HCEO WORKING PAPER SERIES

Prep School for Poor Kids: The Long-Run Impacts of Head Start on Human Capital and Economic Self-Sufficiency

Martha J. Bailey
Shuqiao Sun
Brenden Timpe

Working Paper

2021-031

06/2021

HUMAN CAPITAL AND
ECONOMIC OPPORTUNITY
GLOBAL WORKING GROUP

The University of Chicago
1126 E. 59th Street Box 107
Chicago IL 60637

www.hceconomics.org

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

April 26, 2021

Abstract

This paper evaluates the long-run effects of Head Start using large-scale, restricted 2000-2018 Census-ACS data linked to the SSA's Numident file, which contains exact date and county of birth. Using the county rollout of Head Start between 1965 and 1980 and age-eligibility cutoffs for school entry, we find that Head Start generated large increases in adult human capital and economic self-sufficiency, including a 0.65-year increase in schooling, a 2.7-percent increase in high-school completion, an 8.5-percent increase in college enrollment, and a 39-percent increase in college completion. These estimates imply sizable, long-term returns to public investments in large-scale preschool programs.

JEL Codes: I2, J24, J6

Contact Information

Bailey: Department of Economics, University of California-Los Angeles, 315 Pertola Plaza, Los Angeles, California 90095; Email: marthabailey@ucla.edu; Website: https://sites.google.com/g.ucla.edu/marthajbailey. Sun: The World Bank, 1818 H Street N.W., Washington, D.C. 20433; sqsun@worldbank.org; Timpe: Department of Economics, University of Nebraska, 730 N. 14th Street, Lincoln, Nebraska 68588; btimpe@unl.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. All views expressed in this paper are those of the authors alone and need not necessarily reflect those of the World Bank. 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. Jacob Bastian, Ariel Binder, Dorian Carloni, and Bryan Stuart also contributed substantially to the cleaning of the restricted census data. We are grateful for thoughtful comments from Liz Cascio, Janet Currie, Greg Duncan, Chloe Gibbs, Rita Ginja, Doug Miller, and Gary Solon.

In 1965, the U.S. began a new experiment in the provision of public preschool 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" (U.S. Senate Committee on Labor and Public Welfare, 1964). The ensuing program is the now-famous "Head Start," a "prep school for poor kids" which aimed to help millions of children escape poverty (Levitan, 1969).

More than fifty years later, Head Start is one of the most popular of the War on Poverty's programs, serving more than 1 million children at a cost of $10 billion in 2019.[1] Unlike expensive, small-scale "model" programs such as Perry Preschool and Abecedarian, Head Start's architects prioritized widespread access, calculating that a massive preschool expansion would maximize its poverty-fighting (and political) benefits. Skepticism about the quality of this large-scale preschool program coupled with difficulties in evaluation have generated controversy over its short-term benefits for decades (Currie, 2001; Duncan & Magnuson, 2013; Westinghouse Learning Corporation, 1969). Convincing evidence regarding Head Start's long-term effects 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. Consequently, the best estimates of Head Start's long-term effects are limited by lingering concerns about endogeneity (sibling comparison designs, Currie and Thomas 1995, Garces et al. 2002, Deming 2009) and imprecision (due to measurement error in funding and access, Ludwig and Miller 2007, and small sample sizes, Carneiro and Ginja 2014). Whether Head Start achieved its goal of increasing life opportunities for children remains an open question.

This paper uses large-scale data to estimate Head Start's long-term effects on human capital and economic self-sufficiency. By linking the restricted long-form 2000 Census and 2001-2018 American Community Surveys (ACS) to the exact date and place of birth from the Social Security Administration's (SSA) Numident file, we observe outcomes for one-quarter of U.S. adults as well as a high-quality measure of their access to and eligibility for Head Start as children. The resulting sample is four orders of magnitude

[1] https://eclkc.ohs.acf.hhs.gov/sites/default/files/pdf/no-search/hs-program-fact-sheet-2019.pdf

(or 10,000 times) larger than longitudinal surveys, and information on place of birth and exact date of birth ameliorates measurement error in childhood access to Head Start.

Our research design exploits the county-level roll-out of Head Start programs from 1965 to 1980 at the Office of Economic Opportunity (OEO) (Bailey, 2012; Bailey & Danziger, 2013; Bailey & Duquette, 2014; Bailey & Goodman-Bacon, 2015; Levine, 1970). This approach exploits the well-documented "great administrative confusion" at the OEO (Levine 1970), mitigating problems of measurement error in archival funding data (Barr & Gibbs, 2017) and concerns about the endogeneity of Head Start funding *levels*. An additional strength of our design is that it leverages Head Start's age-eligibility guidelines, comparing cohorts who were age-eligible when it launched (ages 5 and younger) to cohorts born in the same county that were age-*in*eligible (children 6 and older). A key identifying assumption is that 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. We examine changes in this relationship using an event-study specification and summarize them using a three-part, linear spline, which permits a formal test of a trend break at age six—the age where eligibility for the Head Start program abruptly changed.

The results suggest that Head Start increased the human capital and economic self-sufficiency of disadvantaged children. An index of adult human capital rose by 18 percent of a standard deviation among Head Start participants relative to children born in the same county who were age 6 when the program began. Participating children achieved 0.65 more years of education, were 2.7 percent more likely to complete high school, and were 8.5 percent more likely to enroll in college. College completion rates rose by 12 percentage points among participating children, an increase of 39 percent. In addition, Head Start increased economic self-sufficiency in adulthood by 9 percent of a standard deviation—gains driven by increases in the extensive and intensive margins of employment and reductions in adult poverty and public assistance receipt. Children participating in Head Start were 4 percentage points more likely to have worked as adults; they also worked two more weeks per year and three more hours per week on average as adults. Although selection into paid employment results in no measured effect on wage earnings, participation in Head Start appears to have reduced men's public assistance receipt (e.g., disability insurance) by 4.8

percentage points (42 percent) and adult poverty rates among women by 4.4 percentage points (32 percent).

Heterogeneity in Head Start's effects suggests that they reflect, in part, practices outside of a formal preschool curriculum. In particular, health screenings and referrals as well as more nutritious meals appear to be important mechanisms for the program's effects on disadvantaged children. In addition, the effects of Head Start seem to complement greater family and public resources arising from a stronger economy. Overall, Head Start appears to have achieved the goals of its early architects, both increasing children's economic opportunities and reducing poverty.

A final analysis quantifies both the private and public returns to dollars spent on Head Start in the 1960s and 1970s. Rather than using changes in wage income directly, we use the *National Longitudinal Survey of Youth* 1979 (NLSY79) to predict changes in *potential* earnings for the relevant cohorts net of any ability differences (Deming, 2009; Neal & Johnson, 1996). Potential earnings account for negative selection due to Head Start-induced employment increases. This exercise suggests a private internal rate of return to Head Start of 13.7 percent, which is similar for both men and women. Using *only* savings on public assistance expenditures and increases in tax revenue due to higher wage earnings, we find that the public internal rate of return of putting one child through Head Start ranges from 5.4 to 9.1 percent. While the size of the fiscal externality varies with different assumptions about the marginal beneficiary, the bottom line is that America's first national, public preschool program generated sizeable returns over the lifetimes of its first participants. While the results do not imply that all of today's large-scale preschool programs work, they suggest that some less-than-model preschool programs can have lasting effects—a key finding for current policy deliberations (Phillips et al., 2017).

I. The Launch of Head Start in the 1960s and Expected Effects

In the 1960s, the idea that preschool could improve children's cognitive development was revolutionary. Challenging the conventional notion that IQ was immutable and fixed at birth, Joseph McVicker Hunt's (1961) book, *Intelligence and Experiences*, persuasively argued that children's intelligence could be significantly improved by altering their experiences. Benjamin Bloom further emphasized that the first four years of children's lives was a "critical period," noting that "intelligence appears to develop as much from conception to age 4 as it does from age 4 to 18" (1964). This idea 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.

A. A Brief History of Head Start's Launch

Funded by the OEO, Head Start began as an 8-week summer program in 1965. After a successful first summer, President Lyndon Johnson announced that Head Start would become a full-year program for children ages 3 to 5.[2] 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 & Miller, 2007).

Head Start's political popularity led to an even faster launch than other War on Poverty programs. Figure 1 shows the program's quick expansion. 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, federal expenditures on the program reached $326 million, or $2.1 billion in 2018 dollars (OEO, 1970). This early expansion ensured that by 1970, Head Start existed in nearly half of U.S. counties, putting preschool programs within a short drive for 83 percent of children under age six (Online Appendix Table A1).

The exact timing of Head Start's launch depended on many idiosyncratic factors. The OEO's "wild sort of grant-making operation" has been well documented in oral histories (Gillette 1996: 193) as well as in more recent, quantitative analyses (Bailey & Duquette, 2014). In the case of Head Start, other factors were key as well: how excited were local institutions or politicians about the program? Was there adequate and available space to launch? Could the program be integrated within the public school system or would it remain separate? The final result of the grant-making process was that Head Start began earlier in areas that were significantly more populous and urban and, consequently, in areas with higher median family income (Online Appendix Table A2). After accounting for these differences, the roll-out of Head Start was unrelated by many other county characteristics (Online Appendix Table A3). Consistent with the historical evidence that this national program was rushed into existence, pre-existing local characteristics do not

[2] "Preschool" also included five-year-olds, because public kindergarten was not yet universal (E. U. Cascio, 2009).

systematically predict the date that Head Start programs launched.

B. Head Start's Mission

Head Start's architects adopted a holistic approach that aimed to develop children's mental and physical abilities by improving health; self-confidence; verbal, conceptual, and relational skills; and raising parental involvement. Levitan (1969) notes that Head Start's 1966-7 budget included early childhood education (daily activities and transport, 70 percent), health services (including immunizations, screenings and medical referrals), and nutrition (17-20 percent). Parental involvement, social services (e.g., helping families cope with crises), and mental health services accounted for the remaining budget.

The effects of Head Start on adult outcomes could result from its early education components. But the program's health and nutrition services were likely important as well. Head Start's vaccinations and screening (e.g., tuberculosis, diabetes, vision, hearing) and referrals to local physicians may have prevented complications from childhood diseases (Ludwig & Miller, 2007; North, 1979) and helped parents obtain simple, cost-effective technologies to improve learning (e.g., eye glasses and hearing aids or antibiotics to reduce hearing damage from ear infections). Healthy meals and snacks may have also raised children's ability to learn. Early estimates suggest that more than 40 percent of children entering Head Start were receiving less than two-thirds of the recommended allotment of iron, and 10 percent were extremely deprived in terms of their daily calories (Fosburg et al., 1984). Among children who received blood tests in the 1968 full-year program, 15 percent were found to be anemic (DHEW 1970). Reducing these nutritional deficiencies could also translate into significant gains in educational achievement in both the short and longer term (Frisvold, 2015).

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 (Braun & Edwards, 1972; Hechinger, 1966).[3] In addition, many components of Head Start phased in slowly. For instance, the OEO wrote that in 1965, "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 expect it to have reached over 90 percent" (OEO, 1967).

Consequently, Head Start in its earliest years was far from a model preschool program. Nevertheless, even the less-than-ideal implementation of Head Start was likely higher quality than the alternatives available to low-income children in the 1960s (Currie, 2001). Importantly, similar concerns hold today: Head Start's quality score from the National Institute for Early Education Research places Head Start program quality around the median of the score distribution (Espinosa, 2002) but the program may still be much better than informal child care (Loeb, 2016).

II. Evaluations of the Long-Term Effects of Head Start

Previous evaluations of Head Start suggest the program had long-term effects on human capital and economic self-sufficiency.[4] One pioneering approach was the use of family fixed effects with longitudinal data. Building on work by Currie and Thomas (1995), Garces, Thomas, and Currie (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). Bauer and Schanzenbach (2016) use sibling-based comparisons with more recent NLSY cohorts than Deming and find increases in high school completion and college attendance, especially among African-Americans. They

[3] Sizable variation in preschool quality persists today. For an overview see Currie (2001), E. U. Cascio and Schanzenbach (2013), and Duncan and Magnuson (2013).
[4] See reviews of studies of Head Start's short-term effects (E. U. Cascio & Schanzenbach, 2013; Currie, 2001; Duncan & Magnuson, 2013; Gibbs, Ludwig, & Miller, 2014).

also find an increase in receipt of post-secondary credentials, including post-secondary licenses or certificates, associate's degrees, and bachelor's degrees. Well-known critiques caution that sibling comparisons may suffer from sources of endogeneity bias (Bound & Solon, 1999; Griliches, 1979), and the sample identifying the treatment effects may be highly selected (Miller, Grosz, & Shenhav, 2019). In addition, small sample sizes in longitudinal surveys may provide unreliable estimates of Head Start's effects (Miller et al., 2019).

More recent work exploits variation in access to Head Start using four distinct research designs. The path-breaking application of RD in Ludwig and Miller (2007) 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.[5] 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 find a positive though statistically insignificant effect on receiving a high-school diploma as well as suggestive evidence that Head Start *reduced* college enrollment. Finally, De Haan and Leuven (2020) place bounds on Head Start's effects using a partial identification approach and imposing assumptions on the distribution of potential outcomes. Using data from the NLSY, they rule out null and negative effects at low levels of educational attainment and wage income, and they also find that the program's effects are largest among the most disadvantaged and women. However, the procedure limits their ability to characterize the treatment effects precisely.

In work closely related to this paper, three studies make use of county-year variation in Head Start

[5] Also, limited evidence shows the poorest 300 counties were more likely to get funding for Head Start (see Ludwig and Miller 2007: Table II and Pihl 2017).

funding in the 1960s and 1970s to quantify the program's long-term effects. Using a sample of likely eligible children from the NLSY, Thompson (2018) 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 raise adult household income. Focusing on a "high impact" sample in the PSID, Johnson and Jackson (2019) find that exposure to an average level of Head Start and education spending increased years of schooling by about one-third of a year and increased high school graduation rates by 10 percentage points. These children also experienced a 10 log-point increase in adult wages, a 7.6 percentage-point reduction in poverty at ages 20 to 50, and a 2.5 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 (also referred to as "roll-out"). To alleviate concerns about the endogeneity of funding levels and measurement error in the National Archives data, their roll-out design uses a binary measure of Head Start access that is equal to one if funding exceeds the 10^{th} percentile of observed funding per four-year-old. They find evidence of large first-generation effects on women (including a gain of half a year of schooling) and large second-generation effects on their children's high school graduation and completed education.

III. DATA AND RESEARCH DESIGN

This study combines the long-form 2000 Census and 2001-2018 ACS with the SSA's Numident file to shed new light on Head Start's long-term effects. The Census/ACS data represent almost one quarter of the U.S. population and are four orders of magnitude (or 10,000 times) larger than previously used longitudinal samples. Linking these data to the Numident file, a database of administrative records of applications for Social Security cards, adds information on exact date of birth and county of birth (rather than adulthood residence or state of birth) which allows us to approximate Head Start access and age eligibility in childhood.[6] The data's main disadvantage is that they contain no information on family

[6] The Numident place-of-birth variable is a string variable detailing place of birth. We adapt Isen, Rossin-Slater, and Walker (2013) and Black, Sanders, Taylor, and Taylor (2015) to construct a crosswalk between this string variable and county FIPS codes. See Taylor, Stuart, and Bailey (2016) and Online Appendix section 2 for more details.

background.[7] This lack of covariates means that we cannot model many determinants of adult outcomes—which limits precision even in this large dataset—or model treatment effect heterogeneity by childhood characteristics.

