun in more than 500 counties where over half of the nation’s children under age 6 resided. By 1970, Head Start had begun in another roughly 900 counties, and federal expenditures on the program reached $326 million, or $1.9 billion in 2012 dollars (OEO 1970). This early expansion ensured that by 1970, Head Start existed in roughly half of U.S. counties nationwide and covered 83 percent of U.S. children under age six. Because much of this history is described elsewhere, our quantitative description of the program’s roll-out appears in our Appendix.

Head Start’s holistic programming aimed to develop the health and physical abilities of poor children, including their self-confidence and relational skills, verbal and conceptual skills, and parent involvement with children.\(^2\) But the challenges of quickly starting a new national program meant that implementation often deviated from ideals. Not only did Head Start lack curricular standardization, but programs struggled to find high-quality teachers to achieve the suggested pupil-to-teacher ratio of 15:1. As a practical solution many centers relied on para-professionals, most of whom lacked post-secondary education; thirty percent had not finished high school (Hechinger 1966, Braun and Edwards 1972).\(^3\) In addition, many components of Head Start phased in slowly. For instance, in 1965 the OEO wrote that “the proportion of children receiving treatment for conditions discovered in Head Start medical and dental examinations…was probably under 20 percent. It rose to over 65 percent in 1966, and in 1967 we fully

\(^{1}\) “Pre-school” also included five-year-olds, because public kindergarten was not universal in this period (Cascio 2009).

\(^{2}\) For instance, Levitan (1969) notes that in Head Start’s 1966-7 budget, early childhood education (daily activities and transport) comprised about 70 percent; health services (including immunizations, screenings and medical referrals) and nutrition added another 17-20 percent to its budget; and parent involvement, social services (e.g., helping families cope with crises), and mental health services accounted for the remainder. Spending on health and other services increased over time.

\(^{3}\) Sizable variation in pre-school quality persist today. See an overview of this literature by Currie (2001), Cascio and Schanzenbach (2013), and Duncan and Magnuson (2013).
expect it to have reached over 90 percent” (OEO 1967).

These characteristics of Head Start’s implementation imply a pattern in the expected relationship of adult outcomes with a child’s age at the time of the program’s launch, which we plot in Figure 1. First, we expect a limited effect of Head Start on children over age 5 when Head Start launched. The effect may not be zero, because 10 percent of children in full-year Head Start were 6 or older (Vinovskis 2008) and school-age children could benefit from younger siblings’ participation (Garces et al. 2002). Our preferred specification which uses 6 year olds as the control group may, therefore, understate Head Start’s effects.

Second, if Head Start had a causal effect, we expect to see a change in adult outcomes for children under age 5 when Head Start launched, because these cohorts would have been the first to have been age-eligible and have access. Several factors, however, imply that we should not expect an immediate shift in the level of outcomes (akin to a regression discontinuity, RD) but rather a shift in the slope (akin to a regression kink). The program’s gradual implementation means that capacity grew over time—full-year Head Start served only 20,000 children in 1965, but this rapidly increased to 160,000 in 1966, 215,000 in 1967 and 1968, and 257,700 in 1970 (OEO 1965, OEO 1966, OEO 1967, OEO 1968, OEO 1970).4 Quality also increased over time, reflecting hiring better teachers, curricular advances, and auxiliary services (e.g., health). Finally, changes in cumulative exposure captures the fact that some children would have been eligible for up to three years. A child age 5 when Head Start launched could participate for one year, whereas children age 3 at launch would be age-eligible for three years. This combination of gradual implementation and cumulative exposure implies a non-linear shift in the relationship between age at Head Start launch and adult outcomes. Because children age -1 at program launch would have been age-eligible for a fully implemented program for three years, our discussion of results focuses on these cohorts.

---

4 Enrollment in summer Head Start was much higher, but we expect the summer program to have smaller effects than full-year exposure to a Head Start program. Even at the beginning of the program, few experts on the planning committee believed that an 8-week summer program could produce lasting benefits (Vinovskis 2008 citing Edward Zigler). Moreover, in this period, 30 to 40 percent of children in summer Head Start were aged six and older, whereas only 10 percent of those in full-year programs were older than five. See Table 1 in Thompson (forthcoming).
II. **EVIDENCE REGARDING THE LONG-TERM EFFECTS OF HEAD START**

Previous evaluations of Head Start’s long-term effects provide suggestive evidence of the program’s effects on human capital and economic self-sufficiency. One pioneering approach to examining Head Start’s long-term effects was the use of family fixed effects with longitudinal data. Building on work by Currie and Thomas (1995), Garces et al. (2002) used the Panel Study of Income Dynamics (PSID) to compare children who participated in Head Start to their siblings who did not. They show that Head Start increased high school graduation rates and college enrollment among whites and reduced arrest rates among blacks. Using a similar research design for more recent cohorts in the National Longitudinal Survey of Youth (NLSY), Deming (2009) finds that Head Start participation had large and positive effects on a summary index of adult outcomes (including high school graduation, college attendance, “idleness,” crime, teen parenthood, and health status). Well-known critiques caution that sibling comparisons may suffer from well-known sources of endogeneity bias (Griliches 1979, Bound and Solon 1999). In addition, small sample sizes in longitudinal surveys may provide unreliable estimates of Head Start’s effects (Miller et al. 2017).

