The Effect of Early Childhood Education on Adult Criminality: Evidence from the 1960s through 1990s

John Anders[*], Andrew C. Barr[†], Alexander A. Smith[‡]

April 2021[§]

Abstract

We compare the effects of early childhood education on adult criminal behavior across time periods, using administrative crime data that provide significant precision advantages over existing work. We find that improvements in early childhood education led to large (20%) reductions in later criminal behavior, reductions that far exceed those implied by estimates of test score gains in prior studies. While the benefits generated account for a large portion of the costs of the education provided, we find substantial relative gains from the targeting of funds to high poverty areas and to areas without existing access to subsidized care.

[*]Department of Economics, Trinity University

[†]Department of Economics, Texas A&M University

[‡]Department of Social Sciences, United States Military Academy, West Point

[§]We thank Jason Lindo, Mark Hoekstra, and participants at the 2020 NBER Children's Program and the 2018 Southern Economics Association for helpful comments and suggestions. We also thank Laura Hitt and the North Carolina Smart Start organization for sharing their funding data. The opinions expressed herein reflect the personal views of the authors and not those of the U.S. Army or the Department of Defense. All errors are our own.

1 Introduction

Are criminals made or born? This fundamental question not only has important implications for our understanding of criminality, but also is central to any efforts aimed at reducing the large costs that crime imposes on society ($2 trillion annually).[1] Most policies address these costs by focusing on the incapacitation or rehabilitation of criminals. Relatively little is known about the factors that influence an individual's likelihood of becoming a criminal in the first place. The concentration of crime among a small number of perpetrators (less than 6% of the population commit the majority of crime) provides an opportunity for policy interventions to have outsized effects if they can prevent the development of criminals.[2] In fact, some estimates suggest that preventing the development of a single career criminal could result in more than 100 fewer victims of crime each year.[3]

One policy intervention that may influence later criminal behavior is early childhood education. Understanding its relationship to criminality is important because recent attempts to improve the quality and accessibility of early childhood education have been driven in large part by policymakers' belief that those interventions will have large impacts later in life. Positive externalities, such as reductions in adult criminal behavior, are particularly important for thinking about the social return to investments in early childhood education programs: unlike the labor market return to improvements in human capital, these benefits accrue largely to those who are not directly affected by the programs and thus may not be considered in parents' participation decisions. Because of the large social costs of criminal behavior, even modest reductions may warrant large subsidies to early childhood education.

[1] For context, $2 trillion dollars is 17% of annual GDP. (United States. Senate Committee on the Judiciary. Hearing on The Costs of Crime. September 19, 2006 (statement of Jens Ludwig))

[2] Farrington (2006) generates this statistic by tracking the criminal behavior of a set of boys in London. Given the higher propensity to commit crime among males, 6% is likely a substantial overestimate of the share of the population that commits the majority of crime.

[3] Across major crime categories, estimates suggest that a relatively small proportion of individuals (consistently less than 10%) account for the majority of crime. These "career criminals" commit hundreds of crimes each year (authors' calculations from Chaiken and Chaiken (1982)).

Indeed, in one often cited early childhood study, effects on crime account for 40-65% of the estimated benefits of the program (Heckman et al., 2010).