Our sample is comprised of 22.48 million children born from 1950 to 1980 in the continental U.S. We additionally limit our sample to individuals who are in their prime earning years (ages 25 to 54).[8] We collapse these data to means by birth year, survey year, county of birth, and school age.[9] We also weight our regressions using the number of observations in each cell (Solon, Haider, & Wooldridge, 2015). To minimize disclosure concerns at the Census Bureau, we use only observations with non-allocated and non-missing values for all outcomes.[10]

Primary outcomes of interest include summary measures of human capital and economic self-sufficiency, which permit tests of co-movements of related adult outcomes and limit the number of statistical tests (Kling, Liebman, & Katz, 2007). The human capital index includes as subcomponents four binary variables for 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. The index of self-sufficiency includes binary indicators of employment, poverty status, income from public sources, family income, and income from other non-governmental sources; continuous measures of weeks worked, usual hours worked, the log of labor income, log of other income from non-governmental sources, and log ratio of family income to the federal poverty threshold. All subcomponents are coded so that positive values reflect improvements: the binary indicators of poverty status and public-source income receipt are, therefore, reverse coded. The indices standardize each subcomponent and average them, weighting each subcomponent equally. Because changes in the indices do not distinguish an effect driven by a dramatic shift in one subcomponent from an effect driven

[7] For instance, we cannot focus on a high impact sample: adults who were very poor as children and would have been much more likely to participate in Head Start. 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).
[8] We find no evidence that Head Start affected survival to 2000 (see Online Appendix Table A6).
[9] School age is defined using exact date of birth and school-entry age cutoffs and relies on Bedard and Dhuey (2006) as well as the authors' research.
[10] This restriction has minimal effects on our estimates (see Online Appendix Figures A5 and A6), and we find no evidence that this sample restriction induces differential selection by treatment status into our sample.

by changes in all subcomponents, we also examine the subcomponents separately to understand the mechanisms.

A. Measuring Exposure to Head Start

Combining data on 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 & Gibbs, 2017; Johnson & Jackson, 2019; Thompson, 2018). First, we use only variation in the launch of the Head Start program rather than a continuous measure of Head Start spending. This refinement (1) addresses the potential endogeneity of Head Start funding levels to the program's *performance* and (2) sidesteps issues of measurement error in the National Archives grant data (Barr & Gibbs, 2017).[11] Second, we examine changes in outcomes for children who were *age-eligible* for Head Start (ages three to five or younger) relative to those who were age-ineligible (ages six and older) when it launched, allowing for the effects to vary by the number of years each cohort was potentially eligible. 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. Finally, our large dataset allows us to use state-by-birth-year fixed effects to adjust estimates for state economic and policy changes that could have affected children's outcomes independently of Head Start.[12] Our identifying assumption in the analysis that follows 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. (See Online Appendix for more description of our sample.)

B. Event-Study Regression Model

Our research design uses a flexible event-study framework and the roll-out of Head Start to estimate the effect of exposure to Head Start on long-term human capital and economic outcomes,

$$Y_{bct} = \theta_c + \alpha_t + \delta_{s(c)b} + \boldsymbol{Z}'_c b\boldsymbol{\beta} + HeadStart_c \boldsymbol{Age}'_{bs(c)}\boldsymbol{\varphi} + \varepsilon_{bct}. \tag{1}$$

Children's birth years are indexed by b=1950−1980, county of birth by c, and Census/ACS year by t=2000-

[11] Thompson (2018) also tries this strategy but notes that his estimates in the NLSY are statistically insignificant.

[12] For instance, state policies surrounding the funding of education were changing for our cohorts of interest (E. Cascio, Gordon, Lewis, & Reber, 2010). Online Appendix Tables A8-A9 document how the estimated effects of Head Start change in our Census/ACS-Numident dataset using (1) alternative measures of access to Head Start including funding per capita and years of access to Head Start as well as (2) including state-by-birth-cohort fixed effects.

2018. Specifications include fixed effects for county of birth, θ_c, year, α_t, and state-by-birth-year, $\delta_{s(c)b}$, which, respectively, capture time-invariant differences across counties, national changes affecting all cohorts, and changes in state policies that differentially affect birth cohorts. We follow the literature and include county characteristics, $\boldsymbol{Z}_c$ interacted with a linear trend in year of birth, b (Bailey, 2012; Bailey & Goodman-Bacon, 2015; Hoynes, Page, & Stevens, 2011; Hoynes & Schanzenbach, 2009; Hoynes, Schanzenbach, & Almond, 2016). These 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, and account for secular trends toward *worse* outcomes in poor, urban areas over the 1960s and 1970s.

HeadStart is a binary variable equal to 1 if a child was born in a county that received a Head Start grant before 1980, and 0 for children born in counties that never received Head Start grants during our sample period. ***Age*** is a vector of dummy variables for a child's age when Head Start was introduced in her county of birth *c*, with age measured at the school entry cutoff date in her state of birth s.[13] We include individual event-study dummies for -10 to +14 in our estimating equation, and we group event-time less than or equal to -11 or greater than or equal to +15 into single dummies to avoid collinearity. Our event-study coefficients are balanced by county and cohort from event-time -2 through 14, but imbalanced county-cohorts make up a relatively minor share of the sample outside this range so are included in all figures.[14] We omit school age six (that is, age 6 by the school entry cutoff date), because these children would have been age-eligible for first grade and, therefore, unlikely to attend Head Start during the school year. Our point estimates of interest, $\boldsymbol{\varphi}$, describe the evolution of the intent-to-treat (ITT) effects of Head Start on long-term human capital and economic self-sufficiency.

All standard errors are corrected for heteroskedasticity and adjusted for an arbitrary within-birth-

[13] For example, consider a child born October 1, 1960, in a county where Head Start started in fall of 1966. If the state's cutoff for turning age six for first grade entry was December 1, we would code the child as school age six in fall of 1966. However, if the state's age cutoff for first grade entry was September 1, we would code the child as school age five in fall of 1966.

[14] Between ages -3 and -10, we lose about 85 late-adopting counties out of more than 1,500 treated counties in all (see Online Appendix Tables A1), representing about 2 percent of the children born in Head Start counties during our sample period. Our estimates are very similar when we omit 85 counties from the analysis.

county covariance structure (Arellano, 1987; Bertrand, Duflo, & Mullainathan, 2004).[15] In our tables, we also report p-values using the Bonferroni-Holm method to correct for multiple hypothesis testing (Duflo, Glennerster, & Kremer, 2007; Holm, 1979).

C. *Expected Effects of Exposure to Head Start by Age at Launch*

The event-study model is flexible and imposes few restrictions on the relationship between Head Start and adult outcomes. Although economic theory does not make predictions about the magnitudes of the event-study coefficients, the program's phased implementation and the greater potential for some children to enroll (due to multiple years of exposure) imply a *pattern* of estimates.

First, if we assume Head Start had no effect on children who were over age five when the program launched (and no spill-over effects onto older siblings), then the relationship between adult outcomes and Head Start for these children should be zero. This is equivalent to a test for a pre-trend in our analysis and is illustrated as a flat line for children ages six to 14 in each panel of Figure 2.

Second, if Head Start has a positive causal effect on adult outcomes, we expect the outcomes of treated cohorts to *change* relative to the cohorts too old to benefit (i.e., cohorts age ineligible). This change should be apparent as a level shift or a slope change for children under age five when the program launched, because these cohorts would have been the first to have been age-eligible *and* have access. A level shift would be consistent with a one-time, immediate change in Head Start's capacity (in terms of participants) and quality (in terms of services and funding per child) which remained constant (Figure 2A, solid line no markers). However, several institutional features suggest a more gradual change.

1. Head Start's capacity grew as the national program and individual programs matured. Consistent with this prediction, the full-year program served only 20,000 children in 1965 but 160,000 in 1966, 215,000 in 1967 and 1968, and 257,700 in 1970 (OEO, 1965, 1966, 1967, 1968, 1970). (See Online Appendix Figure A2 for the growth in capacity.)

---

[15] We also implement alternative standard error corrections for clustering by birth state and, separately, two-way clustering by birth-county and year (Cameron, Gelbach, & Miller, 2011)—neither of which affects inference. Because the Census Bureau has requested that we reduce disclosures for this project and because these alternative corrections have little effect on our conclusions, we have not disclosed these additional estimates.

2. Program quality also increased over time with better hiring and training of teachers, curriculum development, and the implementation of auxiliary services (e.g., health).[16]
3. We expect larger effects for children who were younger when Head Start launched due to the age structure of program admissions. For instance, a child 5 years old when Head Start launched could participate for at most one year, whereas a 3-year-old child would be age-eligible for three years. This does not necessarily assume that a 3-year-old is more likely enrolled for more than one year, but rather that it is more likely that a child would enroll if s/he had three years to do so.

The combination of phased implementation and cumulative potential access to Head Start implies a shift in the *slope* of the relationship between adult outcomes and a child's age at Head Start's launch around age six (Figure 2B). In addition, differences in the likelihood of enrollment by age could make this relationship more S-shaped because enrollment in the early years was more likely at ages four than five (circle markers).

Another possibility is that Head Start benefitted some older children or had spill-over effects on older children or the older siblings of three- to five-year-olds participating in the program. This implies that the relationship could begin to slope up *at ages older than six* (Figure 2C). This relationship would be consistent with reports that 10 percent of children in full-year Head Start were six or older (Vinovskis 2008), and age-ineligible children could still benefit from their younger siblings' participation (Garces et al., 2002).[17] Because our analysis computes the event-study effects relative to age six, effects of Head Start on older children would appear as the event-study estimates falling below zero.

A third feature in the pattern of intention-to-treat effects is that we expect them to flatten, as in each of the cases in Figures 2A-2C. A leveling off of the intention-to-treat effects reflects the fact that program capacity is reached, full implementation is achieved, and all children born after a certain point will have three years of access to a fully implemented and full-capacity program. Unfortunately, statistics are only

[16] We suspect that the speed of implementation varied with the year of Head Start's implementation—programs starting later could adopt best practices faster. However, the rapid roll-out of Head Start programs limits our ability to test for this heterogeneity. Studies of other War on Poverty programs such as family planning or community health centers suggest that many of these programs reached maturity around 4 to 5 years after launch (Bailey, 2012; Bailey & Goodman-Bacon, 2015).
[17] For the age distribution of children in Head Start, see Thompson (2018)'s Table 1.

published for Head Start programs open in a fiscal year and not for programs by their years in existence. Little systematic evidence characterizes when programs reached full capacity or quality, and the data provide no straightforward predictions about when full implementation—and leveling off—should occur. Our event-study specification allows us to estimate the relationship using the data without imposing assumptions about which of the patterns in Figures 2A-2C – or which combinations of these patterns – we might find.

D. Spline Summary Specification

We summarize our event-study estimates using a spline specification. The spline is especially helpful in cases where we do not have space to present the event studies and for smaller samples and noisier outcomes (by improving precision). The spline also sidesteps the drawbacks of simpler differences-in-differences estimators, which would only fit predictions in Figure 2A and may be difficult to interpret in settings with staggered adoption and dynamic treatment effects (Borusyak & Jaravel, 2018; Callaway & Sant'Anna, 2019; de Chaisemartin & D'Haultfœuille, 2020; Goodman-Bacon, 2018).

Based on Figure 2's predictions, we *restrict* $\boldsymbol{Age}_{bs(c)}$ in equation (1) to be a three-part, linear spline with its components defined by age at the time of Head Start's launch, $a = T_c^* - B_{s(c)}$. One knot of the spline falls at age six—the age at which most children attended primary school and not full-year Head Start. The location of the second knot—the leveling-off point—is not defined by institutional or theoretical considerations, because local capacity growth and quality are unobserved for most Head Start programs. We, therefore, adopt a data-driven approach to inform this choice. We regress a composite of the human capital and economic self-sufficiency indices on our covariates in equation (1) and replace $\boldsymbol{Age}_{bs(c)}$ with different spline specifications, fixing one knot at age six and allowing the other knot to vary between -8 and +5. The idea is to use the outcome data to find the location of the second knot that maximizes the within R-squared. We find that this occurs at $a=-5$, which is consistent with Head Start enrollment increasing through the late 1970s—roughly one decade after it began—although few individual programs launched after 1970. In keeping with the result of this data-driven procedure, we define age –5 as treated with a fully implemented and full-capacity program and focus on this cohort for the main estimates.

We estimate the following specification, constraining $\boldsymbol{\rho}_1$ and $\boldsymbol{\rho}_2$ to ensure that the spline joins at a=6 and -5,

$$(2) \qquad Y_{bct} = \theta_c + \alpha_t + \delta_{s(c)b} + \boldsymbol{Z}'_c b\boldsymbol{\beta} + HeadStart_c(\boldsymbol{D}'_{cb}\boldsymbol{\rho}_1 + a\boldsymbol{D}'_{cb}\boldsymbol{\rho}_2) + \varepsilon_{bct},$$

where $\boldsymbol{D}'_{cb}$ is a vector of dummy variables, $1(-10 \leq a \leq -5)$, $1(-5 \leq a \leq 6)$, $1(6 \leq a \leq 14)$, and the other variables remain as previously defined. For comparability with the event study, we group estimates less than or equal to -11 or greater than or equal to +15 into single dummies.

The spline also has the advantage of embedding formal tests of the research design's identifying assumptions. First, testing for parallel pre-trends is equivalent to testing whether the slope of the spline for ages six to 14 is zero. In some cases, we find a pre-trend that works against finding an effect, so we use the pre-treatment spline to adjust for this apparent violation of parallel trends. This involves extrapolating our estimated pre-trend to younger cohorts and then projecting the treatment effect at -5 relative to this counterfactual, where the standard error of this point estimate accounts for the variance of both spline components as well as their covariance. Because adjustments for pre-trends can be controversial (Freyaldenhoven, Hansen, & Shapiro, 2019), our tables present both unadjusted and adjusted estimates for each index outcome.[18] Second, the spline embeds a test of the post-trend slope (segment from –5 to –10), which we also expect to be zero (net of any pre-trend). Note that the goodness of fit procedure to select the second knot imposes no *ex ante* restrictions on the *slope* of the post-trend, which could take on any value. Third, the spline permits a formal trend-break test at age six, the age at which the relationship between adult outcomes and age at Head Start's introduction is expected to *change.* We report each of these tests in tables alongside our main estimates.

IV. Tests of Identifying Assumptions

The research design relies on two crucial assumptions: (1) receiving a Head Start grant increased participation in Head Start (relevance), and (2) the "parallel trends" assumption that the outcomes of interest

[18] This projection takes the difference between the "phase-in" and "pre-trend" spline slopes and multiplies the result by -11 (the number of years between age 6 and age -5). We calculate the standard error of this linear combination of estimates using standard procedures.

for treated cohorts would have evolved similarly to those of untreated cohorts in the absence of the Head Start program (validity). This section provides evidence regarding both assumptions.

A. How Much Did Head Start Increase Preschool Enrollment?

Administrative data suggest that the launch of a Head Start program significantly increased children's enrollment. The OEO reported that full-year Head Start served over 600,000 children before 1968, rising from 20,000 children in 1965, to 160,000 in 1966, to around 215,000 in 1967 and 1968 (OEO, 1965, 1966, 1967, 1968, 1970).[19] About 257,700 children attended full-year Head Start in 1970. Three-quarters of the children were aged 4 or 5, three-quarters were nonwhite, and 62 percent came from families with less than $4,000 in annual income. Between 1971 and 1978, enrollment increased as funding grew. Directory information suggests that the average county with a Head Start program served roughly 309 children. These sources imply that the average full-year Head Start program served from about 10 percent of resident age-eligible children in 1971 to 15.8 percent in 1978. Including funded summer slots—which could be converted to school year slots under federal rules—brings estimates of Head Start enrollment to 18 percent in 1970, 15 percent in 1971, and 17 percent in 1978.[20]

While there is little doubt that introducing a Head Start program increased children's attendance in the program, the magnitude of this relationship net of crowd-out is crucial for interpreting the ITT effects in equations (1) and (2). If Head Start substituted for private preschool for some children (Bassok, Fitzpatrick, & Loeb, forthcoming; E. U. Cascio & Schanzenbach, 2013; Kline & Walters, 2016), administrative data may overstate the role of Head Start programs in increasing exposure to preschool. To examine this possibility, we use the 1970 Census, which was the first to ask children *younger than age 5* about school enrollment as of February 1—a date *during* the school year, which should capture enrollment

[19] 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 no more than 10 percent of those in full-year programs were older than five. See also Table 1 in Thompson (2018).
[20] U.S. Office of Child Development (1972) notes that there was a "community option in converting funds from summer to full-year programs" beginning in the 1969-1970 academic year (p. 1).

in *full-year* Head Start (Ruggles, Genadek, Grover, & Sobek, 2015).[21] Public-use Census data show that four-year-old children in counties without Head Start programs were 3.4 percentage points less likely to be enrolled in school (16.8 versus 20.2 percentage points, see Figure A1). Five-year-old children were 17 percentage points less likely to be enrolled in school (48.9 versus 65.9 percentage points). These gaps are 5.9 percentage points among four-year-olds and 21.3 percentage points among five-year-olds when looking only at children of mothers with less than a high school education.[22]

We use a linear probability model on the restricted-use Census data, which provides greater detail on place of residence, to adjust these differences for state fixed effects (to account for age-invariant, state-level factors that determine the local supply of preschools) and 1960 county characteristics (share of county population in urban areas, in rural areas, under 5 years of age, 65 or older, nonwhite, with 12 or more years of education, with less than 4 years of education, in households with income less than $3,000, in households with incomes greater than $10,000, local government expenditures, income per capita, and whether the county was among the 300 poorest counties). The regression results show that school enrollment was 14.9 percentage points higher for all five-year-olds, 15.1 percentage points higher for boys, and 14.5 percentage points higher for girls (Online Appendix Table A5). These results are robust to the inclusion (or exclusion) of different covariates (see Online Appendix section 6).