More recent work exploits shifts in access to Head Start using three distinct research designs. Ludwig and Miller (2007) path-breaking application of RD exploited the OEO’s special effort to generate grant proposals from the 300 poorest counties. Comparing the outcomes of children on either side of this threshold, they find evidence that Head Start reduced childhood mortality and increased the receipt of high-school degrees and college enrollment. However, because the 1990 and 2000 Censuses required them to use county of residence in adulthood to proxy for childhood Head Start access, measurement error causes their education results to be sensitive to specification and often statistically insignificant. Carneiro and Ginja (2014) use an RD in state-, year-, and household-based income eligibility cutoffs for more recent Head Start programs. They find that Head Start decreased behavioral problems, the prevalence of some health conditions (including obesity) between the ages of 12 and 17, and crime rates around age 20. They

---

5 See also excellent reviews of studies of Head Start’s short-term effects (Currie 2001, Cascio and Schanzenbach 2013, Duncan and Magnuson 2013, Gibbs et al. 2014).

6 There is also limited evidence showing the poorest 300 counties got were more likely to get grant-funding for Head Start (see Ludwig and Miller 2007: table II and Pihl 2017).
find a positive though statistically insignificant effect on receiving a high-school diploma as well as suggestive evidence that Head Start reduced college enrollment.

In work closely related to this paper, three studies make use of county-year variation in Head Start funding in the 1960s and 1970s to quantify the program’s long-term effects. Using the NLSY, Thompson (forthcoming) finds that greater funding for Head Start at ages 3 to 6 raised college graduation rates, reduced the incidence of health limitations, and tended to raised adult household income. Focusing on a “high impact” sample, Johnson and Jackson (2017) find that an average level of Head Start spending increases the likelihood that children graduated from high school by 8 percentage points and gained 0.40 in years of schooling. These children also experienced an 8 log-point increase in adult wages, a 14.4 log-point increase in adult family income at ages 20 to 50, a 3.5 percentage-point reduction in poverty at ages 20 to 50, and a 2.8 percentage-point reduction in adult incarceration. Finally, Barr and Gibbs (2017) examine the intergenerational effects of Head Start using the NLSY and two research designs: family fixed effects and variation in program availability across birth counties. Pointing out the measurement error in Head Start grant funding amounts in the National Archives data, the latter approach constructs Head Start access (equal to one if the level of Head Start funding exceeds the 10th percentile of observed funding per four year old). Although Thompson (forthcoming) finds little evidence of first-generation effect in the NLSY using a similar access measure, Barr and Gibbs (2017) find evidence of first-generation effects on women as well as large second-generation effects on their daughters’ high school graduation and completed education.

III. DATA AND RESEARCH DESIGN

This study combines the long-form 2000 Census and 2001-2013 ACS with the Numident to shed new light on Head Start’s long-term effects. The Census/ACS data represent almost one third of the U.S. population and are four orders of magnitude larger than previously used longitudinal samples. Another

7 Assuming these effects are driven by enrollment in Head Start and that 40 percent of poor children attended full-year Head Start (their calculation on p. 23) implies a treatment effect on treated children of 0.8 of a year of education, 16 log-point increase in adult wages (of men and women), a 29 percentage point increase in adult family income, 7 percentage point reduction in poverty, and a 5.6 percentage point reduction in adult incarceration. These effects increase by 30 percent if they were also driven by full-year and summer Head Start (p. 23).
advantage of these data is that the Numident contains county of birth (rather than adulthood residence) and exact date of birth which allows a high-quality proxy for Head Start access and age eligibility in childhood. The data’s main disadvantage is that they contain no information on family background, which limits subgroup analyses of the lower-income children heavily represented in the program.8 Our sample is comprised of children born from 1950 to 1980 in U.S. states where the school-entry age cutoff is known. We additionally limit our sample to individuals who are in their prime-earning years from 2000 and 2013 (ages 25 to 54). To minimize disclosure concerns at the Census Bureau, we use only observations with non-allocated and non-missing values for all outcomes.

Combining the launch of Head Start programs from Bailey and Goodman-Bacon (2015) with the Census/ACS-Numident permits two refinements to previously used research designs (Barr and Gibbs 2017, Johnson and Jackson 2017, Thompson forthcoming). First, we use only the launch of the Head Start program rather than Head Start spending. This refinement addresses the potential (1) endogeneity of Head Start funding decisions to the program’s performance and (2) the correlation of year-to-year variation in Head Start funding with that of other OEO programs. It also sidesteps issues of measurement error in the National Archives grant amounts (Barr and Gibbs 2017).9 Second, we examine changes in the relationship of interest for children who were age-eligible for Head Start when it launched (ages 3-5) relative to those who were age-ineligible (ages 6+). Age eligibility is based on exact date of birth in the Numident and school-entry age cutoffs, which alleviates measurement error in defining the potential treatment and control groups. Similar to a regression-kink design, our identifying assumption is that the causal effect of Head Start is the only reason for a change in the relationship between a child’s age at the program’s launch and her outcomes as an adult.