Yet, the limited evidence on the effect of early childhood education on later criminal behavior is mixed, imprecise, and limited largely to cohorts from the 1960s and 1970s (Figure A1a).[4] The evidence rests entirely on evaluations of three early childhood programs: Perry Preschool, the Abecedarian Project, and Head Start. Perry Preschool, a single evaluation of a small-scale high-intensity intervention, provides some of the more compelling evidence, but even in this case there is disagreement among researchers regarding the presence of statistically significant effects (Heckman et al., 2010; Anderson, 2008). A randomized evaluation of a similar program, the Abecedarian Project, indicates no overall effect of the program on crime (Campbell et al., 2012; Anderson, 2008; García et al., 2019). Furthermore, while these studies provide rigorous evidence driven by random assignment, both rely on very small samples with substantial attrition to support their conclusions.[5] Evidence on the effects of Head Start on criminal behavior is also mixed and relies on small samples, self-reported crime data, and sibling comparison approaches that may underestimate the effects of community-level program availability in the presence of spillovers (Deming, 2009; Garces et al., 2002). Johnson and Jackson (2017) overcome some of these limitations by leveraging the rollout of Head Start to estimate its effects on incarceration in the PSID, but these data have their own limitations. Furthermore, as we discuss in the next section and Appendix B, in all of these studies the underlying sample sizes limit the confidence we can place in the resulting estimates. For example, to identify the simple mean effect size of early childhood education on criminality in the literature, the statistical power of the Head Start studies ranges from 0.09 to 0.13. Under balanced priors on the likelihood of an effect, this implies false discovery rates of 42 to 54 percent. Finally, the early childhood context and counterfactual care

[4]The construction of Figure A1 is discussed in detail in Appendix B.

[5]At adult follow up, the Perry experiment had 123 members and the Abecedarian experiment had 111, both somewhat reduced from initial samples.

options have changed dramatically from the 1960s to the present day, suggesting that even if we had a convincing sense of the role of early childhood education in preventing criminal behavior in earlier periods, the same relationship may not exist in recent years.

We contribute to this literature by using administrative crime data to examine the effects of two large-scale early childhood education programs operating in two different time periods. Our use of the universe of administrative criminal conviction data from North Carolina results in estimates based on the criminal behavior of hundreds of thousands (versus a hundred or thousands) with corresponding increases in precision.[6] As we discuss and support with simulation exercises in Appendix C, this improved precision reduces the likelihood that any statistically significant estimated effects are false positives (i.e., the false discovery rate) and reduces the expected upward bias in the magnitude of reported effects. Our added statistical power allows us to be more confident that early childhood education reduces criminal behavior, that this reduction is likely considerably smaller than some prior estimates suggest, and that it is concentrated in counties with high poverty levels and limited access to alternative care options.

Our approach also substantially improves external validity over prior estimates of long-run crime effects that focus on small-scale single-site interventions and/or treatments that occurred roughly 50 years ago. It enables a direct comparison of adult crime effects of two different interventions implemented in very different contexts, while adding to the limited evidence on the effectiveness of *modern* early childhood education programs by demonstrating that these effects extend past the schooling years.[7] This is important because, as we discuss below, existing evidence suggests that short-term effects on test scores are not predictive of

[6] Our samples are based on the behavior of 1.5 million individuals (Head Start) and 1.4 million individuals (Smart Start) versus at most 4,536 individuals in the case of prior studies (see column (4) of Appendix Table B1 and Figure B2 for information on sample sizes of prior studies). Correspondingly, for example, our estimate of the effect of Head Start is more precise than a precision weighted average of the estimates from the five prior studies combined.

[7] For example, Fitzpatrick (2008, 2010); Ladd et al. (2014) show short-term effects on test scores and maternal labor force participation

long-run effects on criminal behavior.

Our unique data provide two additional advantages. First, the increased statistical power enables us to explore the extent of heterogeneity in effects across areas with different levels of poverty and access to existing care. This allows us to provide a potential explanation for the pattern of results in the literature and to inform the debate between universal and more targeted early education policies. Second, the detailed information available in the data allow us to better quantify crime benefits of the programs by directly estimating the discounted social benefits generated by reductions of various types of crimes.

Our investigation exploits the staggered introduction of the Head Start and Smart Start programs in North Carolina. The Head Start program, funded and administered through the U.S. Department of Health and Human Services, has been an integral part of early childhood education nationwide since the late 1960s. It is the largest early childhood education program in the United States, with an annual enrollment that has grown from 400,000 during its early years to nearly a million participants today. Head Start was designed to focus on the "whole child" by providing a number of wrap-around services alongside education (Ludwig and Miller, 2007). The Smart Start program was implemented in North Carolina in 1993 to address concerns about the preparation of children for school. The program helps parents pay for child care, improves the quality of early care and education programs, provides tools that help parents support their children, and ensures that children have access to preventative health care. The multi-faceted nature of Smart Start has many similarities with the early years of the Head Start program; in particular, both program implementations allowed considerable flexibility with respect to how funding was spent. Unlike Head Start, individuals do not enroll directly in the Smart Start program. Instead, the program operates by funding a variety of programs at the community level, and critically, these funds are not explicitly targeted towards poor children the way Head Start funds are.