The 14.9-percentage-point increase in enrollment in the Census is between the 10 and 18 percent estimates from administrative data, suggesting minimal crowd-out. The estimate of 14.9 percentage points is also in the range of other studies. Garces et al. (2002) estimates the national Head Start participation rate was between 10 percent and 17 percent for the 1964 to 1970 cohorts in the PSID (p. 1002) and Ludwig and Miller (2007) estimate that children's enrollment in Head Start was 17 percentage points higher at the 300-poorest county discontinuity in the 1988 National Educational Longitudinal Survey.

[21] Although the 1960 Census asked about school enrollment (including kindergarten), the question was only reliably asked of children ages 5 and older which precludes analysis of preschool aged children.
[22] Note that Head Start was not exclusively for poor kids in the 1960s and 1970s. To encourage interaction between poor children and those from less disadvantaged backgrounds, OEO policy allowed 15 percent, and later 10 percent, of children to come from families that did not meet its poverty criteria. Roughly two-thirds of children in the full-year 1969 and 1970 programs came from families in which the mother had less than a high school education, although the mothers of about 7 percent of children had attended or graduated from college.

Based on this evidence, we use the estimate of 14.9 percentage points from the Census to transform the ITT effects from the spline specifications into average treatment-effects-on-the-treated (ATET). We also construct confidence intervals using a parametric bootstrap procedure with 10,000 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.[23] Rather than impose independence between the magnitude of the Head Start take-up estimates (first stage) and ITT estimates (reduced form), we sample counties with replacement, estimate separate regressions for the first stage and reduced form, and repeat the procedure 1,000 times to compute the correlation between first-stage and reduced-form effects. This produces an estimated correlation of 0.07, which modestly reduces the 95-percent confidence intervals for these estimates.

B. Did Head Start's Launch Correspond to Other Policy Changes?

The parallel-trends assumption is also central to our analysis. Our event-study and spline specifications provide visual and formal statistical tests for parallel *pre*-trends, which we present alongside the results in the next section. However, the parallel-trends assumption additionally requires that there were no confounding shocks or policy changes that occurred at the same time or just after Head Start began.

To provide additional evidence regarding the validity of our research design, we use information on other War on Poverty programs compiled from the National Archives and estimate regressions similar to equation (1). In particular, we replace the dependent variable with an indicator for receiving funding in county c in fiscal year t, or $Y_{ct} = \theta_c + \delta_{s(c)t} + \boldsymbol{X}'_{ct}\boldsymbol{\beta} + \sum_k \pi_k 1(t - T_c^* = k) + \varepsilon_{cbt}$. We include county and state-year fixed effects, θ_c and $\delta_{s(c)t}$, and county-level covariates, $\boldsymbol{X}_{ct}$, as we do in our main regression specifications. Our variable of interest is event-time, $t - T_c^* = k$, the year of observation relative to the date Head Start launched (the year before Head Start began, $k = -1$, is omitted).

Figure 3 shows the relationship between the launch of Head Start and the launch of other OEO programs. As expected, 100 percent of treated counties in our sample first received a Head Start grant in

[23] This procedure follows Efron and Tibshirani (1993) (pp. 53-6) and Johnston and DiNardo (1997) (pp. 365-6).

event-year 0. This is by design. The share of counties receiving a Head Start grant tapers off to 70 percent after around five years—this reflects the fact that some counties received multi-year grants and also the fact that not all of the early programs continued.

For our estimates of Head Start's effects to be confounded by other federal programs, grants for other programs would need to happen around the time Head Start launched in event-year 0. Figure 3, however, finds little evidence of such a relationship. Analyses of Food Stamps, Community Health Centers (CHCs), and other child health programs show no such pattern. The one program that shows a small change in funding after Head Start began is the Community Action Program (CAP) health project. Most CAP health grants were for the Emergency Food and Medical Services program (EFMS), later known as the Community Food and Nutrition Program, which was designed to provide food and medical supplies to counteract malnutrition and starvation and connect poor families to programs like Food Stamps. Importantly for this paper, the CAP health program the program was not directly targeted toward children or families with children. EFMS assistance was targeted at a broader population, including the elderly, and CAP health grants also included initiatives such as treatment of alcoholism and outreach to elderly individuals who were eligible for Medicare. In addition, it reached far fewer communities and was much smaller in funding at around $8 in 2013 dollars per person versus annual Head Start funding of nearly $1,500 per 4-year-old during the same period. In short, we find no evidence that the parallel-trends assumption is violated. Although we cannot rule out funding changes in programs we do not measure, these results support the validity of our research design.

V. HEAD START'S EFFECTS ON HUMAN CAPITAL

Figure 4 plots the event-study estimates for the human capital index and its subcomponents, with the x-axis plotting the ages of cohorts of children when Head Start launched (i.e., -10 is the cohort born 10 years after Head Start began, whereas +14 is the cohort born 14 years before the program launched).[24] The bold, solid line plots the event-study estimates (95-percent, point-wise confidence intervals in dashed lines),

[24] Note that cohorts to the left of -2 are slightly imbalanced. See Online Appendix for analyses of program roll-out by cohort and age. Accounting for this imbalance has negligible effects on the results.

and the dashed gray line plots the corresponding spline estimate. Consistent with the patterns in Figure 2B and 2C and the validity of the parallel-trends assumption, the human capital index and each of its components exhibit little relationship to adult outcomes for cohorts too old to benefit from Head Start (i.e., age 6 or older when it launched). In many cases, the slight trend works *against* the analysis finding effects. Positive estimates for pre-trends suggest that cohorts with access to Head Start would have been *worse* off than their older peers in the absence of the program.[25] However, the human capital index and its subcomponents exhibit a striking trend-break around age 6, suggesting that access to Head Start improved the human capital of adults.

Figure 4 also suggests our spline specification provides an informative parameterization of the intent-to-treat effects of Head Start on long-run outcomes. Using the spline estimates, Table 1 translates the ITT-effects into ATETs and summarizes statistical tests of the pre- and post-trends and the trend-break at age six; all estimates except those for years of schooling are multiplied by 100 to place them in percentage-point units. The ITT-spline estimate at –5 (the cohort that was exposed to a full capacity and fully mature program) suggests that Head Start significantly improved adult human capital. The ITT estimate at –5 shows that the standardized index was 2.1 percent of a standard deviation higher for the fully exposed cohort (column 2) without accounting for a slightly positive (off-setting) pre-trend in human capital accumulation. The remaining ITT estimates in Table 1 assume that estimated pre-trend is informative of the counterfactual trend after Head Start's launch. This assumption increases the estimated effects on human capital to 2.7 percent of a standard deviation for the fully exposed cohort. If we assume these gains came solely from attendance in full-year Head Start (and not through indirect channels such as sibling or peer participation in Head Start), this amounts to an increase of 18 percent of a standard deviation for treated children (column 6 divides the ITT estimate in column 2 by the take-up estimate of 0.149). Supporting the visual impression in the event-study plots, the data fail to reject the null of no pre-trend (slope of spline for ages six to 14 is equal to zero, column 3) but reject the equality of slopes in splines on both sides of age six

[25] See the insightful discussion by Rambachan and Roth (2020) regarding the importance of pre-trends in differences-in-differences analysis.

(column 4). This evidence of a trend break supports the interpretation that Head Start had a causal and positive effect on children's long-term outcomes. Finally, the data show that the intention-to-treat effects of Head Start level off for children with access to a full-capacity and mature program (column 5).

Underlying the striking change in children's human capital are large effects on some of the most commonly studied outcomes in the preschool literature, including high school graduation and college enrollment. Table 1 shows that treated children were 2.4 percentage points more likely to complete high school/GED (column 6)—a 2.7-percent increase relative to the control mean (column 7). The magnitude of this estimate is precisely estimated, but smaller than other estimates of Head Start's effects in the literature. Figure 5A shows that the effect is roughly two-thirds the size of Garces et al. (2002)'s sibling comparison in the PSID and half the size of Thompson (2018)'s spending design in the NLSY. It is one quarter the size of Deming et al. (2009)'s sibling comparison for Head Start in the 1990s for more recent cohorts. In addition, it is one-sixth the size of Johnson and Jackson (2019)'s spending design estimates for the very disadvantaged sample in the PSID; and one-seventh the size of Ludwig and Miller (2007)'s RD estimates using the Census. It falls just below, but within the 95th percent confidence interval of, the lower bound on the average causal effect estimated by De Haan and Leuven (2020). Although this paper's estimate of the effect of Head Start on high school completion falls within the 95-percent confidence intervals of other studies, this fact reflects the imprecision of many of those estimates.

Table 1 also shows a statistically significant effect of Head Start on college enrollment. Among participants, Head Start raised college enrollment by 5.4 percentage points, or 8.5 percent (column 6, column 7). Figure 5B shows that this estimate is three-fifths 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 0.054 is only slightly smaller than Deming (2009)'s NLSY sibling comparison for Head Start in the 1990s, but less than half the size of the estimate obtained by Bauer and Schanzenbach (2016) using an updated NLSY sample. These estimates are one-quarter to one-fifth the size of those found for the Abecedarian Project (Barnett & Masse, 2007; Currie, 2001; Duncan & Magnuson, 2013).

While our estimated treatment effects tend to be smaller than others in the literature, our analysis suggests that commonly used measures of Head Start access and alternative specifications yield smaller results than our splines when used with our data and control variables. Standard differences-in-differences estimators take a variance-weighted average of treatment effects across all years after Head Start's launch (including the early years in which programs were small and of lower quality) and yield an ATET of 1.2 to 1.3 percentage points for high school completion (Online Appendix Table A9, columns 2 and 3 of panel B, scaled by our first-stage estimate of 0.149) versus our estimate of 2.4 percentage points, which captures the effect of a full-capacity and mature Head Start program relative to unexposed cohorts. For college attendance, the differences-in-differences estimator yields an ATET of around 3.6 to 4 percentage points versus 5.4 percentage points using our approach (Online Appendix Table A9, columns 2 and 3 of panel D). Using Head Start funding per capita to measure Head Start access yields even smaller estimates. This reflects measurement error in Head Start county funding per capita in the National Archives data and the fact that grant amounts are not highly correlated with program capacity or quality (Online Appendix Figure A8).[26] The strength of our empirical strategy is that it minimizes the impact of measurement error in historical grant data (by using program launch dates) while explicitly allowing for the gradual implementation of Head Start, the potential for differential pre-trends, and the state-level changes in education policy and economic circumstances using state-by-birth-year fixed effects.[27]

Our large-scale data also permit a novel evaluation of the effects of Head Start on other dimensions of human capital, including college completion or higher degrees, which few previous studies have had power to detect. Table 1 shows that participating children were 12 percentage points more likely to graduate from college, an increase of 39 percent (column 7, ATET; column 4, trend break statistically significant at

[26] This is well documented. Ludwig and Miller (2007) note that "the accuracy of these data [on Head Start funding from the National Archives] is less than perfect given poor documentation and some obvious errors. In the end, only data from 1968 and 1972 were usable, in the sense that the electronic data matched published Head Start and other federal spending figures at the national and state levels. Even here measurement error arises from complications such as providers that run Head Start programs in multiple counties but are listed as receiving federal funds only in the county with the organization's headquarters." Barr and Gibbs (2017) document similar problems and went to great lengths to clean the National Archives data. Measurement error in National Archives grant data for other War on Poverty programs has led other papers to use a roll-out design rather than funding per capita (Bailey, 2012; Bailey & Goodman-Bacon, 2015; Bailey, Hoynes, Rossin-Slater, & Walker, 2020; Bailey, Malkova, & McLaren, 2018).

[27] We provide further discussion of these estimates, including the implications of state-by-birth-year fixed effects, in section 8 of the Online Appendix.

1-percent level). Similarly, completion of professional or doctoral degrees increased by 2.6 percentage points among treated children. These gains across the education distribution are summarized in a 0.65-year increase in schooling. This estimate is larger than Johnson and Jackson (2019)'s estimate of 0.44 years for very disadvantaged children (which may reflect the specification of Head Start access as funding per capita described above).[28]

Large effects on college and higher degrees may be surprising, given that few other studies of preschool have documented effects on post-secondary education. This lack of evidence may, in part, reflect the small longitudinal samples or the small scale of model preschool programs. Differences in the participating children may also matter. Abecedarian and Perry's participants were *very* disadvantaged children and mostly black, and Perry's participants had low IQs.[29] In contrast, Head Start was not exclusively for poor, African-American, or low-IQ children. Consequently, Head Start's participants in the 1960s and 1970s likely faced fewer socio-economic and cognitive disadvantages and less racism on average relative to participants in other model preschool programs. Differences in the background characteristics of Head Start's participants make it less surprising that they experienced gains in post-secondary education.

Because analyses of model preschool programs have found different educational effects for boys and girls, Table 2 stratifies our sample by sex. Among participating men, the human capital index increased by a statistically significant 15 to 17 percent of a standard deviation (ATET, column 6). These effects are almost identical across the unadjusted and pre-trend adjusted estimates, because any pre-trend appears very small. For participating men, high school completion rose by 3.1 percent, college attendance rose by 11 percent, and college completion rose by 37 percent (% change in ATET, column 7). The high school estimates are smaller than others in the literature, but the college attendance estimates are larger. Head Start cumulatively raised years of education among treated men by 0.60 years and the likelihood of completing

[28] Jackson and Johnson's ITT estimate is 0.07721 per $1000 spent per poor 4-year-old. We translate this into an ATET by multiplying the coefficient by 4.23 (the average Head Start spending per poor 4-year-old measured in thousands) and dividing by 0.75 (their estimate of take-up among income-eligible four-year-olds in counties with Head Start programs), so that 0.07721*4.23/0.75=0.44.

[29] The model Perry Preschool Program, which focused on lower IQ children, had no measured effects on postsecondary outcomes (Anderson, 2008).

a professional/doctoral degree by 3.2 percentage points (ATET, column 6). Men participating in Head Start as children were 19 percent more likely to hold professional jobs.

The human capital index increased by a similar amount among participating women at 19 percent of a standard deviation (ATET, column 6) after adjusting for the pre-trend. The unadjusted estimates are smaller, because the off-setting pre-trend for this group is larger. Among women participating in Head Start as children, completion of high school (or a GED) rose by 2.9 percent (% change in ATET, column 7) and college attendance rose by 5.8 percent (although the trend-break is not statistically significant). Changes in women's human capital index appear driven by increases in higher degrees, including a 12-percentage-point increase in college completion and a 2.1-percentage-point increase in professional degree attainment (ATET, column 6). Overall, years of schooling among participating women rose by 0.70 years and their likelihood of holding a professional job rose by 13 percentage points.