We implement these comparisons within the following event-study framework,

\[
Y_{bct} = \theta_c + \alpha_t + \delta_{s(c)b} + Z'_{cb}\beta + \text{HeadStart}_{c} Agg_{bs(c)}\varphi + \epsilon_{bct}.
\]

Children’s birth years are indexed by \( b = 1950 - 1980 \), county of birth by \( c \), and Census/ACS year by

---

8 In 1970, 62 percent of Head Start’s participants were from families with annual incomes less than the poverty line for a family of four (~$4,000) (OEO 1970).
9 Thompson (forthcoming) also tries this strategy but notes that his estimates in the NLSY are statistically insignificant.
Specifications include fixed effects for county of birth, \( \theta_c \), year, \( \alpha_t \), and state-by-birth-year, \( \delta_{s(c)b} \), which, respectively, capture time-invariant differences across counties, national changes affecting all individuals, and changes in state policies that differentially affect birth cohorts. Although covariates matter little, we follow the literature and include county characteristics interacted with a linear trend, \( Z_{cb} \) (Hoynes et al. 2011, Bailey 2012, Bailey and Goodman-Bacon 2015), but these covariates matter little.\(^{10}\) Standard errors are corrected for heteroskedasticity and adjusted for an arbitrary within-county covariance structure (Arellano 1987, Bertrand et al. 2004). See the Appendix for more detailed variable descriptions and robustness checks.

Our outcomes of interest are summary measures of human capital and economic self-sufficiency, which permits tests of co-movements of related adult outcomes and limits the number of statistical tests (Kling et al. 2007).\(^{11}\) A shortcoming of this approach is that, because indices weight each component equally, large changes in one dimension are averaged with potentially opposite-signed or zero effects in other dimensions. We, therefore, also examine the individual index components. The human capital index includes four binary variables indicating achievement of a given level of education or greater: high school or GED, some college, a 4-year college degree, and a professional or doctoral degree; years of schooling, and an indicator for working in a professional occupation. Our index of self-sufficiency includes binary indicators of employment, poverty status, and income from public sources, weeks worked, usual hours worked, and the log of labor income.

Our point estimates of interest, \( \phi \), capture the intent-to-treat (ITT) effects of Head Start on long-term human capital and labor-force productivity outcomes. The set of these effects describe the evolution of outcomes for children born in counties, \( c \), that received Head Start programs (\( HeadStart = 1 \)) with dummy variables capturing the children’s age relative to the state school-entry cutoff at the time the program began, \( T_c^* \). We treat school age (age 6 before the school entry cut-off date) as the omitted category, because these

\(^{10}\) County characteristics include the 1960 poverty rate, log county population, population share over age 65, under age 5, living in an urban setting, and non-white.

\(^{11}\) Following Kling et al. (2007), we standardize outcome measures for each individual using the mean and standard deviation for in the control group (ages 6-7 at launch), and we recode outcomes so that increases indicate improvements in human capital.
children would have been unlikely to have attended Head Start rather than public school. Although the event study estimates range from $-15$ (1965–1980) to 30 (1980–1950), we focus on years $-1$ to $+15$, which are compositionally balanced across counties.

Based on Figure 1’s predictions, we also summarize the event-study estimates using a three-part spline with knots at age 6 and -1. The spline has the added benefit of allowing us to conduct formal trend-break tests at age 6 to characterize whether the large increase in age-eligibility for Head Start at age 5 is associated with a change in adult outcomes.

These ITT effects average over all children in a county, regardless of whether they participated in Head Start. To compare our estimates to the literature, we generate average treatment-effects-on-treated children (ATET) by dividing by 0.151, our best estimate of the effect of Head Start launch on program enrollment using both administrative and 1970 Census data. (See Appendix for more details.) We construct confidence intervals using a parametric bootstrap procedure using 10,000 independent draws from normal distributions with means and standard deviations equal to the point estimates and standard errors from the reduced-form and first-stage estimates (Efron and Tibshirani 1993). Because we cannot resample from data used in other Head Start papers, we also use this method to recalculate confidence intervals for other estimates in the literature as needed.

IV. HEAD START’S EFFECTS ON HUMAN CAPITAL

Table 1 summarizes this evidence for all outcomes in the human capital index for the fully-exposed cohort (age -1 at Head Start’s launch). Column 2 shows the event-study estimate, while column 4 contains the spline estimate. The results suggest that Head Start dramatically improved adult human capital. The standardized index is about one-thirteenth of a standard deviation (column 2 / standard deviation in column 1) higher for the cohort overall and half a standard deviation for treated children (column 6 / column 1). Although the spline trades flexibility for parsimony, it fits the data well: column 4’s estimates are identical to column 2 to the hundredth. Column 5 presents the test-statistic for the equality of spline segments for ages 6 and older and -1 to 5. For the human capital index, the data reject the null of no trend-break at the 1-percent level. Figure 2 summarizes the event-study results for a selection of outcomes driving changes in the index. The x-axis is the child’s age at the time Head Start launched, and the dashed lines plot the spline
segments. Consistent with a causal interpretation, each index component shows a visual trend break at age 6 with increasing ITT estimates as anticipated in Figure 1.