To estimate the effect of early childhood education on adult crime, we leverage within-

county variation in exposure and funding levels generated by the rollout of each program, along with individual-level administrative data for the universe of convicted criminals in North Carolina between 1972 and 2018. These administrative data are particularly well suited to our estimation strategy because they include each criminal's county of birth. Thus we are able to link early childhood policy exposure to later criminal outcomes and can overcome a variety of measurement and endogeneity difficulties that likely inhibited earlier research.[8] We combine these data with birth counts to construct birth-county by birth-cohort conviction rates, which we then link with information on early childhood education funding and availability in each county and year.

Our estimates indicate that early childhood education reduces adult criminality. In particular, we find that both programs are responsible for reducing the conviction rate by approximately 20% in high poverty areas. Head Start availability reduces by 1.3 percentage points the likelihood of a serious conviction by age 35, but only in high-poverty counties.[9] Smart Start, which operated decades after Head Start and often in different counties, generates reductions in later criminal behavior on the same order of magnitude, reducing the likelihood of a serious criminal conviction by 0.6 percentage points (12 percent) overall by age 24.

The internal validity of these estimates is supported by the nature of the introduction of both programs, which suggests that program availability and funding was likely not related to an individual's propensity for criminal behavior.[10] Consistent with this, there is little relationship between baseline county characteristics and exposure for either program, and

[8]Most administrative crime datasets do not contain county of birth. Using these data would likely require researchers interested in the early childhood environment to make assumptions about the relationship between the location of arrest and earlier residence, thus introducing measurement error which would bias the effect estimates toward zero, as well as potentially introducing bias due to endogenous migration.

[9]This is perhaps not surprising given the focus of the Head Start program on poor children and the resulting concentration of funding among high-poverty counties. Head Start funding per capita is between three and four times larger in high- versus low-poverty counties (Appendix Figure A3).

[10]For example, Smart Start's pilot partnerships were chosen so as to be representative of North Carolina's diversity and geography.

our estimates are robust to a variety of standard checks that allow for differential trends in criminality across counties and the inclusion of controls for potential confounders.[11] The legitimacy of the identification strategy is further bolstered by event study estimates that show no significant program "impact" in the years prior to its introduction in a given county but a reduction in criminality in the years afterward. While our data only cover crimes committed in North Carolina, we find no evidence of differential out-of-state migration as a result of Head Start availability or Smart Start funding levels in early childhood.[12] Given the modest level of migration out of one's state of birth (and the lack of evidence of any differential out-migration), we view our estimates as providing a lower bound for the overall effect on criminal behavior.

Though not a targeted program, Smart Start's reductions in crime are larger in high-poverty counties (1.2 percentage points, or 23 percent) and are larger for blacks than whites, perhaps as a result of differential pre-existing access to high-quality care. The reductions generated by Smart Start are also larger in counties without access to Head Start, suggesting that the effectiveness of additional early education funding may diminish as funding increases. Both patterns of differential effects are consistent with the general patterns in the literature. More broadly, the concentration of effects in high-poverty and underserved areas has important implications for the growing number of cities and states considering the provision of universal Pre-K. Our results suggest much smaller average effects than those implied by a number of older studies of programs implemented in very high-poverty, underserved

[11] In the case of Head Start, for example, the estimates are robust to the inclusion of time-varying county-level controls for the availability of other War on Poverty Programs as well as birth county trends. Furthermore, Head Start availability is unrelated to other policy changes shown to affect crime (e.g., removal of lead from gasoline, changes to compulsory schooling law ages in North Carolina, or the legalization of abortion), which occurred at the state level and generally affected different cohorts of individuals.