Table 3 further breaks down these estimates by broad race categories. The patterns for whites are similar to those in Tables 1 and 2, because whites constitute 87 percent of the overall sample. We use "nonwhite" as a second group. Nonwhite is not a racial identity and combines the different experiences of many groups, but the aggregation confers greater statistical power by combining the experiences of the 13 percent of our sample who were on average more economically and socially disadvantaged in childhood and more likely to face discrimination in labor markets as adults. While small sample sizes for nonwhites leave us unable to distinguish effects statistically between these groups, the pattern of estimates is consistent with the idea that the program offered a "head start" to children with fewer resources. Overall effects on the human capital index are comparable, increasing by 16 percent of a standard deviation for nonwhites and 19 percent for whites (ATET, column 6). Although the pre-trend adjustment matters little for whites, it is especially important for nonwhites, as their human capital outcomes were trending more negatively before the Head Start program started. However, an examination of the subindices suggests that human capital gains among whites represent increases in postsecondary education with weaker effects on high school graduation. In contrast, changes among nonwhites were driven by improvements lower in the educational distribution, with greater effects on completion of high school (or a GED), but smaller effects on professional or higher degrees. Years of schooling among participating nonwhites rose by 0.53 of a year, and the likelihood of holding a professional job rose by 5.6 percentage points (ATET, column 6). Online Appendix Table A10-A12 estimates effects separately by sex-race subgroups for the interested reader.

VI. HEAD START'S EFFECTS ON ECONOMIC SELF-SUFFICIENCY

The substantial effects of Head Start on human capital suggest that economic self-sufficiency may have also improved. Figure 6 plots the event-study estimates for the economic self-sufficiency index and its subcomponents. The index of economic self-sufficiency exhibits a similar pattern to the human capital index overall, but the event-study estimates are noisier and the results are more sensitive to pre-trend adjustments. Adjusting for offsetting pre-trends sharply increases some of the estimated effects, because subcomponents related to employment were worsening in (typically more urban) locations where Head Start was set up for younger cohorts. Our discussion, therefore, relies on the pre-trend adjusted figures in Table 4, which summarizes the estimates for the spline and trend-break test. Note, however, that estimates for the index subcomponents "in poverty" and "received public assistance" show little evidence of pre-trends and are not sensitive to pre-trend adjustments.

Based on pre-trend adjusted estimates, the economic self-sufficiency index was 9.2 percent of a standard deviation higher for children who attended a full-capacity and mature Head Start program (ATET, column 6). Head Start participants were 5.3 percent (4.4 percentage points) more likely to work in the previous year, and they worked 2.3 weeks and 3 hours more on average (column 6). The program also decreased the likelihood of adult poverty by 23 percent and receipt of public assistance income by 27 percent among participants (column 7). These subcomponents of the index are individually statistically significant at the 5-percent level when using the Bonferroni-Holmes method to account for multiple hypothesis testing. The fact that labor income and family income relative to the poverty line do not increase by a statistically significant amount is perhaps surprising. However, this finding is consistent with Head Start raising employment among less skilled individuals, who tended to earn less than the average worker already in the labor force. Negative selection would tend to off-set wage gains due to increases in human capital and experience. We return to this point and use a potential wages framework to quantify the effect of Head Start on wages after accounting for selection.

Ample theoretical reasons suggest that the self-sufficiency of men and women may have changed in different ways. Whereas Head Start's effects on men's human capital may have led them to increase employment (e.g., the substitution effect dominates the income effect), the program's effects on women's human capital may have helped them marry a higher-earning spouse and, potentially, work less for pay and more in the household (e.g., the income effect dominates).

Table 5 investigates these differences. For male participants in Head Start, the self-sufficiency index increased by an insignificant 3.4 percent of a standard deviation (ATET, column 6). For men, adjusting for pre-trends has little effect. Subcomponents of the index reveal meaningful changes in both the extensive and intensive margins of labor supply. The ATET of Head Start on male employment was a gain of 2.5 percentage points and 2.4 more hours per week (column 6). Although there is little evidence that Head Start decreased poverty for men, the program registers a sizable 42-percent reduction in public assistance receipt. This finding reinforces the suggestion that Head Start is shifting men from disability assistance into paid employment. Consistent with the negative selection implied by changes in men's employment, men participating in Head Start show little evidence of increases in their average wages, although the 95-percent confidence interval fails to rule out wage gains of over 7 percent. To investigate the implications of this negative selection further, we assume all new labor market entrants come from the least skilled part of the wage distribution. If we eliminate the lowest-earning individuals to offset the employment gains for men shown in Table 5, the NLSY shows that mean wages would have been as much as 9.2 log points higher.[30] Because we find a 2.7 log-point decrease in annual wages, this implies that Head Start led to an estimated 6.5-percent wage increase. Although we find no evidence of a concerning pre-trend for any outcome for men (column 3), the trend-break is not statistically significant for some outcomes (column 4). The noisiness of the estimates suggests some caution in drawing causal conclusions.

[30] We use the NLSY79 to examine how much log wages would change if we truncate the lowest-earning 2.7 percent of men who report attending Head Start as children. We obtain 2.7 percent by dividing the estimated ATET in Table 5A (column 6, "Worked last year") by the ATET plus the control-group mean employment rate. The idea behind this exercise is to reverse the selection into the sample caused by new labor market entrants, allowing us to examine how large the wage effects would be without this selection. The difference between our observed effect and the effect without lower truncation provides an estimate of Head Start's impact on mean wage earnings. If labor-market entrants were less negatively selected, the selection-corrected wage effects would be smaller, so this estimate can be viewed as an upper bound.

The pattern of estimates differs for women and the estimates are more sensitive to adjustments for pre-trends. Women's self-sufficiency index increased by 5.1 percent of a standard deviation without adjusting for off-setting pre-trends and 12 percent with this adjustment. Their employment in the last year was 5.4 percentage points higher among women participating in a fully implemented Head Start program, and the average participant worked an average of 2.7 more weeks per year and 2.9 more hours per week—an increase of roughly 196 annual hours.[31] Similarly, the estimates suggest that female participants in Head Start earned around 9.7 percent more, and their public assistance receipt declined by 10 percent, but neither effect is statistically significant at conventional levels (columns 6 and 7). As was the case for men, the trend-break is not statistically significant after accounting for multiple hypothesis testing, suggesting caution in the causal interpretation (column 4). The most striking changes for women occurred in their poverty rates, which were a highly statistically significant 4.4 percentage points, or 32 percent, lower among Head Start participants (columns 6 and 7).

Online Appendix Table A11 breaks these estimates down by race. Because the estimates for nonwhites are highly imprecise, we omit these estimates from the main paper. Our sample sizes are too small for our research design to draw conclusions.

A final set of estimates examines the effect of Head Start on adult incarceration—an important domain in economic self-sufficiency and also a domain where pre-school programs like Head Start have shown effects. For example, using a family fixed-effects design, Garces et al. (2002) estimated that Head Start decreased the probability of being booked or charged with a crime by 5 percentage points, or nearly 12 percentage points for African-Americans, although other studies using this research design have failed to replicate this finding (Deming, 2009; Miller et al., 2019). Johnson and Jackson (2019) suggest a decrease of roughly 3 percent in the share of Head Start attendees who were ever incarcerated. Barr and Gibbs (2017) also find intergenerational effects of Head Start on criminal history: a reduction of 15.6 percentage points for the female children of mothers who attended Head Start with less than a high school education. Research

[31] 2.7 additional weeks x [30.2 mean hours (column 1) +2.9 additional hours] + 2.9 additional hours x 36.9 weeks (column 1) ~196 annual hours.

has also found that the Perry Preschool program reduced criminal activity: Anderson (2008) found that by the teenage years, female students were 34 percentage points less likely to have a juvenile record, while estimates for men were statistically insignificant but too imprecise to rule out large effects.

Table 6 shows that our research design finds small, insignificant effects of Head Start on incarceration after adjusting for pre-trends (see estimates without this adjustment in Online Appendix section 10). The 95-percent confidence intervals for the ATETs (column 6) often do not contain estimates found in previous studies. However, we advise caution in interpreting this effect, because of limitations in the Census data. Importantly, previous studies have used longitudinal data or administrative crime data to measure the share of beneficiaries who were *ever* arrested or convicted of a crime. The only available measure of criminal activity in the Census/ACS is *incarceration* at the time of observation. Therefore, our data show that Head Start had very small, if any, effects on current incarceration and, potentially, longer spells of imprisonment. As for differences in effects from Perry Preschool students, this may reflect differences in the Head Start and Perry enrollees that were previously described.

VII. HETEROGENEITY IN HEAD START'S LONG-RUN EFFECT

This final section seeks to understand the mechanisms for Head Start's effects by examining how the estimates vary with access to other public programs and local economic conditions. Heterogeneity in Head Start's effects related to certain local programs or differences in community characteristics can shed light on how the program worked. However, we caution readers not to interpret these effects as causal, because other programs and community characteristics were not randomized. Our analysis interacts a binary indicator for whether cohorts lived in counties with "high" or "low" exposure to a program with the spline in equation (2), where "high" is equal to 1 for counties above the median in the characteristics and 0 otherwise. For parsimony and precision, we present only the human capital and economic self-sufficiency indices as the dependent variables. Additionally, we scale these effects by estimates of differential take-up to translate the ITT estimates into ATETs. Uncertainty about how program enrollment varied in these more disaggregated units means these estimates tend to be imprecise.

We first investigate the hypothesis that Head Start's long-run effects are driven by complementarities with other health programs for disadvantaged children. If health screening and referrals to health services (a sizable share of Head Start's budget) played a role in driving long-term effects, we would expect Head Start's effects to be larger for children with greater access to these services through community health centers (CHCs) and/or Medicaid.[32] Table 7 provides evidence consistent with this mechanism, showing that the ATETs of Head Start for human capital were more than three times as large in states where children had access to Medicaid (31 percent increase relative to 10 percent for less-exposed children). Head Start's ATETs on economic self-sufficiency are smaller in locations with greater eligibility for Medicaid, which is consistent with health services bringing more previously disabled workers into the labor market and the selection effects on wages for men in Table 5). As in Table 6, incarceration effects are too imprecisely estimated to find meaningful differences. On the other hand, greater access to CHCs is correlated with somewhat smaller effects, which suggests that Head Start may have substituted for CHC services for disadvantaged children, although the effects are not statistically different at conventional levels.

Head Start may also have affected adult outcomes by providing healthy meals and snacks, improving child nutrition which increased both health and learning. If nutrition is an important mechanism for Head Start's long-run effects, we would expect the program's effects to be *smaller* for children with greater access to Food Stamps, which provided a substitute for healthy meals at Head Start.[33] Consistent with healthy food mattering, Table 7 shows that children participating in Head Start with more access to Food Stamps experienced smaller—although still statistically significant—gains in human capital and economic self-sufficiency.

The OEO's larger effort to set up Head Start programs in the poorest 300 counties could also lead the Head Start program to be more intensive in these areas (Ludwig & Miller, 2007), leading the program to have larger treatment effects. For this test, we report the poorest 300 counties in the column for "above

[32] To construct high and low CHC exposure, we use data from Bailey and Goodman-Bacon (2015). We use the first year in which the program began in the county to construct the number of person-years an individual in each county and cohort would have been exposed to CHCs between ages 0 and 5. For Medicaid, we use Goodman-Bacon's (2018) data on the year states adopted Medicaid and the share of children covered by Aid for Families with Dependent Children (AFDC) in the year of Medicaid's launch. Because AFDC recipients were categorically eligible for Medicaid, coverage rates are strongly correlated with take-up of Medicaid in the early years. By combining the year of launch with the child AFDC rate for each county and cohort, we can obtain a measure of person-years of access to Medicaid.

[33] We use data from Almond, Hoynes, and Schanzenbach (2011). Access to Food Stamps is constructed in the same manner as access to CHCs.

median" and report the effects for counties outside this group in the column "below median." While Table 7 shows little evidence of differential effects on human capital across these two groups of counties, children in poorer counties benefitted more from the program in terms of their economic self-sufficiency—although this difference is imprecise.

A final hypothesis is that Head Start's effects should be larger in areas with greater subsequent economic growth. Strong economic growth would increase the resources of children's parents, expand the provision of public goods (such as schools), and create stronger incentives for children to invest in themselves, as children could expect higher and more certain returns. In addition, strong economic growth should be associated with more and better jobs. Rather than using actual economic growth (which may be endogenous to Head Start's effects on local human capital), we use predicted economic growth between 1965 and 1985.[34] The results in Table 7 show that the ATETs of Head Start for human capital were twice as large in areas with strong predicted economic growth than in areas with weaker predicted economic growth. The ATETs of Head Start for economic self-sufficiency were fifty percent larger in areas with strong predicted economic growth. In short, children benefitted the most from Head Start in areas with stronger economies, which likely created complementary public and family resources and jobs.

All in all, these results suggest that Head Start's long-run effects may be driven by many factors beyond a preschool curriculum, including health screenings and referrals and more nutritious meals for a population that otherwise may have been under-nourished and had little access to health care. The effects of Head Start also appear to be complementary to the family and public resources arising from a stronger economy.

VIII. NEW EVIDENCE ON THE LONG-TERM RETURNS TO HEAD START

Over the past 20 years, substantial evidence has accumulated that *model* preschool programs have sizable economic returns (Almond & Currie, 2011; Cunha & Heckman, 2007; Duncan & Magnuson, 2013;

---

[34] We use data from the Bureau of Economic Analysis (BEA) and County Business Patterns (CBP). We create a county-level panel of log real earnings from 1965 through 1985. We then regress growth in real earnings between 1965 and 1985 on a number of county characteristics from the 1960 Census: log of total population, share of a county in farmland, share of the population living in an urban setting, share black, share under age 5, share over age 65, and share living in poverty. We use predicted growth from this regression to select counties that were likely to have high economic growth from 1965-1985 and those that were not.

Heckman, Moon, Pinto, Savelyev, & Yavitz, 2010). However, convincing evidence on the long-run returns to larger-scale, public preschool has remained sparse (Phillips et al., 2017).

Using large-scale restricted Census/ACS data, this paper provides new evidence on the long-term effects of Head Start, the nation's longest-running, large-scale public preschool program. We find that Head Start had large effects on participants' human capital. Head Start children were 2.7 percent more likely to complete high school and 8.5 percent more likely to enroll in college. Their college completion rates rose by 12 percentage points. A second finding is that Head Start increased the economic self-sufficiency of adults, reducing the incidence of adult poverty by 23 percent and public assistance receipt by 27 percent. Heterogeneity tests suggest that these long-run effects reflect many aspects of the Head Start program beyond its academic curriculum: health screenings and referrals and more nutritious meals appear to be important mechanisms for the program's effects on disadvantaged children. In addition, a stronger economy increased Head Start's effect, likely due to its complementarity with growth in family and community resources as well as on the availability of jobs.

A full accounting of the costs and benefits of Head Start is beyond the scope of this paper, but we use data from the NLSY79 to summarize the implications of our estimates for program participants. We follow Deming (2009) and Neal and Johnson (1996) and use *potential* earnings rather than actual earnings, motivated by our evidence that Head Start affected employment and, therefore, observed wages via selection. We regress log earnings on components of our human capital and economic self-sufficiency indices for individuals born from 1957 to 1965 (ages 14 to 22 in 1979), a time frame that overlaps our Census/ACS analysis. We also flexibly control for ability using the AFQT. (Although AFQT is not available in the Census/ACS, using this as a covariate helps mitigate omitted variables bias in ability in the education and earnings relationship.) We then combine these regression estimates with our estimated treatment effects of Head Start to calculate the effect of Head Start on potential earnings. The NLSY79 suggest a private internal rate of return to Head Start of 13.7 percent. In addition, Head Start generates a fiscal externality due to savings on public assistance expenditures (estimated at $9,967 per public-program beneficiary in the Survey of Income and Program Participation) and additional tax revenue generated from wage gains (estimated at $576 to $2341 annually per Head Start participant using data from the NLSY and the National Bureau of Economic Research's Taxsim model[35]). While these figures do not capture all of Head Start's impacts on public expenditures, they suggest the internal rate of return to the government from these factors is 5.4 to 9.1 percent. The size of the fiscal externality varies with assumptions about who the marginal beneficiary is, but the bottom line is that Head Start easily pays for itself and generates sizeable returns when taking account of its long-run effects. (See Online Appendix section 11 for more details, regression results, and discussion.)