Panels A and B of Figure 2 summarize event-study estimates for two of the most commonly studied human-capital outcomes in the literature: high school graduation and college enrollment. Table 1 shows that treated children were 2.1 percent more likely to complete high school/GED. This is much smaller than effects reported in related studies of Head Start’s effects in the 1960s. Figure 3A shows that it is roughly-half the size of Garces et al. (2002)’s sibling comparison in the PSID and Thompson (forthcoming)’s spending design using the NLSY. In addition, it is one-fifth the size of Johnson and Jackson (2017)’s estimates for a very disadvantaged sample in the PSID; and one-ninth the size of Ludwig and Miller (2007)’s RD estimates using the Census. (It is one quarter the size of Deming et al. (2009)’s sibling comparison for Head Start in the 1990s.) Although our estimate falls within the confidence intervals of previous studies, this reflects the large uncertainty associated with those estimates.

Similarly, Table 1 shows a statistically significant effect of Head Start on college enrollment. Head Start raised college enrollment by 1.3 percent, which translates into 8.7 percent among treated children. Figure 3B shows that our estimate is one half the size of Garces et al. (2002) and one quarter the size of Ludwig and Miller (2007). (The magnitude of the increase in college enrollment of 0.05 is only slightly smaller than Deming (2009)’s NLSY sibling comparison for Head Start in the 1990s.) Consistent with the visually evident trend-break in Figure 2B, we reject a trend-break at age 6 at 1-percent level.

In addition to generating more precise estimates for these commonly studied outcomes, our data permit a novel exploration of Head Start on other dimensions of human capital, including college completion or higher degrees. Table 1 shows that children with access to with Head Start achieved 2.8 percent higher college graduation rates—implying an 18.5 percent increase among treated children. Consistent with a causal interpretation, column 5 shows evidence of a trend break at the 1-percent level. These effects are one-quarter to one-fifth the size of the model Abecedarian Project (Currie 2001, Barnett

---

12 The similarity of column 2 for some college and college completion is correct. This similarity does not imply that every person that Head Start induced to enter college completed it. Rather, Head Start likely helped some individuals enter and others complete college.
and Masse 2007, Duncan and Magnuson 2013). Similarly, completion of professional or graduate degrees increases by 50 percent among treated children. Overall, these gains at virtually every level of educational attainment are captured by an increase of 0.29 years of schooling for treated children. This estimate is much lower than Johnson and Jackson (2017)’s estimate of 0.52 years for very disadvantaged children, but it is highly statistically significant and does not appear driven by a pre-trend.\(^{13}\)

These results may be surprising, given that no other study of preschool has documented effects on post-secondary education. Note, however, that this lack of evidence may reflect the small longitudinal samples or the small scale of model pre-school programs. Differences in the participating children may also matter. Abecedarian and Perry’s participants were all very disadvantaged children and mostly black, and Perry’s participants had low IQs.\(^{14}\) In contrast, Head Start participants were less economically disadvantaged and roughly half were African American in 1970. It is, therefore, likely that many Head Start participants would have faced fewer socio-economic disadvantages and less racism, which makes is less surprising that Head Start may have influenced post-secondary educational outcomes.

Because analyses of model preschool programs have found larger educational effects for boys, Table 1 also presents effects separately by sex. Among participating men, the human capital index increased by a statistically significant over half of a standard deviation. For this group, high school completion rose by 2.8 percent, college attendance rose by 12.5 percent, and college completion rose by 26.5 percent. The high school estimates are, again, smaller than others in the literature: Deming (2009) estimates an effect of 11.4 percentage points, while Johnson and Jackson’s estimates imply an ATET of 10.6 percentage points for their very disadvantaged sample. However, the college attendance estimates are much larger than Deming’s estimate of a statistically insignificant increase of 0.02 percentage points. Together, these shifts raised years of education among treated men by half of a year and the likelihood of completing a professional or doctoral degree by 58 percent, or 2 percentage points. The evidence suggests that men

\(^{13}\) Jackson and Johnson’s ITT estimate is 0.0967 per $1000 spent per poor 4-year-old. They translate this into an ATET by multiplying the coefficient by $4000 (the average Head Start spending per poor 4-year-old), so that 0.0967*4/0.75=0.52

\(^{14}\) The model Perry Preschool Program, which focused on lower IQ children, had no measured effects on postsecondary outcomes (Anderson 2008).
treated with Head Start were 19 percent more likely to hold professional jobs, although the trend-break is weakly significant.