[12] Across a variety of approaches and subsamples, our estimates indicate a small and non-significant relationship between childhood Head Start availability and the likelihood of living in one's state of birth. Assuming similar patterns of criminality among North Carolina leavers and stayers, our upper bound estimate of additional migration can explain at most 5% of our estimated effect. Additional analyses support Ladd et al.'s (2014) conclusion that Smart Start funding did not result in differential migration out of North Carolina. We discuss migration concerns further in Section 5.3.

settings, suggesting large potential gains from a more targeted approach.

A critical question for evaluating the benefits of more recent programs, where long-run outcomes cannot yet be observed, is the extent to which long-run crime effects are predictable based on earlier impacts on commonly measured short-run outcomes such as test scores. Our work underscores a recent body of evidence suggesting that effects on cognitive measures alone may not be predictive of long-run effects on criminal behavior. Under very conservative assumptions, the estimated effects of Smart Start on criminal behavior are more than 7 times the effect implied by the 3rdgrade test score estimates and the conditional correlation between test scores and crime. Effects of Head Start on educational attainment underpredict effects on criminal behavior by an even larger margin. These results suggest that early childhood education likely imparts important non-cognitive skills that are not captured by test scores.

Finally, we show that the discounted benefits generated by early childhood education's later crime reduction account for a substantial portion of the costs of the education provided. In fact, these benefits pay for nearly the full cost of the programs in high poverty counties, when accounting for multiple offenses per conviction. This is especially noteworthy because later crime reduction was not the stated objective of either program, and these benefits likely accrue in large part to those not directly affected by the programs. Taken together, our results support recent efforts to expand and improve early childhood education, but further point to large potential gains from targeting these efforts toward areas with the greatest need.

2 Evidence on the Origins of Criminal Behavior

Research on the developmental factors that influence the likelihood that an individual will become a criminal is limited, with many studies focusing solely on adolescence. A number of evaluations of the Moving to Opportunity project provide mixed evidence on the effect of

neighborhood environment on criminal behavior, while studies of assignment to foster care do suggest that family environment plays an important role in affecting both contemporaneous and later criminal behavior (Sanbonmatsu et al., 2011; Doyle, 2007, 2008).[13] Several studies have focused on the relationship between secondary education and crime, suggesting that additional years of schooling, increases in school quality, and changes in the composition of school peers can affect the likelihood of criminal behavior several years later (Lochner and Moretti, 2004; Deming, 2011). Because these adolescent treatments occur at an age when individuals typically first decide to engage in crime, they may directly affect the costs or benefits of crime (e.g. through direct exposure to crime or criminal peers) rather than influencing an individual's development.[14]

Research on earlier periods of development is somewhat less common, yielding mixed evidence of effects. Emerging evidence suggests an important role for early childhood health and nutrition. Evaluations of the Nurse-Family Partnership Program and the Food Stamp program suggest significant effects of early health interventions on adolescent or adult criminal behavior (Olds et al., 1998, 2007; Barr and Smith, 2018).

Still fewer studies examine the role of early childhood education, and the evidence is mixed, imprecise, and limited largely to cohorts from the 1960s and 1970s. Figure A1a consolidates information on effect size and estimate precision from a comprehensive review of studies that contain estimates of the causal effect of an early childhood education program on the likelihood that an individual will become a criminal.[15] The evidence rests largely on

[13] While early evaluations of the program found mixed evidence of effects on involvement with the criminal justice system at different ages (Katz et al., 2001; Kling et al., 2005; Ludwig and Kling, 2007), Sanbonmatsu et al. (2011) indicates no clear pattern of significant effects on arrests or delinquent behavior. Any effects that exist appear to be a result of current neighborhood conditions rather than the neighborhood that one grew up in. Doyle (2008) finds that those on the margin of placement are two to three times more likely to enter the criminal justice system as adults if they are placed in foster care.