These estimates are likely to be conservative for several reasons. First, our research design differences out spill-over effects on siblings age six and older, which tends to reduce the estimated effect sizes. Second, reports of income and public assistance receipt may be severely underreported in major national surveys (Bound, Brown, & Mathiowetz, 2001; Meyer, Mok, & Sullivan, 2015), suggesting estimates of Head Start's effect on public assistance may be understated. Third, our estimates may understate the savings accruing through reduced reliance of social safety-net programs if Head Start participants were less likely to receive other public assistance benefits such as Medicaid or the Medicare coverage that often accompanies Social Security Disability Insurance, which are not in our data (Autor & Duggan, 2006). Finally, estimates of the returns to Head Start neglect improvements in outcomes not measured here. For instance, they ignore the extent to which more education engenders better health, longevity, or well-being not captured in wages or employment. An analysis that included these additional outcomes would tend to strengthen the conclusion that Head Start achieved its goal of reducing adult poverty and delivered sizable returns to public dollars spent in the 1960s and 1970s, with potentially even larger social returns.

The long-run returns to today's public preschool programs may be different for a number of reasons. The curriculum is different, the target population is different, and the alternative programs and resources available to poor children are radically different than in the past. Of course, researchers will need to wait another 50 years to evaluate the long-run effects of today's preschool programs. In the meantime,

[35] https://taxsim.nber.org/

the sizable returns to the "less-than-model" Head Start preschool program of the 1960s suggest productive avenues for improving the lives of disadvantaged children today.

IX. References

Almond, D., & Currie, J. (2011). Human capital development before age five. *Handbook of labor economics, 4*, 1315-1486.

Almond, D., Hoynes, H. W., & Schanzenbach, D. W. (2011). Inside the War on Poverty: The Impact of Food Stamps on Birth Outcomes. *Review of Economics and Statistics, 93*(2), 387-403.

Anderson, M. A. (2008). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association, 103*(484), 1481-1491.

Arellano, M. (1987). Computing Robust Standard Errors for Within-Groups Estimators. *Oxford Bulletin of Economics and Statistics, 49*(4), 431-434. Retrieved from http://www.blackwellpublishing.com.proxy.lib.umich.edu/journal.asp?ref=0305-9049

Autor, D. H., & Duggan, M. G. (2006). The Growth in the Social Security Disability Rolls: A Fiscal Crisis Unfolding. *Journal of Economic Perspectives, 20*(3), 71-96. doi:10.1257/jep.20.3.71

Bailey, M. J. (2012). Reexamining the Impact of U.S. Family Planning Programs on Fertility: Evidence from the War on Poverty and the Early Years of Title X. *American Economic Journal: Applied Economics, 4*(2), 62-97.

Bailey, M. J., & Danziger, S. (2013). The Legacies of the War on Poverty. In M. J. Bailey & S. Danziger (Eds.), *Legacies of the War on Poverty*. New York: Russell Sage Foundation.

Bailey, M. J., & Duquette, N. J. (2014). How Johnson Fought the War on Poverty: The Economics and Politics of Funding at the Office of Economic Opportunity. *Journal of Economic History, 74*(2), 351-388. Retrieved from www-personal.umich.edu/~baileymj/Bailey-Duquette.pdf

Bailey, M. J., & Goodman-Bacon, A. J. (2015). The War on Poverty's Experiment in Public Medicine: The Impact of Community Health Centers on the Mortality of Older Americans. *American Economic Review, 105*(3), 1067-1104.

Bailey, M. J., Hoynes, H. W., Rossin-Slater, M., & Walker, R. (2020). Is the Social Safety Net a Long-Term Investment? Large-Scale Evidence from the Food Stamps Program. *NBER Working Paper 26942*.

Bailey, M. J., Malkova, O., & McLaren, Z. (2018). Do Family Planning Programs Increase Children's Opportunities? Evidence from the War on Poverty and the Early Years of Title X. *Journal of Human Resources*. Retrieved from http://www-personal.umich.edu/~baileymj/Bailey_Malkova_McLaren.pdf

Barnett, W., & Masse, L. N. (2007). Comparative benefit-cost analysis of the Abecedarian program and its policy implications. *Economics of Education Review, 26*(1), 113-125. Retrieved from https://EconPapers.repec.org/RePEc:eee:ecoedu:v:26:y:2007:i:1:p:113-125

Barr, A., & Gibbs, C. R. (2017). *Breaking the Cycle? Intergenerational Effects of an Anti-Poverty Program in Early Childhood*. Retrieved from https://curry.virginia.edu/sites/default/files/files/EdPolicyWorks_files/61_Anti_Proverty_Effects_ECE.pdf

Bassok, D., Fitzpatrick, M., & Loeb, S. (forthcoming). Does State Preschool Crowd-Out Private Provision? The Impact of Universal Pre-Kindergarten on the Childcare Sector in Oklahoma and Georgia. *Journal of Urban Economics*.

Bauer, L., & Schanzenbach, D. W. (2016). The Long-Term Impact of the Head Start Program. *Hamilton Project - Brookings*. Retrieved from https://www.hamiltonproject.org/assets/files/long_term_impact_of_head_start_program.pdf

Bedard, K., & Dhuey, E. (2006). The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects *Quarterly Journal of Economics, 121*(4), 1437-1472.

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-in-Differences Estimates? *Quarterly Journal of Economics, 119*(1), 249-275. Retrieved from http://www.mitpressjournals.org/loi/qjec

Black, D. A., Sanders, S. G., Taylor, E. J., & Taylor, L. J. (2015). The Impact of the Great Migration on Mortality of African Americans: Evidence from the Deep South. *American Economic Review, 105*(2), 477-503.

Bloom, B. S. (1964). *Stability and Change in Human Characteristics*. New York: Wiley.

Borusyak, K., & Jaravel, X. (2018). Revisiting Event Study Designs. *SSRN Working Paper*. Retrieved from http://dx.doi.org/10.2139/ssrn.2826228

Bound, J., Brown, C., & Mathiowetz, N. (Eds.). (2001). *Measurement Error in Survey Data* (Vol. 5). Amsterdam: Elsevier.

Bound, J., & Solon, G. (1999). Double Trouble: On the Value of Twins-Based Estimation of the Return to Schooling *Economics of Education Review, 18*(2), 169-182.

Braun, S. J., & Edwards, E. P. (1972). *History and Theory of Early Childhood Education*. Belmont, CA: Wadsworth.

Callaway, B., & Sant'Anna, P. H. C. (2019). Difference-in-Differences with Multiple Time Periods. *SSRN Working Paper*. Retrieved from http://dx.doi.org/10.2139/ssrn.3148250

Cameron, A. C., Gelbach, J. B., & Miller, D. L. (2011). Robust Inference With Multiway Clustering. *Journal of Business & Economic Statistics, 29*(2), 238-249. doi:10.1198/jbes.2010.07136

Carneiro, P., & Ginja, R. (2014). Long-Term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start. *American Economic Journal: Economic Policy, 6*(4), 135-173. doi:10.1257/pol.6.4.135

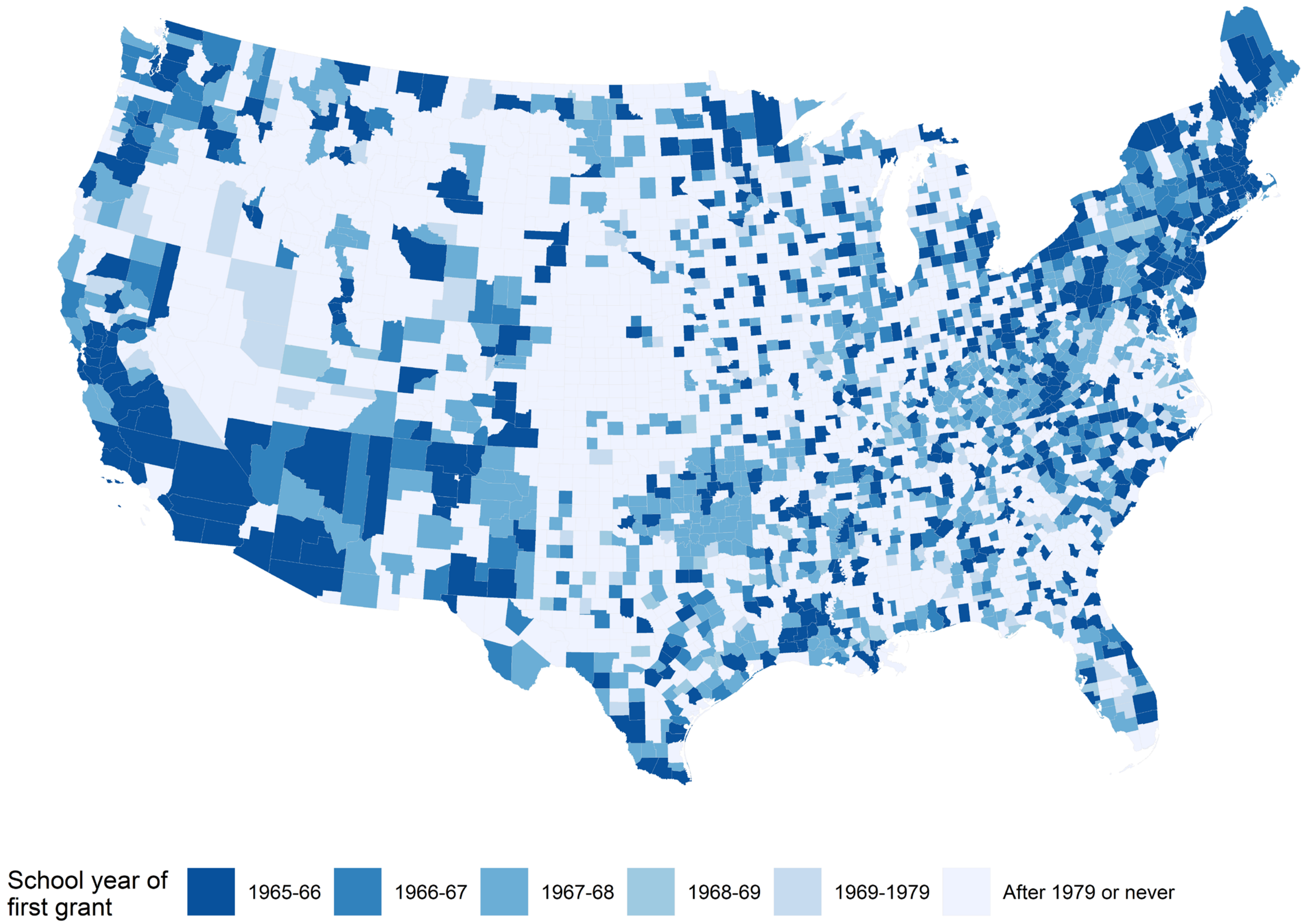
Cascio, E., Gordon, N., Lewis, E., & Reber, S. J. (2010). Paying for Progress: Conditional Grants and the Desegregation of Southern Schools. *Quarterly Journal of Economics, 125*(1), 445-482.

Cascio, E. U. (2009). Maternal Labor Supply and the Introduction of Kindergartens into American Public Schools. *Journal of*

Human Resources, 44(1), 140-170.
Cascio, E. U., & Schanzenbach, D. W. (2013). The Impacts of Expanding Access to High-Quality Preschool Education. *Brookings Papers on Economic Activity*, 127-178.
Cunha, F., & Heckman, J. (2007). The Technology of Skill Formation. *American Economic Review, 97*, 31-47.
Currie, J. (2001). Early Childhood Education Programs. *Journal of Economic Perspectives, 15*(2), 213-238.
Currie, J., & Thomas, D. (1995). Does Head Start Make a Difference? . *American Economic Review, 85*(3), 341-364.
de Chaisemartin, C., & D'Haultfœuille, X. (2020). Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. *American Economic Review, 110*(9), 2964-2996. doi:10.1257/aer.20181169
De Haan, M., & Leuven, E. (2020). Head Start and the Distribution of Long-Term Education and Labor Market Outcomes. *Journal of Labor Economics, 38*(3), 727-765. Retrieved from https://EconPapers.repec.org/RePEc:ucp:jlabec:doi:10.1086/706090
Deming, D. (2009). Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start *American Economic Journal: Applied Economics, 1*(3), 111-134.
Duflo, E., Glennerster, R., & Kremer, M. (2007). Chapter 61 Using Randomization in Development Economics Research: A Toolkit. In T. P. Schultz & J. A. Strauss (Eds.), *Handbook of Development Economics* (Vol. 4, pp. 3895-3962): Elsevier.
Duncan, G. J., & Magnuson, K. (2013). Investing in Preschool Programs. *Journal of Economic Perspectives, 27*(2), 109-132. Retrieved from http://www.aeaweb.org/articles?id=10.1257/jep.27.2.109
Efron, B., & Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*. New York: Chapman & Hall.
Espinosa, L. M. (2002). High-Quality Preschool: Why We Need It and What it Looks Like. *NIEER Preschool Policy Matters, 1*. Retrieved from http://nieer.org/wp-content/uploads/2016/08/1.pdf
Fosburg, L. B., Goodrich, N. N., Fox, M. K., Granahan, P., Smith, J., Himes, J. H., & Weitzman, M. (1984). The Effects of Head Start Health Services: Report of the Head Start Health Evaluation. In *Report Prepared for the Administration for Children, Youth and Families, U. S. Department of Health and Human Services*. Cambridge, MA: Abt Associates.
Freyaldenhoven, S., Hansen, C., & Shapiro, J. M. (2019). Pre-event Trends in the Panel Event-Study Design. *American Economic Review, 109*(9), 3307-3338. doi:10.1257/aer.20180609
Frisvold, D. E. (2015). Nutrition and cognitive achievement: An evaluation of the School Breakfast Program. *Journal of Public Economics, 124*, 91-104. doi:https://doi.org/10.1016/j.jpubeco.2014.12.003
Garces, E., Thomas, D., & Currie, J. (2002). Longer Term Effects of Head Start. *American Economic Review, 92*(4), 999-1012.
Gibbs, C., Ludwig, J., & Miller, D. L. (2014). Does Head Start Do Any Lasting Good? In M. J. Bailey & S. Danziger (Eds.), *Legacies of the War on Poverty* (pp. 39-65). New York, NY: Russell Sage Foundation.
Goodman-Bacon, A. J. (2018). Difference-in-Differences with Variation in Treatment Timing. *NBER Working Paper 25018*. Retrieved from https://www.nber.org/papers/w25018
Griliches, Z. (1979). Sibling Models and Data in Economics: Beginnings of a Survey *Journal of Political Economy, 87*(5, Part 2), S37-S64.
Hechinger, F. M. (Ed.) (1966). *Pre-School Education Today: New Approaches to Teaching Three-, Four-, and Five-Year-Olds*. New York: Doubleday.
Heckman, J. J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The Rate of the Return to the High Scope Perry Preschool Program *Journal of Public Economics, 94*(1), 114-128.
Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics, 6*(2), 65-70. Retrieved from http://www.jstor.org/stable/4615733
Hoynes, H. W., Page, M. E., & Stevens, A. H. (2011). Can Targeted Transfers Improve Birth Outcomes? Evidence from the Introduction of the WIC Program. *Journal of Public Economics, 95*(7), 813-827.
Hoynes, H. W., & Schanzenbach, D. W. (2009). Consumption Responses to In-Kind Transfers: Evidence from the Introduction of the Food Stamp Program. *American Economic Journal: Applied Economics, 1*(4), 109-139.
Hoynes, H. W., Schanzenbach, D. W., & Almond, D. (2016). Long Run Impacts of Childhood Access to the Safety Net. *American Economic Review, 106*(4), 903-934.
Hunt, J. M. (1961). *Intelligence and Experience*. New York: Ronald Press.
Isen, A., Rossin-Slater, M., & Walker, W. R. (2013). Every Breath You Take-- Every Dollar You'll Make: The Long-Term Consequences of the Clean Air Act of 1970. *NBER Working Paper 19858*.
Johnson, R., & Jackson, C. K. (2019). Reducing Inequality through Dynamic Complementarity: Evidence from Head Start and Public School Spending. *American Economic Journal - Economic Policy*, 310-349.
Johnston, J., & DiNardo, J. E. (1997). *Econometric Methods*. New York: McGraw-Hill.
Kline, P., & Walters, C. (2016). Evaluating Public Programs with Close Substitutes: The Case of Head Start. *Quarterly Journal of Economics, 131*(4), 1795-1848.
Kling, J. R., Liebman, J. B., & Katz, L. F. (2007). Experimental analysis of neighborhood effects. *Econometrica, 75*(1), 83-119.
Levine, R. A. (1970). *The Poor Ye Need Not Have with You: Lessons from the War on Poverty*. Cambridge: MIT Press.
Levitan, S. A. (1969). *The Great Society's Poor Law: A New Approach to Poverty*. Baltimore: Johns Hopkins Press.
Loeb, S. (2016). Missing the Target: We Need to Focus on Informal Care Rather than Pre-School. *Economic Studies at Brookings, 1*(19).
Ludwig, J., & Miller, D. L. (2007). Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design. *Quarterly Journal of Economics, 122*(1), 159-208.
Meyer, B. D., Mok, W. K. C., & Sullivan, J. X. (2015). Household Surveys in Crisis. *Journal of Economic Perspectives, 29*(4).
Miller, D. L., Grosz, M. Z., & Shenhav, N. a. (2019). Selection into Identification in Fixed Effects Models, with Application to Head Start. *NBER Working Paper 26174*.