Similar to results in model pre-school programs, the human capital index increased by half as much for women: only 28 percent of a standard deviation among treated women. In addition, the relationships of other index components with Head Start exposure were smaller and sometimes statistically insignificant. Completion of high school (or a GED) rose by 1.5 percent, and college attendance rose by 5.5 percent, although the trend-break at age 6 is not statistically significant at conventional levels. Changes in the human capital index appear more directly affected by changes in college completion and professional degrees, which rose by 10.7 percent and 33 percent, respectively. Overall, treated women’s years of education rose by 0.16 years and their likelihood of holding a professional job rose by 9 percent.

V. HEAD START’S EFFECTS ON LABOR-MARKET OUTCOMES

The substantial effects of Head Start on human capital naturally raise the question of whether the program improved labor-market outcomes. When examining the economic self-sufficiency aggregated over both sexes, we find that Head Start improved treated adults’ self-sufficiency by one-sixth of a standard deviation—an estimate that is not statistically different from zero (we omit these results for brevity). This null result likely reflects the fact that men and women’s work effort (and likely selection) changed in offsetting ways. Head Start’s effects on men’s human capital may have led them to increase employment (if the substitution effect of higher potential wages dominates), but the reverse may be true for women (if greater education enables marriage to higher-earning men). Head Start could lead women to lower labor-force participation by raising the income of husbands.

To investigate these effects, Table 2 examines the components of the self-sufficiency index by sex. Stratifying by sex significantly reduces sample sizes, so we focus our analysis on the spline estimates. For treated men, the self-sufficiency index increased by over one-seventh of a standard deviation. Notably, we find positive effects of Head Start exposure on both the extensive and intensive margins of men’s labor-force participation. Column 6 shows that treated men were 2.1 percent more likely to have worked for pay, and they worked an average of 5 more days in the previous year (1 week) and 1 additional hour per week. Consistent with these estimates reflecting the causal effect of Head Start, we find evidence of a marginally
significant trend-break at age 6. Interestingly, the combined effects of increased human capital and labor-force participation do not appear to have affected annual wages—perhaps reflecting the role of negative selection into work. Although Head Start reduced the receipt of income from public sources by 34 percent, it appears to have little effect on men’s poverty.

As expected, the pattern is different for women. Although the self-sufficiency index for treated women increased by around one-seventh of a standard deviation, their labor-force participation was slightly lower. Column 6 shows that treated women were 1.5 percent less likely to work for pay. Women also worked fewer weeks and hours, although these estimates are not statistically different from zero. These reductions in work appear to have increased annual wages of working women by around 4 percent, which is consistent with Head Start engendering positive selection as less-skilled women opt out. Finally, Table 2 suggests that that Head Start reduced treated women’s receipt of income from public sources by 30 percent and poverty rates by 16 percent.

Consistent with the intended effects of the program, the effects of Head Start appear more clearly at the lower end of the skill distribution. Aggregating over both men and women, exposure to Head Start decreased the likelihood of adult poverty by 12 percent (1.2 percentage points) among treated individuals. In addition, treated children were 29 percent (3.2 percentage points) less likely to receive income from public programs as adults.

VI. REASSESSING THE LONG-TERM ECONOMICS OF HEAD START ON HUMAN CAPITAL

Over the past 20 years, substantial evidence has accumulated that model preschool programs may have sizable economic returns (Cunha and Heckman 2007, Heckman et al. 2010, Almond and Currie 2011, Duncan and Magnuson 2013). However, the convincing evidence on the long-run returns to larger-scale, public preschool has remained sparse (Phillips et al. 2017).

This paper provides novel estimates of the long-term effects of Head Start, the nation’s longest running, large-scale preschool program on a variety of human capital and self-sufficiency outcomes. Children exposed to Head Start achieved 0.29 more years of schooling, reflecting the fact that they were 2.1 percent more likely to complete high school, 8.7 percent more likely to enroll in college, and 18.5 percent more likely to complete college. Although effects on high school graduation are smaller than in
previous work, these estimates also reveal the cost-effectiveness of Head Start as a higher-education intervention. Because Head Start in its early years cost roughly $4,400 per child, the implied cost per college enrollee is $80,686 ($=4,400/0.054)—which compares favorably to other programs intending to increase college attendance and graduation (Dynarski et al. 2013). 15

A second finding is that Head Start improved the adult self-sufficiency of disadvantaged children. Even the imperfect implementation of the program in the 1960s decreased participants’ public assistance receipt by 29 percent and adult poverty by 12 percent. The latter result provides an optimistic assessment of the value of Head Start in decreasing poverty. Using the supplemental poverty measure for 2012, Temporary Assistance for Needy Families cost $10.24 billion in 2013 dollars and reduced poverty rates of non-elderly adults by 0.65 percent (0.10/15.5) (Short 2013). Ignoring offsetting behavioral changes and deadweight loss, this implies that TANF reduced poverty by 1 percent for every $15.8 billion spent annually. By comparison, Head Start reduced poverty by 1 percent per year for every $166 million spent ($2 billion /12) spent over the 1965 to 1985 period.