[14] Deming (2011) suggests peer effects as one explanation for the effect of school quality on criminal behavior. Bayer et al. (2009) estimate criminal peer effects more directly, showing that juvenile offenders assigned to the same facility affect each other's subsequent criminal behavior.

[15] Appendix B provides the construction and details of these estimates.

evaluations of three early childhood programs: Perry Preschool, the Abecedarian Project, and Head Start.

Perry Preschool and the Abecedarian Project were pilot interventions that enrolled cohorts from the later 1960s and 1970s, respectively. Randomized evaluations of these somewhat resource-intensive programs provide mixed evidence. Heckman et al. (2010) finds that Perry Preschool participation led to large reductions in criminal behavior, but the Campbell et al. (2002) evaluation of the Abecedarian program indicates limited effects of the program on crime. While these studies provide rigorous evidence driven by random assignment, both rely on small sample sizes from single sites to support their conclusions. Other studies that adjust for multiple hypothesis testing suggest that neither program had statistically significant effects on adult crime for boys or girls at the 5 percent level (Anderson, 2008), or that the effect is robust only for participating girls (García et al., 2019). Additionally, it is not clear that these estimates readily generalize to larger scale programs with relatively less disadvantaged children, lower spending per pupil, and broader potential for important spillover effects (List et al., 2019).[16]

2.1 The Head Start Program

The largest early childhood education program in the United States, Head Start began as a summer program in 1965. It quickly expanded to a year-round program in the following year. Head Start's mission was to "[provide] the children of the poor with an equal opportunity to develop their full potential" (Office of Child Development, 1970). To that end, it was designed to focus on the "whole child" by providing a number of wrap-around services alongside education (Ludwig and Miller, 2007). These additional services included: providing nutritious meals and snacks, access to social workers, mental health and dental treatment, immunizations, and health screenings.

[16] As we discuss further below, these types of spillovers either attenuate or elude RCT estimates.

Head Start served a decidedly disadvantaged population in the early years of the program; the median family income of children enrolled in Head Start was less than half that of all families in the U.S. (Office of Child Development 1968). Quasi-experimental evidence focused on this period suggests that Head Start has had important long-term effects for the cohorts of children who participated. Leveraging sibling comparisons and discontinuities in grant-writing assistance and program eligibility, a number of studies have documented increased educational attainment, better health, and higher earnings (Carneiro and Ginja, 2014; Deming, 2009; Garces et al., 2002; Ludwig and Miller, 2007), even in the presence of short-term test-score fadeout (Deming, 2009).[17] More recent evidence indicates that the positive effects of the program persist into later ages (Thompson, 2017), interact positively with school funding levels (Johnson and Jackson, 2017), and even spill over into the next generation (Barr and Gibbs, 2018).

Two of the earlier studies include criminal behavior in their investigations of the long-run effects of Head Start but yield conflicting evidence. Using the CNLSY, Deming (2009) finds no effect of Head Start participation on criminal behavior. Using the PSID, Garces et al. (2002) find that Head Start participation reduces later criminality, but only among black participants. However, in their replication of this study, Miller et al. (2019) estimate an overall *positive* effect of Head Start participation on criminal behavior and an effect on black children that is half as large as the earlier study and not statistically significant. When they reweight this estimate to account for non-random selection into the identifying sample, the magnitude is halved again and is far from statistically significant. While effects on crime are not the focus of any of these papers, these estimates should be interpreted cautiously given the well-known issues with underreporting in self-reported measures of criminal behavior (Hindelang et al., 1981).[18] Moreover, these studies use family fixed-effects designs, where the

[17] Gibbs et al. 2011 provide a more comprehensive review of the Head Start literature.

[18] As Deming notes "previous research has found that self-reported crime data (unlike the arrest records used in the Perry Preschool study) are highly unreliable, both in the NLSY and in other data sources

choice to send one child to Head Start, and not their sibling, may be related to characteristics of the child or to the parents' circumstances at the time, potentially biasing the estimates.