Neal, D., & Johnson, W. (1996). The Role of Premarket Factors in Black-White Wage Differences. *Journal of Political Economy, 104*(5), 869-895. Retrieved from https://EconPapers.repec.org/RePEc:ucp:jpolec:v:104:y:1996:i:5:p:869-95
North, A. F. (1979). Health Services in Head Start. In E. Zigler & J. Valentine (Eds.), *Project Head Start: A Legacy of the War on Poverty*. New York: Free Press.
OEO. (1965). *1st Annual Report: A Nation Aroused*.
OEO. (1966). *2nd Annual Report: The Quiet Revolution*.
OEO. (1967). *3rd Annual Report: The Tide of Progress*.
OEO. (1968). *4th Annual Report: As the Seed is Sown*.
OEO. (1970). *Annual Report Fiscal Years 1969-70*.
Phillips, D. A., Lipsey, M. W., Dodge, K. A., Haskins, R., Bassok, D., Burchinal, M. R., . . . Christina, W. (2017). *The Current State of Scientific Knowledge on Pre-Kindergarten Effects*. The Brookings Institution. Washington, DC. Retrieved from https://www.brookings.edu/wp-content/uploads/2017/04/duke_prekstudy_final_4-4-17_hires.pdf
Rambachan, A., & Roth, J. (2020). An Honest Approach to Parallel Trends. *Harvard University Working Paper*. Retrieved from https://scholar.harvard.edu/files/jroth/files/roth_jmp_honestparalleltrends_main2.pdf
Ruggles, S., Genadek, K., Grover, J., & Sobek, M. (2015). *Integrated Public Use Microdata Series (Version 6.0) [Machine-Readable database]*. Retrieved from: http://doi.org/10.18128/D010.V6.0
Solon, G., Haider, S. J., & Wooldridge, J. M. (2015). What are We Weighting For? *Journal of Human Resources, 50*(2), 301-316.
Taylor, E. J., Stuart, B., & Bailey, M. (2016). Summary of Procedure to Match Numident Place of Birth County to GNIS Places. *U.S. Census Bureau, 1284 Technical Memo #2, May 6*.
Thompson, O. (2018). Head Start's Long Run Impact: Evidence form the Program's Introduction. *Journal of Human Resources, 53*, 1100-1139. Retrieved from https://drive.google.com/file/d/0B9lVhfNKAH86YWh1bGlWbDc5c2M/view
U.S. Office of Child Development, R. a. E. D. (1972). Project Head Start, 1969-1970 : A Descriptive Report of Programs and Participants. In. Washington, D.C.: Government Printing Office.
U.S. Senate Committee on Labor and Public Welfare. (1964). *Economic Opportunity Act of 1964*. Washington, DC: U.S. Government Printing Office
Westinghouse Learning Corporation. (1969). *The Impact of Head Start: An Evaluation of the Effects of Head Start on Children's Cognitive and Affective Development (Executive Summary)*. Washington, D.C.: Office of Economic Opportunity, Retrieved from https://eric.ed.gov/?id=ED036321

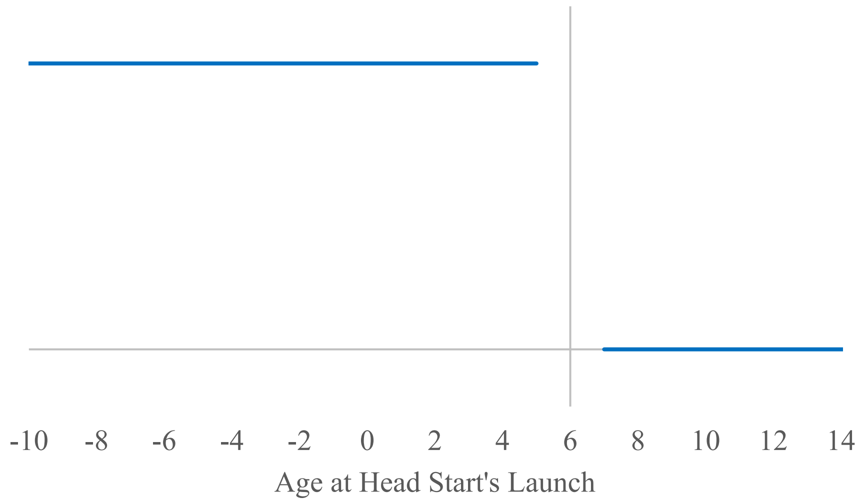
Figure 1. The Launch of Head Start between 1965 and 1980



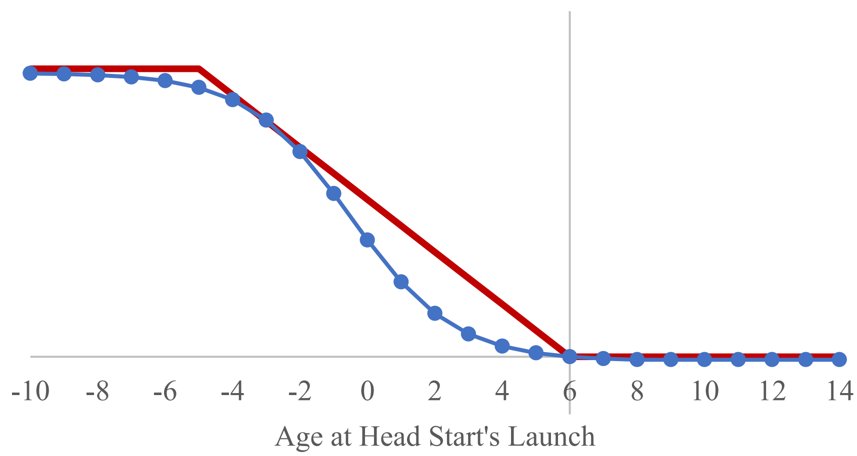
Notes: Counties are grouped by the fiscal year that Head Start launched between 1965 and 1980. Data on federal grants are drawn from the National Archives and Records Administration (NARA). See Bailey and Duquette (2014) and Bailey and Goodman-Bacon (2015) for details on data and variable construction.

Figure 2. The Expected Pattern of Head Start's Effects on Adult Outcomes under Different Assumptions

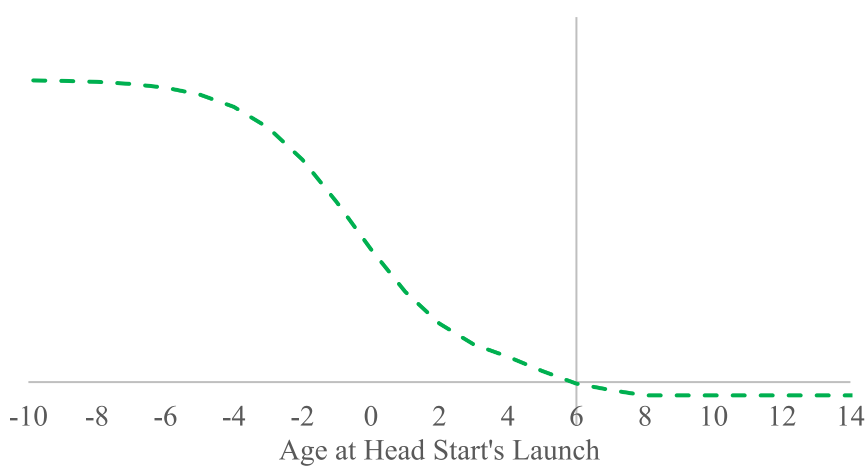
A. Constant Intention-to-Treat Effect: No Growth in Program Capacity or Quality; No Sibling Spill-Overs



B. Increasing Intention-to-Treat Effect: Growth in Program Capacity and Quality; No Sibling Spill-Overs

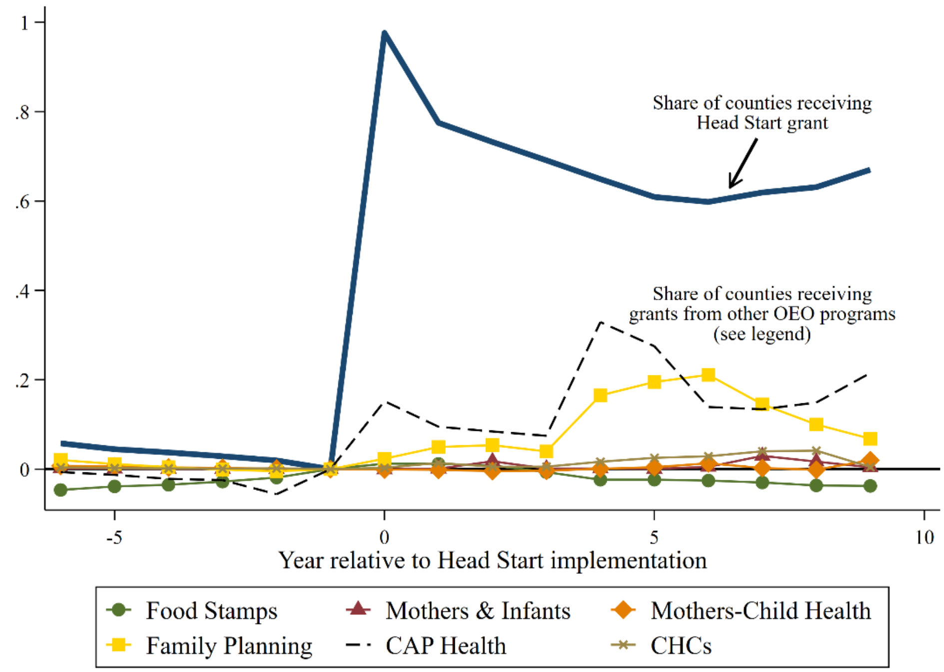


C. Increasing Intention-to-Treat Effect: Growth in Program Capacity and Quality & Sibling Spill-Overs



Notes: Figures show hypothetical effects of Head Start by cohort's age (on the date of the school age entry cutoff) when Head Start launched in their county. See text for discussion.

Figure 3. Funding for Other OEO Programs Relative to the Year Head Start Began



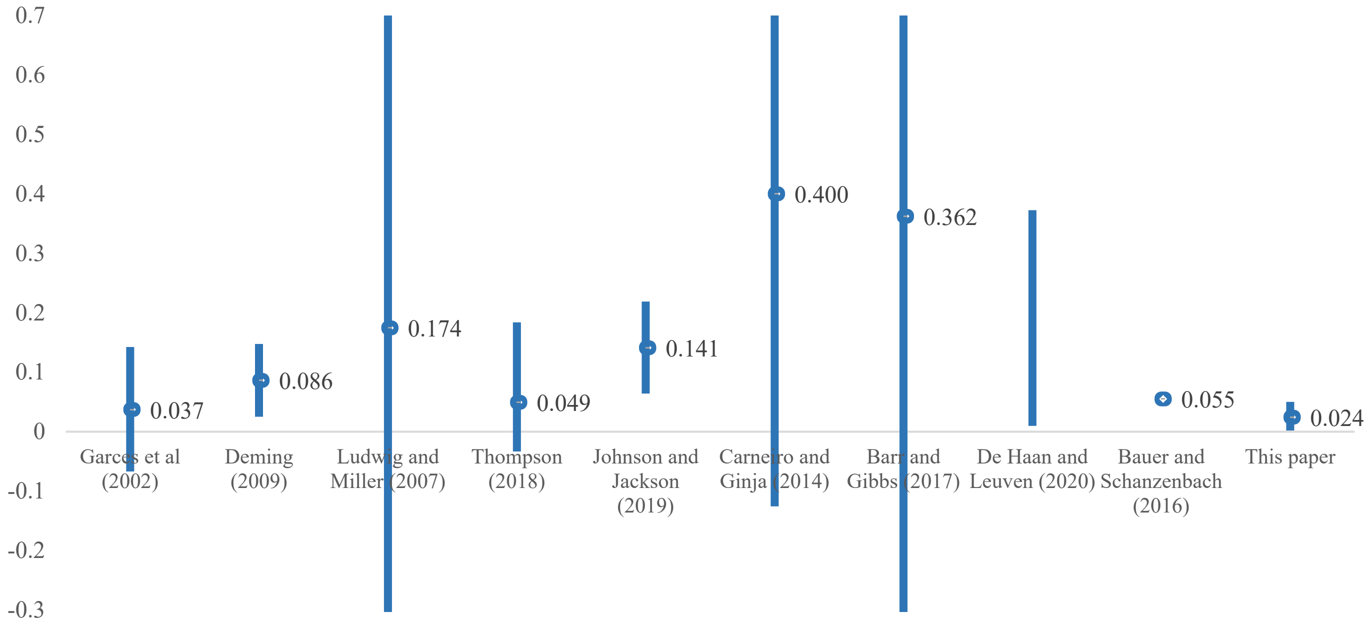
Notes: Dependent variables are binary variables for whether a county received a grant for the indicated program in the indicated year. Data on federal grants and programs are drawn from the NARA.

Figure 4. The Effect of Head Start on Adult Human Capital

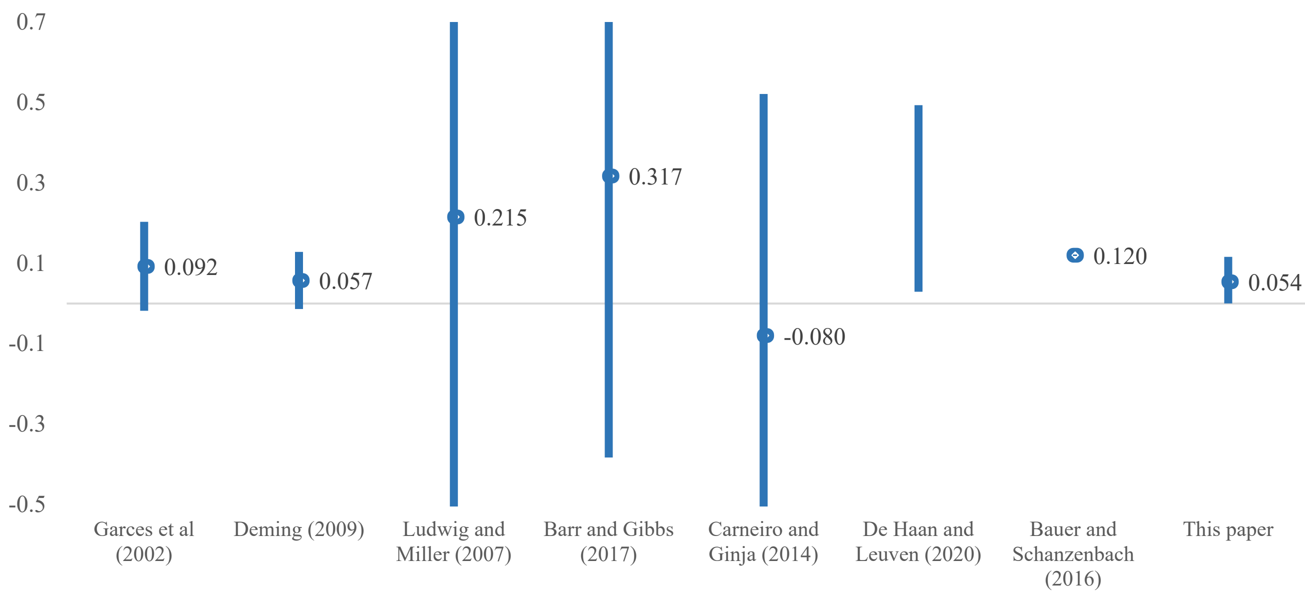
Notes: The figures plot event-study estimates of φ for different outcomes using the specification in equation (1). Long-dashed lines show predicted values from the spline specification in equation (2). Short-dashed lines show 95-percent, point-wise confidence intervals for each event-study estimate. See text for more details.