Several reasons suggest that these estimates are conservative. First, our research design differences out any effects on older siblings. Second, our estimates of Head Start’s launch on participation may be too large, implying that the ATET effects are larger than we show. Third, reports of income and public assistance receipt may be severely underreported in major national surveys by half (Bound et al. 2001, Meyer et al. 2015), our estimates of the effects of Head Start in our Census/ACS sample may be understated. Fourth, other pre-school studies suggest that Head Start may have affected outcomes that we cannot quantify with the Census/ACS, such as incarceration, health, and longevity. Finally, Head Start today is arguably a higher quality program than when it launched, with costs almost twice as high at $8,000 per child. For all of these reasons, the returns to today’s Head Start program may be larger than in the 1960s. Future research should continue to document the effects of Head Start on other outcomes and also conduct cost-benefit analyses to quantify the potentially large social returns to investments in Head Start.

15 All figures in this section are given in 2013 dollars. Head Start enrollment and funding are 1965-1985 averages from https://eclkc.ohs.acf.hhs.gov/sites/default/files/pdf/head-start-federal-funding-funded-enrollment-history-eng.pdf
VII. REFERENCES


Figure 1. Expected Effects of Head Start on Adult Outcomes by Cohort’s Age at Launch

- Cohorts age -1 at Head Start's launched were exposed to a fully implemented for 3 years.
- Cohorts too old to be directly affected by Head Start (this ignores sibling spillovers).

Age at Head Start Introduction

- Quality of Head Start implementation (0-100%)
- Cohort's cumulative years of potential access to Head Start
- Potential effects of Head Start (quality x cumulative access)
Figure 2. The Effect of Head Start on Adult Human Capital and Self-Sufficiency

A. High School Graduation or GED

B. Some College or More

C. College Completion or More

D. Prof. or Doc. Degree

E. Not in Poverty

F. No Income from Public Sources
**Figure 3. The Magnitude of Head Start’s Effects on Education across Studies**

**A. Effects of Head Start on High School Graduation**

![Graph A](image)

**B. The Effects of Head Start on College Enrollment**

![Graph B](image)

Notes: Circles indicate show implied or reported ATET from different studies. For sibling fixed effect studies, the ATET is directly reported in the papers. For other studies, we scale the reported ITT by the reported first-stage estimate and constructed 95-percent confidence intervals using the standard error of the first stage as described in the text. Because Johnson and Jackson (2017) does not report a standard error on the first stage, the confidence interval reported in Panel A does not include that uncertainty. We limited the y-axis range so that the confidence intervals for most studies could be read from the figure. The confidence intervals for Ludwig and Miller (2007) fall outside the y-axis range and are [-0.54, 1.47] in panel A and [-0.67, 1.82] in panel B. Bars indicate the reported 95-percent confidence interval for sibling fixed-effect models or constructed for the ITT studies as described in the text. See Appendix for more details on the exact figures used. *Johnson and Jackson (2017) sample includes PSID individuals born to parents in the bottom quartile of the income distribution.
### Table 1. The Effect of Head Start on Adult Human Capital