As with the small-scale RCT studies, family fixed effect evaluations of early childhood programs capture only direct effects on participants rather than broader effects on the community. Even these direct estimates may be biased downward if there are important spillovers between siblings (in the case of family fixed effects) or between peers (in the case of single-site RCTs in small communities).[19] These types of spillovers may be particularly important in the context of the non-cognitive benefits provided by early childhood education (List et al., 2019). To better capture the overall community-level effects, we focus on the plausibly exogenous variation in Head Start access by county and year (see Figure 1), and use unique administrative crime data with the offender's county of birth for everyone convicted of a crime in North Carolina in the last four decades.[20] Because the Head Start program rolled out quickly, and grant funds were distributed directly to local grantees, the program became available in different counties at different times. There was substantial variation in the year of adoption among counties with similar baseline characteristics, generating plausibly exogenous variation in Head Start access (Barr and Gibbs, 2017; Thompson, 2017).

This rollout strategy is also used by Johnson and Jackson (2017) in an innovative empirical test for the presence of dynamic complementarities that suggests large effects of Head Start on the likelihood of ever being incarcerated, with these effects substantially magnified by the level of K-12 spending. While not the focus of their study, these incarceration measures have their own set of limitations. Incarceration measures in the PSID rely either on *retrospective self-reports* of incarceration (in the 1995 crime history module) or whether

(Lochner and Moretti 2004; Kling, Ludwig, and Katz et al. 2005)"

[19]For example, a back of the envelope calculation for Perry Preschool suggests that nearly all age-eligible children in poverty in Ypsilanti were part of the experimental sample.

[20]This strategy captures peer effects within the same birth cohort, but may still underestimate the total effect if there are spillovers to older cohorts or to individuals born in different counties. We believe the potential for this type of spillover is modest in our context.

a *family member reports* that the individual in question is unavailable to respond because they are in prison or jail. The crime measure may therefore be subject to reporting concerns similar to those in prior Head Start studies. Additionally, the family report data only indicate whether the individual was reported to be in prison or jail *at the time of the interview*, likely missing many periods of incarceration, particularly during later years when the PSID was fielded every other year.[21]

2.1.1 Statistical Power for Effects on Criminal Behavior in Previous Studies

In addition to the mixed nature of the evidence, uncertainty about the internal validity of the approaches, and questions about the effect of measurement error in self-reported criminal behavior, the underlying sample sizes in the survey data limit the extent to which the resulting estimates inform our understanding of the relationship between early childhood education and adult criminal behavior. We illustrate this at length in Appendix C for the studies depicted in Figure A1a. Simulation exercises using the rollout approach demonstrate the wide distribution of estimates one is likely to obtain with samples of the size available in the PSID, NLSY79, and CNLSY, as well as the limited statistical power available to identify effects of the magnitude that we find in our North Carolina population data. This low level of power implies a surprisingly high (40 percent) false discovery rate given balanced priors on the likelihood of an effect.[22] Indeed, if we conduct our same rollout analysis in the NLSY79, we estimate a null effect of Head Start availability with *positive* coefficients for those subgroups more likely to be eligible.

This lack of statistical power in the existing evidence base is not a function of the empirical approach. In Appendix Table C3 we summarize the statistical power of each study to identify the average published effect size, following the guidance provided in the meta-

[21]This would attenuate the resulting estimates toward zero, but could also bias estimates if there are different crime-age profiles across counties (since the incarceration of later cohorts may be undermeasured).

[22]See Appendix C for further discussion of these points.

analysis literature.[23] For example, to identify the simple mean effect size per $1,000 in the literature, the statistical power of the Head Start studies ranges from 0.09 to 0.13. Under balanced priors on the likelihood of an effect, this implies false discovery rates of 42 to 54 percent. These false discovery concerns regarding *effects on criminality* are not limited to the Head Start literature; the randomized evaluations of the more intensive interventions are only modestly better powered. This does not imply that prior estimates of the effect of early childhood education on adult criminality are incorrect, but it does suggest that a better-powered study could dramatically narrow the confidence interval for these effects.[24] Our use of administrative data on the universe of criminal convictions in North Carolina results in estimates based on the criminal behavior of hundreds of thousands (versus a hundred or thousands) and corresponding increases in precision.