Figure 5. The Magnitude of Head Start's Effects on Education across Studies

A. Effects of Head Start on High School Graduation

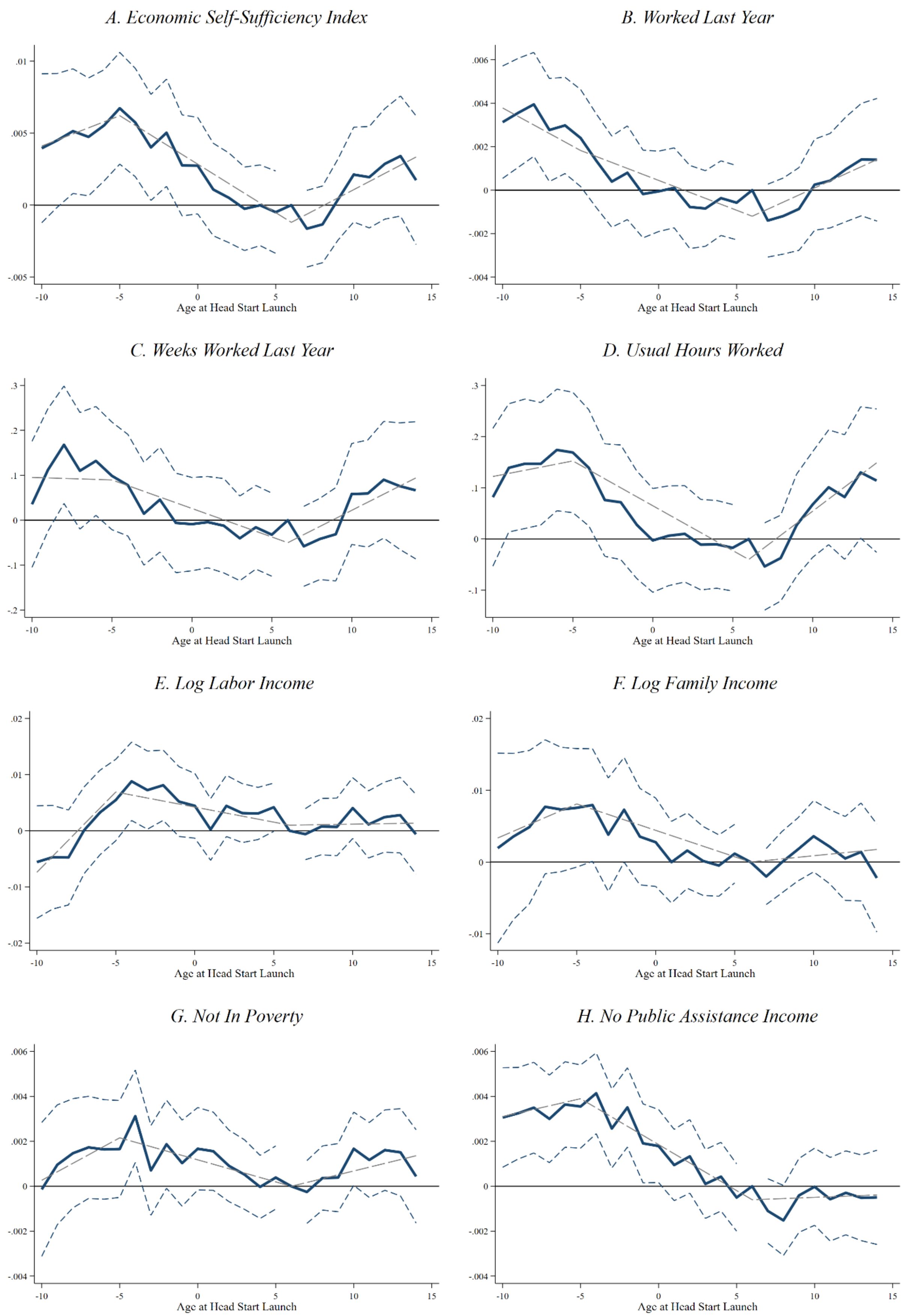


B. Effects of Head Start on College Enrollment



Notes: Circles indicate the reported or derived ATET from different studies. For sibling fixed effect studies, the ATET is directly reported in the papers. For studies reporting an ITT effect, estimates have been converted to an ATET by dividing by the reported first-stage estimate. Bars indicate the 95-percent confidence interval as reported for sibling fixed-effect models or as constructed for the ITT studies using a parametric bootstrap procedure using 10,000 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 & Tibshirani, 1993). Because Johnson and Jackson (2019) do not report a standard error on the first stage, the confidence interval reported for this study in Panel A does not include this first-stage uncertainty. Bauer and Schanzenbach (2016) do not report standard errors, so confidence intervals are omitted from the figure. 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. The confidence intervals for Carneiro and Ginja (2014) are [-0.12,1.21] in panel A and [-0.74,0.52] in panel B. The confidence interval for Barr and Gibbs (2017) is [-0.38,2.08] in panel B. Johnson and Jackson (2019) and Thompson (2018) sample likely eligible samples of the PSID and NLSY79: individuals born to parents in the bottom quartile of the income distribution, and parents with no college education, respectively. Barr and Gibbs (2017) estimates derived from a sample of women with mothers without a high school diploma.

Figure 6. The Effect of Head Start on Adult Economic Self-Sufficiency



Notes: See Figure 4 notes.

Table 1. The Effect of Head Start on Adult Human Capital

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Control mean (s.d.)	ITT estimate (s.e.) [BH p-val]	Slope of pre-trend[1] (s.e.) [BH p-val]	Test of trend break at age 6 (p-val) [BH p-val]	Slope of post-trend[1] (s.e.) [BH p-val]	ATET [95% CI]	ATET % change
Human capital		2.1	0.051	24	0.010	14	
		(0.35)	(0.040)	(<0.001)	(0.075)	[8.6,21]	
Human capital (pre-trend adjusted)		2.7			-0.040	18	
		(0.54)			(0.082)	[9.7,28]	
Subindices (pre-trend adjusted)							
Completed high school/GED	92	0.36	0.0039	4.3	-0.016	2.4	2.7%
	(28)	(0.17)	(0.011)	(0.037)	(0.029)	[0.40,5.2]	
		[0.076]	[1.0]	[0.075]	[1.0]		
Attended some college	64	0.80	-0.014	3.9	0.000	5.4	8.5%
	(48)	(0.41)	(0.027)	(0.048)	(0.055)	[0.092,11]	
		[0.076]	[1.0]	[0.075]	[1.0]		
Completed college	30	1.7	0.030	24	-0.026	12	39%
	(46)	(0.36)	(0.027)	(<0.001)	(0.052)	[6.7,19]	
		[<0.001]	[0.81]	[<0.001]	[1.0]		
Prof. or doc. degree	2.9	0.39	0.016	13	-0.0049	2.6	91%
	(17)	(0.11)	(0.0075)	(<0.001)	(0.012)	[1.2,4.6]	
		[0.0010]	[0.20]	[0.0010]	[1.0]		
Years of schooling	14	0.096	0.0023	25	-0.0014	0.65	4.8%
	(2.5)	(0.019)	(0.0014)	(<0.001)	(0.0029)	[0.40,1.01]	
		[<0.001]	[0.58]	[<0.001]	[1.0]		
Has a professional job	35	1.4	0.033	21	-0.020	9.5	27%
	(48)	(0.30)	(0.022)	(<0.001)	(0.038)	[5.3,15]	
		[<0.001]	[0.58]	[<0.001]	[1.0]		

Notes: For the index and all binary outcome variables, means, standard deviations, point estimates, and standard errors in percentage-point units. In column 1, the control mean and standard deviation are calculated using the cohorts from Head Start counties who were ages 6 and 7 at the time the program was launched in the county. All estimates in columns 2-5 come from spline specification (equation 2). Column 2 presents the estimated intention-to-treat (ITT) effect evaluated for those with full exposure to a fully implemented program. Column 3 presents the spline pre-trend estimate (children ages 6 and older when Head Start was implemented). Column 4 presents the F-statistic and p-value for the test of a trend-break in the spline for children age 6 when Head Start was implemented. Column 5 presents an estimate of the spline post-trend slope (children ages -5 through -10 when Head Start was implemented). The ATET estimate in column 6 divides the ITT effect in column 2 by the estimated effect of receiving a Head Start grant on school enrollment (0.149, s.e. 0.022) and uses a bootstrap to obtain confidence intervals (see discussion in text). Column 7 computes the percentage increase implied by the ATET relative to the control mean (the ratio of column 6 to column 1) for components of the index. The BH p-values presented in columns 2-5 in brackets use the Bonferroni-Holm method to account for multiple hypothesis testing of individual outcomes within an index. [1] Pre-trend-adjusted estimates are constructed by subtracting the estimated pre-trend (column 3) from the slope of the middle (column 2) and post-trend (column 5) segments of the spline. For the subindices and pre-trend adjusted index, the "post-trend" estimates in column 5 have been adjusted for the pre-trend.

Table 2. The Effect of Head Start on Adult Human Capital, by Sex

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Control mean (s.d.)	ITT estimate (s.e.) [BH p-val]	Slope of pre-trend[1] (s.e.) [BH p-val]	Test of trend break at age 6 (p-val) [BH p-val]	Slope of post-trend[1] (s.e.) [BH p-val]	ATET [95% CI]	ATET % change
A. Men							
Human capital		2.3	0.021	15	0.034	15	
		(0.39)	(0.047)	(<0.001)	(0.076)	[9.4,23]	
Human capital (pre-trend adjusted)		2.5			0.014	17	
		(0.65)			(0.088)	[7.7,28]	
Subindices (pre-trend adjusted)							
Completed high	91	0.42	0.0056	3.2	-0.0089	2.8	3.1%
school/GED	(29)	(0.23)	(0.015)	(0.072)	(0.033)	[0.06,6.4]	
		[0.10]	[1.0]	[0.10]	[1.0]		
Attended some college	61	1.0	-0.0045	3.8	0.0031	6.6	11%
	(49)	(0.51)	(0.034)	(0.050)	(0.062)	[0.023,14]	
		[0.10]	[1.0]	[0.10]	[1.0]		
Completed college	29	1.6	0.013	15	-0.0079	11	37%
	(46)	(0.43)	(0.032)	(<0.001)	(0.057)	[5.1,19]	
		[<0.001]	[1.0]	[<0.001]	[1.0]		
Prof. or doc. degree	3.4	0.48	0.016	9.2	0.0047	3.2	95%
	(18)	(0.16)	(0.011)	(0.0025)	(0.017)	[1.2,5.8]	
		[0.0095]	[0.91]	[0.010]	[1.0]		
Years of schooling	14	0.090	0.0014	16	0.00011	0.60	4.4%
	(2.5)	(0.023)	(0.0017)	(<0.001)	(0.0031)	[0.28,1.0]	
		[<0.001]	[1.0]	[<0.001]	[1.0]		
Has a professional job	33	0.98	-0.024	6.6	0.045	6.5	19%
	(47)	(0.38)	(0.027)	(0.010)	(0.047)	[1.7,12]	
		[0.031]	[1.0]	[0.031]	[1.0]		
B. Women							
Human capital		2.0	0.071	21	-0.03	14	
		(0.38)	(0.043)	(<0.001)	(0.089)	[8.5,22]	
Human capital (pre-trend adjusted)		2.8			-0.10	19	
		(0.60)			(0.098)	[11,31]	
Subindices (pre-trend adjusted)							
Completed high	93	0.39	0.0068	4.3	-0.028	2.7	2.9%
school/GED	(26)	(0.19)	(0.013)	(0.037)	(0.032)	[0.15,5.7]	
		[0.074]	[0.68]	[0.075]	[1.0]		
Attended some college	67	0.56	-0.031	1.8	0.00060	3.9	5.8%
	(47)	(0.42)	(0.028)	(0.18)	(0.062)	[-2.6,11]	
		[0.18]	[0.68]	[0.18]	[1.0]		
Completed college	31	1.8	0.035	19	-0.039	12	39%
	(46)	(0.41)	(0.029)	(<0.001)	(0.064)	[6.3,20]	
		[<0.001]	[0.68]	[<0.001]	[1.0]		
Prof. or doc. degree	2.5	0.30	0.015	8.4	-0.019	2.1	85%
	(15)	(0.11)	(0.0069)	(0.0039)	(0.013)	[0.68,3.9]	
		[0.012]	[0.15]	[0.012]	[0.69]		
Years of schooling	14	0.10	0.0028	22	-0.0031	0.70	5.1%
	(2.4)	(0.0215)	(0.0015)	(<0.001)	(0.0035)	[0.4,1.1]	
		[<0.001]	[0.26]	[<0.001]	[1.0]		
Has a professional job	37	1.8	0.083	25	-0.080	13	34%
	(48)	(0.37)	(0.025)	(<0.001)	(0.047)	[7.4,20]	
		[<0.001]	[0.006]	[<0.001]	[0.53]		

Notes: Pre-trend-adjusted estimates are constructed by subtracting the estimated pre-trend (column 3) from the slope of the middle (column 2) and post-trend (column 5) segments of the spline. [1] For the subindex estimates, the "post-trend" estimates in column 5 have been adjusted for the pre-trend estimate shown in column 3. Column 6 scales by an estimated take up of 0.151 (s.e. 0.022) for men and 0.145 (s.e. 0.022) for women (Appendix Table A5). See also Table 1 notes.

Table 3. The Effect of Head Start on Adult Human Capital, by Race

	(1) Control mean (s.d.)	(2) ITT estimate (s.e.)	(3) Slope of pre-trend[1] (s.e.)	(4) Test of trend break at age 6 [p-val]	(5) Slope of post-trend[1] (s.e.)	(6) ATET [95% CI]	(7) ATET % change
A.White							
Human capital		2.2	0.031	21	-0.007	16	
		(0.39)	(0.040)	[<0.001]	(0.079)	[9.4,26]	
Human capital (pre-trend adjusted)		2.5			-0.038	19	
		(0.54)			(0.079)	[9.7,31]	
Subindices (pre-trend adjusted)							
Completed high	93	0.25	-0.0070	2.3	-0.013	1.9	2.0%
school/GED	(26)	(0.17)	(0.010)	[0.13]	(0.029)	[-0.33,4.9]	
Attended some college	65	0.77	-0.018	3.6	-0.0018	5.8	9.0%
	(48)	(0.40)	(0.026)	[0.057]	(0.053)	[-0.14,13]	
Completed college	31	1.7	0.023	20	-0.025	13	40%
	(46)	(0.37)	(0.029)	[<0.001]	(0.049)	[6.8,23]	
Prof. or doc. degree	3.0	0.38	0.013	11	0.00080	2.8	96%
	(17)	(0.11)	(0.0079)	[0.0011]	(0.011)	[1.2,5.5]	
Years of schooling	14	0.091	0.0016	23	-0.0014	0.69	5.0%
	(2.4)	(0.0191)	(0.0015)	[<0.001]	(0.0028)	[0.37,1.2]	
Has a professional job	37	1.4	0.029	20	-0.032	11	29%
	(48)	(0.32)	(0.024)	[<0.001]	(0.040)	[5.8,18]	
B. Nonwhite							
Human capital		0.69	0.25	9.0	0.01	3.2	
		(0.68)	(0.075)	[0.0027]	(0.15)	[-2.8,10]	
Human capital (pre-trend adjusted)		3.4			-0.24	16	
		(1.1)			(0.18)	[5.6,29]	
Subindices (pre-trend adjusted)							
Completed high	87	0.91	0.095	2.8	-0.058	4.2	4.9%
school/GED	(34)	(0.54)	(0.037)	[0.095]	(0.069)	[-0.78,10]	
Attended some college	59	1.1	0.070	2.4	-0.094	5.3	9.0%
	(49)	(0.74)	(0.051)	[0.12]	(0.11)	[-2.2,13]	
Completed college	25	2.1	0.12	14	-0.16	10	40%
	(43)	(0.58)	(0.039)	[<0.001]	(0.097)	[4.5,16]	
Prof. or doc. degree	2.6	0.39	0.024	3.2	-0.035	1.8	71%
	(16)	(0.22)	(0.013)	[0.075]	(0.030)	[-0.21,4.0]	
Years of schooling	13	0.11	0.0089	8.1	-0.0068	0.53	4.0%
	(2.6)	(0.040)	(0.0028)	[0.0044]	(0.0061)	[0.16,0.95]	
Has a professional job	29	1.2	0.095	3.8	-0.066	5.6	19%
	(45)	(0.61)	(0.041)	[0.051]	(0.098)	[-0.17,12]	

Notes: Pre-trend-adjusted estimates are constructed by subtracting the estimated pre-trend (column 3) from the slope of the middle (column 2) and post-trend (column 5) segments of the spline. [1] For the subindex estimates, the "post-trend" estimates in column 5 have been adjusted for the pre-trend estimate shown in column 3. Column 6 scales by an estimated take up of 0.132 (s.e. 0.023) for whites and 0.214 (s.e. 0.028) for nonwhites (Appendix Table A5). See also Table 1 notes.