<table>
<thead>
<tr>
<th>Dependent Variables</th>
<th>(1) Control Mean (St. Dev.)</th>
<th>(2) Event Study at -1 (s.e.)</th>
<th>% Gain</th>
<th>Spline break at 6</th>
<th>F-stat on p-value</th>
<th>ATET Census 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A. Men and Women</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human capital index</td>
<td>0.014 (0.198)</td>
<td>0.015 (0.003)</td>
<td>7.6c</td>
<td>0.015</td>
<td>14.02</td>
<td>0.100</td>
</tr>
<tr>
<td>Completed high school/GED</td>
<td>0.92 (0.078)</td>
<td>0.003 (0.001)</td>
<td>0.3</td>
<td>0.002</td>
<td>2.18</td>
<td>0.19</td>
</tr>
<tr>
<td>Attended some college</td>
<td>0.62 (0.140)</td>
<td>0.008 (0.002)</td>
<td>1.3</td>
<td>0.008</td>
<td>4.84</td>
<td>0.54</td>
</tr>
<tr>
<td>Completed 4 year college</td>
<td>0.29 (0.122)</td>
<td>0.008 (0.002)</td>
<td>2.8</td>
<td>0.009</td>
<td>11.66</td>
<td>0.54</td>
</tr>
<tr>
<td>Prof. or doc. degree</td>
<td>0.028 (0.037)</td>
<td>0.002 (0.001)</td>
<td>7.5</td>
<td>0.002</td>
<td>7.36</td>
<td>0.14</td>
</tr>
<tr>
<td>Years of schooling</td>
<td>13.57 (0.695)</td>
<td>0.043 (0.011)</td>
<td>0.3</td>
<td>0.049</td>
<td>13.06</td>
<td>0.287</td>
</tr>
<tr>
<td>Has a professional job</td>
<td>0.35 (0.121)</td>
<td>0.007 (0.002)</td>
<td>2.1</td>
<td>0.007</td>
<td>4.97</td>
<td>0.048</td>
</tr>
<tr>
<td><strong>B. Men</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human capital index</td>
<td>0.014 (0.247)</td>
<td>0.0206 (0.004)</td>
<td>8.3c</td>
<td>0.020</td>
<td>11.79</td>
<td>0.136</td>
</tr>
<tr>
<td>Completed high school/GED</td>
<td>0.91 (0.104)</td>
<td>0.0038 (0.002)</td>
<td>0.4</td>
<td>0.003</td>
<td>1.92</td>
<td>0.25</td>
</tr>
<tr>
<td>Attended some college</td>
<td>0.59 (0.176)</td>
<td>0.0111 (0.003)</td>
<td>1.9</td>
<td>0.010</td>
<td>5.67</td>
<td>0.074</td>
</tr>
<tr>
<td>Completed 4 year college</td>
<td>0.29 (0.150)</td>
<td>0.0116 (0.003)</td>
<td>4.0</td>
<td>0.011</td>
<td>9.99</td>
<td>0.077</td>
</tr>
<tr>
<td>Prof. or doc. degree</td>
<td>0.034 (0.052)</td>
<td>0.003 (0.001)</td>
<td>8.8</td>
<td>0.003</td>
<td>4.73</td>
<td>0.020</td>
</tr>
<tr>
<td>Years of schooling</td>
<td>13.52 (0.795)</td>
<td>0.0623 (0.013)</td>
<td>0.5</td>
<td>0.063</td>
<td>12.15</td>
<td>0.413</td>
</tr>
<tr>
<td>Has a professional job</td>
<td>0.34 (0.153)</td>
<td>0.0098 (0.003)</td>
<td>2.9</td>
<td>0.009</td>
<td>2.54</td>
<td>0.065</td>
</tr>
<tr>
<td><strong>C. Women</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human capital index</td>
<td>0.015 (0.228)</td>
<td>0.0096 (0.004)</td>
<td>4.2c</td>
<td>0.012</td>
<td>6.50</td>
<td>0.064</td>
</tr>
<tr>
<td>Completed high school/GED</td>
<td>0.93 (0.087)</td>
<td>0.0021 (0.001)</td>
<td>0.2</td>
<td>0.002</td>
<td>1.42</td>
<td>0.014</td>
</tr>
<tr>
<td>Attended some college</td>
<td>0.65 (0.163)</td>
<td>0.0054 (0.002)</td>
<td>0.8</td>
<td>0.006</td>
<td>1.20</td>
<td>0.036</td>
</tr>
<tr>
<td>Completed 4 year college</td>
<td>0.29 (0.147)</td>
<td>0.0047 (0.003)</td>
<td>1.6</td>
<td>0.008</td>
<td>4.63</td>
<td>0.031</td>
</tr>
<tr>
<td>Prof. or doc. degree</td>
<td>0.022 (0.042)</td>
<td>0.0011 (0.001)</td>
<td>5.0</td>
<td>0.001</td>
<td>3.78</td>
<td>0.007</td>
</tr>
<tr>
<td>Years of schooling</td>
<td>13.62 (0.795)</td>
<td>0.024 (0.013)</td>
<td>0.2</td>
<td>0.038</td>
<td>4.91</td>
<td>0.159</td>
</tr>
<tr>
<td>Has a professional job</td>
<td>0.37 (0.150)</td>
<td>0.0051 (0.003)</td>
<td>1.4</td>
<td>0.005</td>
<td>4.35</td>
<td>0.034</td>
</tr>
</tbody>
</table>
Notes: In column 1, the control mean and standard deviation are calculated using the cohorts ages 6 and 7 at the time Head Start was launched. Column 2 presents the estimated intention-to-treat (ITT) effect evaluated at birth cohort of full exposure (-1, see figure 4). Column 3 computes the percentage increase implied by the ITT estimate relative to the control mean (the ratio of column 2 to column 1) for all components of the index. For the index itself, column 3 computes the change in the ITT estimate from column 2 divided by the standard deviation of the control mean in column 1. Column 4 presents the ITT spline estimate evaluated at -1. Column 5 presents the F-statistic and p-value for the test of a trend-break in the spline at age 6. The ATET estimate in column 6 divides the ITT effect at -1 by the estimate of receiving a Head Start grant on school enrollment at school age 5 (0.151 for the full sample; see Appendix Table A2).
### Table 2. The Effect of Head Start on Adult Labor Market Outcomes