2.2 North Carolina's Smart Start Program

Our approach also substantially improves external validity over prior estimates of long-run crime effects that focus on small-scale single-site interventions and/or treatments that occurred roughly 50 years ago. While the initial rollout of Head Start in the 1960s and 1970s provides a compelling source of variation, it is not clear whether these early estimates can be generalized to more recent periods. Indeed, the extreme poverty, hunger, and poor early childhood health that were prevalent during this period are uncommon in recent decades and the counterfactual early childhood education landscape has changed dramatically.[25] To assess the more general potential for early childhood education to reduce criminality, we examine the impact of the Smart Start program in North Carolina.

[23] This will be an overestimate of the true average effect size in the presence of a publication preference for statistically significant effects. Indeed, as we discuss in Appendix C, many of these estimates do not replicate across studies, sometimes even when using the same data set.

[24] This is not a function of the empirical approach; the family fixed effects designs have similar power limitations in these samples.

[25] For example, the enrollment rate of low-income 4 year old children has more than doubled from the 1960 to the 1990s.

The Smart Start program was created in 1993 to address concerns about the school-readiness of children in North Carolina. The program helps parents pay for child care, improves the quality of early care and education programs, provides tools that help parents support their children, and ensures that children have access to preventative health care. The multi-faceted nature of this program is similar in many ways to the early years of the Head Start program. But unlike Head Start, individuals do not enroll directly in the Smart Start program. Instead, the program operates by funding a variety of initiatives at the community level. And critically, the funds are not required to be targeted toward poor children.

Like Head Start, Smart Start was rolled out to different counties at different times, resulting in significant variation in funding across birth cohorts within similar counties (Figure 1). Ladd et al. (2014) use this variation in county-year funding to demonstrate sizable short-term effects of Smart Start funding on 3rd grade test scores (around 5 percent of a standard deviation for the average level of funding in a county at ages zero to five). However, we are not aware of any studies of the longer-term effects of Smart Start. We adopt the identification strategy of Ladd et al. (2014) to provide the first estimates of the effect of Smart Start on an adult outcome, criminal behavior.

2.3 Do Cognitive Effects in Children Predict Long-Run Criminal Behavior Effects?

A critical question for evaluating the benefits of more recent programs, where long-run outcomes cannot yet be observed, is the extent to which long-run effects are predictable based on earlier impacts on commonly measured short-run outcomes such as test scores. The nature of the relationship between cognitive skills and criminal outcomes is not well understood. Indeed, we are unaware of any evidence that a program that improves cognitive skills in childhood (as Smart Start and Head Start have) will necessarily produce reductions

in adult criminal behavior. In fact, there is a variety of evidence that non-cognitive skills are much more closely related to impulsive and criminal behavior than cognitive skills (i.e., test scores), and that each skill type can be influenced independently (Heckman et al. 2006; Hill et al. 2011; Heckman et al 2010; Jackson 2018).[26] Deming (2009), finds large effects of Head Start participation on measures of cognitive ability (0.13 SD at ages 7-10), but no effect on criminal behavior and Heckman et al. (2013) decompose the effects of Perry Preschool on eventual criminal behavior and show that they operate entirely through a non-cognitive channel (externalizing behavior). Simply put, there is evidence of early childhood education affecting cognitive skills but not criminal behavior, as well as evidence of effects on criminal behavior but not cognitive skills.[27]