Table 4. The Effect of Head Start on Adult Economic Self-Sufficiency

	(1) Control mean (s.d.)	(2) ITT estimate (s.e.) [BH p-val]	(3) Slope of pre-trend[1] (s.e.) [BH p-val]	(4) Test of trend break at age 6 (p-val) [BH p-val]	(5) Slope of post-trend[1] (s.e.) [BH p-val]	(6) ATET [95% CI]	(7) ATET % change
ESS index		0.74	0.057	15	0.042	5.0	
		(0.17)	(0.024)	(<0.001)	(0.040)	[2.5,8.0]	
ESS index (pre-trend adjusted)		1.4			-0.015	9.2	
		(0.35)			(0.054)	[4.1,15]	
Subindices (pre-trend adjusted)							
Worked last year	84	0.66	0.033	8.1	-0.072	4.4	5.3%
	(36)	(0.23)	(0.016)	(0.0045)	(0.024)	[1.7,8.5]	
		[0.022]	[0.21]	[0.022]	[0.015]		
Weeks worked last year	40	0.34	0.018	7.3	-0.019	2.3	5.6%
	(20)	(0.12)	(0.0086)	(0.0068)	(0.013)	[0.64,4.3]	
		[0.027]	[0.21]	[0.027]	[0.65]		
Usual hours works per week	35	0.45	0.024	14	-0.018	3.0	8.7%
	(18)	(0.12)	(0.0078)	(<0.001)	(0.012)	[1.4,5.3]	
		[0.0013]	[0.017]	[0.0013]	[0.65]		
Log labor income	11	0.0064	0.0000	1.4	0.0028	0.043	
	(0.98)	(0.0053)	(0.00038)	(0.23)	(0.00075)	[-0.022,0.12]	
		[0.23]	[1.0]	[0.23]	[0.0012]		
Log family income/poverty	5.8	0.011	0.0002	3.0	0.0007	0.071	
	(0.93)	(0.0060)	(0.00042)	(0.082)	(0.0010)	[-0.017,0.16]	
		[0.16]	[1.0]	[0.16]	[1.0]		
In poverty*	12	-0.40	-0.017	6.1	-0.021	-2.7	-23%
	(32)	(0.16)	(0.011)	(0.014)	(0.027)	[-5.2,-0.71]	
		[0.040]	[0.52]	[0.041]	[1.0]		
Received public assistance*	12	-0.48	-0.0027	9.8	-0.013	-3.2	-27%
	(33)	(0.15)	(0.010)	(0.0017)	(0.019)	[-5.7,-1.2]	
		[0.010]	[1.0]	[0.010]	[1.0]		

Notes: *In poverty and received public program income are reverse-coded when used in the self-sufficiency index. Pre-trend-adjusted estimates are constructed by subtracting the estimated pre-trend (column 3) from the slope of the middle (column 2) and post-trend (column 5) segments of the spline. [1] For the subindex estimates, the "post-trend" estimates in column 5 have been adjusted for the pre-trend estimate shown in column 3. See also Table 1 notes.

Table 5. The Effect of Head Start on Adult Economic Self-Sufficiency, by Sex

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Control mean (s.d.)	ITT estimate (s.e.) [BH p-val]	Slope of pre-trend[1] (s.e.) [BH p-val]	Test of trend break at age 6 (p-val) [BH p-val]	Slope of post-trend[1] (s.e.) [BH p-val]	ATET [95% CI]	ATET % change
A. Men							
ESS index		0.60	-0.0080	1.4	0.096	4.0	
		(0.20)	(0.030)	(0.24)	(0.051)	[1.1,7.2]	
ESS index (pre-trend adjusted)		0.51			0.10	3.4	
		(0.44)			(0.054)	[-2.8,9.6]	
Subindices (pre-trend adjusted)							
Worked last year	90	0.38	0.0016	3.2	0.0038	2.5	2.8%
	(30)	(0.21)	(0.014)	(0.076)	(0.024)	[0.035,5.9]	
		[0.38]	[1.0]	[0.38]	[1.0]		
Weeks worked last year	44	0.19	0.0017	2.5	0.023	1.3	2.9%
	(17)	(0.12)	(0.0078)	(0.11)	(0.014)	[-0.26,3.0]	
		[0.45]	[1.0]	[0.45]	[0.49]		
Usual hours works per week	40	0.36	0.0094	8.2	0.026	2.4	5.9%
	(17)	(0.13)	(0.0080)	(0.0042)	(0.015)	[0.66,4.5]	
		[0.025]	[1.0]	[0.025]	[0.49]		
Log labor income	11	-0.0041	-0.00077	0.34	0.0039	-0.027	
	(0.88)	(0.0070)	(0.00049)	(0.56)	(0.00091)	[-0.12,0.071]	
		[1.0]	[0.80]	[1.0]	[<0.001]		
Log family income/poverty	5.9	0.0069	0.000010	1.0	0.00029	0.045	
	(0.87)	(0.0068)	(0.00049)	(0.31)	(0.0010)	[-0.052,0.14]	
		[0.93]	[1.0]	[0.93]	[1.0]		
In poverty	8.9	-0.070	0.0016	0.14	-0.024	-0.47	-5.2%
	(28)	(0.19)	(0.012)	(0.71)	(0.026)	[-3.0,1.8]	
		[1.0]	[1.0]	[1.0]	[1.0]		
Received public assistance	11	-0.72	-0.019	12	-0.017	-4.8	-42%
	(32)	(0.21)	(0.014)	(<0.001)	(0.025)	[-8.3,-2.0]	
		[0.0042]	[1.0]	[0.0042]	[1.0]		

Notes: *In poverty and received public program income are reverse-coded when used in the self-sufficiency index. Pre-trend-adjusted estimates are constructed by subtracting the estimated pre-trend (column 3) from the slope of the middle (column 2) and post-trend (column 5) segments of the spline. [1] For the subindex estimates, the "post-trend" estimates in column 5 have been adjusted for the pre-trend estimate shown in column 3. Column 6 scales by an estimated take up of 0.151 (s.e. 0.022) for men and 0.145 (s.e. 0.022) for women (Appendix Table A5). See also Table 1 notes.

Table 5. The Effect of Head Start on Adult Economic Self-Sufficiency, by Sex (Continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Control mean (s.d.)	ITT estimate (s.e.) [BH p-val]	Slope of pre-trend[1] (s.e.) [BH p-val]	Test of trend break at age 6 (p-val) [BH p-val]	Slope of post-trend[1] (s.e.) [BH p-val]	ATET [95% CI]	ATET % change
B. Women							
ESS index		0.74	0.097	15	-0.011	5.1	
		(0.22)	(0.033)	(<0.001)	(0.048)	[2.1,9.1]	
ESS index (pre-trend adjusted)		1.8			-0.11	12	
		(0.47)			(0.055)	[5.9,21]	
Subindices (pre-trend adjusted)							
Worked last year	79	0.79	0.058	5.1	-0.15	5.4	6.9%
	(41)	(0.35)	(0.024)	(0.024)	(0.039)	[0.28,12]	
		[0.12]	[0.093]	[0.12]	[0.00]		
Weeks worked last year	37	0.39	0.030	4.3	-0.064	2.7	7.2%
	(22)	(0.19)	(0.013)	(0.037)	(0.021)	[0.16,5.6]	
		[0.15]	[0.093]	[0.15]	[0.013]		
Usual hours works per week	30	0.42	0.032	6.2	-0.062	2.9	9.7%
	(18)	(0.17)	(0.011)	(0.013)	(0.018)	[0.63,5.9]	
		[0.075]	[0.028]	[0.075]	[0.0039]		
Log labor income	10	0.014	0.00071	3.5	0.0014	0.097	
	(1.01)	(0.0076)	(0.00050)	(0.063)	(0.0010)	[-0.0030,0.21]	
		[0.19]	[0.47]	[0.19]	[0.55]		
Log family income/poverty	5.7	0.011	0.00024	2.3	0.0011	0.079	
	(0.98)	(0.0076)	(0.00050)	(0.13)	(0.0013)	[-0.025,0.20]	
		[0.27]	[0.63]	[0.27]	[1.0]		
In poverty	14	-0.64	-0.034	8.2	-0.011	-4.4	-32%
	(35)	(0.22)	(0.015)	(0.0042)	(0.037)	[-8.5,-1.4]	
		[0.029]	[0.11]	[0.029]	[1.0]		
Received public assistance	13	-0.19	0.015	0.83	-0.0041	-1.3	-10%
	(33)	(0.21)	(0.014)	(0.36)	(0.025)	[-4.9,1.5]	
		[0.36]	[0.52]	[0.36]	[1.0]		

Notes: *In poverty and received public program income are reverse-coded when used in the self-sufficiency index. Pre-trend-adjusted estimates are constructed by subtracting the estimated pre-trend (column 3) from the slope of the middle (column 2) and post-trend (column 5) segments of the spline. [1] For the subindex estimates, the "post-trend" estimates in column 5 have been adjusted for the pre-trend estimate shown in column 3. Column 6 scales by an estimated take up of 0.151 (s.e. 0.022) for men and 0.145 (s.e. 0.022) for women (Appendix Table A5). See also Table 1 notes.

Table 6. The Effect of Head Start on Incarceration in Adulthood

	(1) Control mean (s.d.)	(2) ITT estimate (s.e.)	(3) Slope of pre-trend[1] (s.e.)	(4) Test of trend break at age 6 [p-val]	(5) Slope of post-trend[1] (s.e.)	(6) ATET [95% CI]	(7) ATET % change
Full Sample	1.2	-0.058	-0.0021	0.92	0.0018	-0.39	-32.6%
	(11)	(0.061)	(0.0046)	[0.34]	(0.0080)	[-1.2,0.49]	
White females	0.17	0.032	0.00080	0.76	0.0017	0.25	146.9%
	(4.1)	(0.037)	(0.0025)	[0.38]	(0.0044)	[-0.26,0.87]	
White males	1.3	-0.17	-0.0095	2.47	0.014	-1.2	-94.6%
	(11)	(0.11)	(0.0073)	[0.12]	(0.012)	[-3.1,0.26]	
Nonwhite females	0.50	-0.13	-0.014	0.46	-0.0035	-0.63	-126.7%
	(7.1)	(0.19)	(0.013)	[0.50]	(0.020)	[-2.6,1.3]	
Nonwhite males	6.5	0.17	-0.018	0.050	0.076	0.7	11.4%
	(25)	(0.75)	(0.055)	[0.82]	(0.083)	[-7.0,8.0]	

Notes: Dependent variable is a binary variable for being incarcerated at the time of observation. Incarceration is measured using group quarters residence in 2000 Census and 2005-2018 American Community Survey. All estimates are shown in percentage point units. Pre-trend-adjusted estimates are constructed by subtracting the estimated pre-trend (column 3) from the slope of the middle (column 2) and post-trend (column 5) segments of the spline. [1] For the subindex estimates, the "post-trend" estimates in column 5 have been adjusted for the pre-trend estimate shown in column 3. Column 6 scales by an estimated take-up of 0.149 (s.e. 0.022) for full sample, 0.130 (s.e. 0.023) for white females, 0.135 (s.e. 0.023) for white males, 0.208 (s.e. 0.029) for nonwhite females, and 0.223 (s.e. 0.029) for nonwhite males. See also Table 1 notes.

Table 7. Heterogeneity in the Effect of Head Start, by Local Programs and Economic Circumstances

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Intention-to-treat effects			Effects on pre-school enrollment		Average treatment effect on treated children (ATET)	
	Above median	Below median	F-test of difference [p-value]	Above median	Below median	Above median	Below median
A. Human capital index							
Medicaid exposure	3.8	1.8	5.1	12	18	31	10
	(0.92)	(0.63)	[0.02]	(2.4)	(3.0)	[17,57]	[2.6,20]
CHC exposure	2.5	2.4	2.6	20	11	12	22
	(0.59)	(0.58)	[0.10]	(3.4)	(2.5)	[6.3,21]	[12,39]
Food Stamps exposure	1.8	3.3	9.8	12	17	15	20
	(0.65)	(0.62)	[0.0017]	(2.6)	(3.2)	[3.9,29]	[12,34]
Poorest 300 counties	-1.8	2.8	0.28	7.3	13	-24	22
	(2.0)	(0.55)	[0.59]	(2.6)	(2.3)	[-110,39]	[13,38]
Predicted economic growth	3.3	2.2	2.304	14	13	23	17
	(0.69)	(0.60)	[0.13]	(3.3)	(2.5)	[12,45]	[8.3,30]
B. Economic self-sufficiency index							
Medicaid exposure	0.55	1.9	0.10	12	18	4.5	11
	(0.63)	(0.41)	[0.75]	(2.4)	(3.0)	[-6.5,16]	[5.6,18]
CHC exposure	1.8	0.93	5.9	20	11	8.8	8.5
	(0.41)	(0.39)	[0.016]	(3.4)	(2.5)	[4.4,15]	[1.4,19]
Food Stamps exposure	0.86	1.8	8.3	12	17	7.0	11
	(0.41)	(0.44)	[0.0039]	(2.6)	(3.2)	[0.29,16]	[5.5,21]
Poorest 300 counties	-0.15	1.4	0.08641	7.3	13	-2.1	11
	(1.6)	(0.35)	[0.77]	(2.6)	(2.3)	[-65,56]	[5.2,21]
Predicted economic growth	2.1	0.85	0.63	14	13	15	6.8
	(0.42)	(0.41)	[0.43]	(3.3)	(2.5)	[8.4,31]	[0.64,16]
C. Incarceration							
Medicaid exposure	-0.069	-0.040	0.052	12	18	-0.56	-0.23
	(0.10)	(0.075)	[0.82]	(2.4)	(3.0)	[-2.2,1.3]	[-1.1,0.67]
CHC exposure	-0.075	-0.029	0.44	20	11	-0.37	-0.27
	(0.073)	(0.066)	[0.51]	(3.4)	(2.5)	[-1.1,0.36]	[-1.6,1.1]
Food Stamps exposure	-0.060	0.012	0.83	12	17	-0.49	0.072
	(0.072)	(0.073)	[0.36]	(2.6)	(3.2)	[-1.7,0.70]	[-0.84,1.1]
Poorest 300 counties	0.61	-0.060	2.8	7.3	13	8.4	-0.48
	(0.41)	(0.061)	[0.094]	(2.6)	(2.3)	[-2.7,35]	[-1.5,0.50]
Predicted economic growth	-0.085	-0.028	0.65	14	13	-0.61	-0.22
	(0.075)	(0.068)	[0.42]	(3.3)	(2.5)	[-2.0,0.60]	[-1.3,0.88]

Notes: Point estimates and standard errors are in percentage-point units. The ITT estimates are constructed by subtracting the estimated pre-trend from the slope of the middle and post-trend segments of the spline as described in the text. ATETs are constructed by dividing the group-specific ITT estimate of Head Start's effect on long-run outcomes by the group-specific estimated first stage. Results for the 300 poorest counties are reported in the column for "above median" with results for other counties reported in "below median."

[Click here for Online Appendices]