<table>
<thead>
<tr>
<th>Dependent Variables</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Control Mean</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (St. Dev.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log family income/poverty</td>
<td>0.035</td>
<td>0.005</td>
<td>2.2+</td>
<td>0.005</td>
<td>1.28</td>
<td>0.030</td>
</tr>
<tr>
<td>(0.208)</td>
<td>(0.003)</td>
<td></td>
<td>[0.26]</td>
<td>[-0.009,0.078]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worked last year</td>
<td>0.91</td>
<td>0.003</td>
<td>0.3</td>
<td>0.003</td>
<td>2.33</td>
<td>0.019</td>
</tr>
<tr>
<td>(0.108)</td>
<td>(0.001)</td>
<td></td>
<td>[0.13]</td>
<td>[0.006,0.038]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weeks worked last year</td>
<td>44.79</td>
<td>0.153</td>
<td>0.3</td>
<td>0.168</td>
<td>2.46</td>
<td>1.013</td>
</tr>
<tr>
<td>(6.03)</td>
<td>(0.078)</td>
<td></td>
<td>[0.12]</td>
<td>[0.002,2.319]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usual weekly hours</td>
<td>41.31</td>
<td>0.148</td>
<td>0.4</td>
<td>0.172</td>
<td>2.94</td>
<td>0.980</td>
</tr>
<tr>
<td>(6.04)</td>
<td>(0.088)</td>
<td></td>
<td>[0.09]</td>
<td>[-0.156,2.419]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log labor income</td>
<td>10.88</td>
<td>0.003</td>
<td>0.0</td>
<td>0.002</td>
<td>0.33</td>
<td>0.022</td>
</tr>
<tr>
<td>(0.301)</td>
<td>(0.004)</td>
<td></td>
<td>[0.57]</td>
<td>[-0.031,0.082]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log family income/poverty</td>
<td>5.91</td>
<td>0.004</td>
<td>0.1</td>
<td>0.002</td>
<td>0.07</td>
<td>0.023</td>
</tr>
<tr>
<td>(0.302)</td>
<td>(0.005)</td>
<td></td>
<td>[0.79]</td>
<td>[-0.043,0.098]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In poverty*</td>
<td>0.070</td>
<td>-0.000</td>
<td>-0.0</td>
<td>-0.001</td>
<td>0.01</td>
<td>-0.002</td>
</tr>
<tr>
<td>(0.095)</td>
<td>(0.001)</td>
<td></td>
<td>[0.90]</td>
<td>[-0.016,0.012]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recvd. public program income*</td>
<td>0.11</td>
<td>-0.005</td>
<td>-0.5</td>
<td>-0.005</td>
<td>9.93</td>
<td>-0.031</td>
</tr>
<tr>
<td>(0.110)</td>
<td>(0.002)</td>
<td></td>
<td>[0.00]</td>
<td>[-0.065,-0.004]</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>B. Women</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-sufficiency index</td>
<td>0.020</td>
<td>0.004</td>
<td>2.1+</td>
<td>0.003</td>
<td>1.50</td>
<td>0.026</td>
</tr>
<tr>
<td>(0.189)</td>
<td>(0.003)</td>
<td></td>
<td>[0.22]</td>
<td>[-0.013,0.073]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Worked last year</td>
<td>0.80</td>
<td>-0.002</td>
<td>-0.2</td>
<td>-0.002</td>
<td>0.15</td>
<td>-0.012</td>
</tr>
<tr>
<td>(0.129)</td>
<td>(0.002)</td>
<td></td>
<td>[0.70]</td>
<td>[-0.042,0.015]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weeks worked last year</td>
<td>37.66</td>
<td>-0.097</td>
<td>-0.3</td>
<td>-0.093</td>
<td>0.03</td>
<td>-0.642</td>
</tr>
<tr>
<td>(6.80)</td>
<td>(0.109)</td>
<td></td>
<td>[0.87]</td>
<td>[-2.296,0.824]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Usual weekly hours</td>
<td>30.49</td>
<td>-0.108</td>
<td>-0.4</td>
<td>-0.092</td>
<td>0.00</td>
<td>-0.715</td>
</tr>
<tr>
<td>(5.81)</td>
<td>(0.097)</td>
<td></td>
<td>[0.95]</td>
<td>[-2.217,0.584]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log labor income</td>
<td>10.31</td>
<td>0.006</td>
<td>0.1</td>
<td>0.007</td>
<td>3.94</td>
<td>0.041</td>
</tr>
<tr>
<td>(0.339)</td>
<td>(0.005)</td>
<td></td>
<td>[0.05]</td>
<td>[-0.024,0.119]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log family income/poverty</td>
<td>5.81</td>
<td>0.008</td>
<td>0.1</td>
<td>0.008</td>
<td>0.68</td>
<td>0.055</td>
</tr>
<tr>
<td>(3.17)</td>
<td>(0.005)</td>
<td></td>
<td>[0.41]</td>
<td>[-0.010,0.137]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>In poverty*</td>
<td>0.12</td>
<td>-0.003</td>
<td>-0.3</td>
<td>-0.003</td>
<td>4.21</td>
<td>-0.019</td>
</tr>
<tr>
<td>(0.108)</td>
<td>(0.002)</td>
<td></td>
<td>[0.04]</td>
<td>[-0.050,0.007]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recvd. public program income*</td>
<td>0.12</td>
<td>-0.005</td>
<td>-0.6</td>
<td>-0.003</td>
<td>0.67</td>
<td>-0.033</td>
</tr>
<tr>
<td>(0.115)</td>
<td>(0.001)</td>
<td></td>
<td>[0.41]</td>
<td>[-0.055,-0.018]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: *For the economic self-sufficiency index, column 3 computes the change in the ITT estimate in column 2 relative to the standard deviation of the control mean in column 1. *In poverty and received public program income are reverse coded when used in the self-sufficiency index. See also Table 1 notes.

[Click here for Online Appendices]