This lack of relationship between effects on cognitive measures and effects on adult criminal behavior in prior studies may not generalize to all cognitive measures, such as those used by Ladd et al. (2014) to estimate Smart Start effects. However, Deming (2011) uses the same North Carolina test score measures as Ladd et al. (2014) to demonstrate long-run effects of school choice on criminal behavior, despite the gains in school quality generating no effect on test scores, suggesting that the crime effects are not operating through the cognitive channel in this context. Indeed, even the conditional correlation between the North Carolina test scores and criminal behavior, which is likely partially driven by omitted variable bias, suggests a weak relationship.[28] If we use the conditional correlations between test scores and criminal behavior and assume a persistent effect on test scores from 3rd through 5th grade,

[26]Jackson (2018) shows that teachers can have effects on test scores and behaviors, and that these value-added measures are only weakly correlated. This suggests that many teachers who raise test scores do not improve behaviors. Along similar lines, he shows that about 75 percent of the variation in his behavior index is unrelated to test scores.

[27]More generally, Gibbs et al. (2011) suggest that while test score fade-out is observed in studies of Head Start, Perry Preschool as well as kindergarten class-size, the absence of cognitive effects in certain grades does not preclude significant long-run gains.

[28]For example, a consistent standard deviation reduction in test scores across grades 3 through 5 generates a relative odds ratio of 1.3, whereas ever being suspended in each of those grades generates an odds ratio of 3.6 (Deming 2011).

we can infer a conservative upper bound on the magnitude of Ladd et al.'s implied effect of Smart Start on criminal behavior of -0.088 pp, a reduction of 1.75 percent of the mean.[29] In other words, the effects of Smart Start on the cognitive channel imply an extremely modest reduction in crime.

A natural question is whether the richer set of observed human capital accumulation effects in the Head Start literature might suggest larger impacts on criminal behavior than those implied by test scores alone. Indeed, the positive effect of Head Start on educational attainment and the negative relationship between educational attainment and criminal behavior hints that effects on crime might be expected or even entirely predictable. However, a more careful effort to combine prior estimates of the effect of Head Start on educational attainment with causal estimates of the effect of education on crime suggests a different conclusion. Previously estimated effects on educational attainment imply reductions in crime of under 0.8 percent of the mean, an extremely small effect relative to the observed effects of Head Start on other margins.[30]

The different roles of cognitive and non-cognitive skills in determining adult outcomes suggests that the evidence on the test score (or subsequent educational attainment) effects of modern early childhood education programs might not successfully predict these programs' impacts on criminal behavior. To better predict these longer-run behavioral effects, it is likely that researchers will need to develop better short-run *behavioral* measures. This is an area of future research that will be of great practical benefit to policymakers.

[29] We use the estimates in Appendix Table III of Deming (2011) to calculate odds ratios, which we apply to our measures. We assume Smart Start effects operate by moving individuals in the range between -2 and -1 SD and apply the estimates from Ladd et al. (2014) accordingly.

[30] We combine the mean effect on years of education and high-school graduation across published studies of Head Start (scaled to the implied effect given the participation rates in our high-poverty counties) with the respective estimates of the effect of education on crime from Lochner and Moretti (2004).

3 Data

Our primary data source is administrative conviction data from the state of North Carolina. We use these data, along with birth records, to calculate rates of conviction by county-year birth cohort. We use Head Start and Smart Start funding information to construct measures of program exposure at the same level. We link the measures of program exposure to the conviction rates at the county-year birth cohort level to estimate the effect of each early childhood education program on later adult crime.

3.1 North Carolina Data

We obtained data containing public information on all individuals convicted of a crime in North Carolina between 1972 and 2018 from the North Carolina Department of Public Safety.[31] The administrative data contain information on the type of crime, including the statute of the offense and whether it was a felony, as well as the name, dates of birth, sex, and race of the perpetrator. An important advantage of the North Carolina data over other state criminal databases is the inclusion of county of birth for each individual. Combining information on criminals' years and counties of birth with birth counts obtained from the North Carolina Department of Health and Human Services allows us to construct conviction rates for birth cohorts of individuals born in North Carolina. For example, to generate the cohort conviction rate by age 35 for children born in county c in 1961, we divide the number of individuals born in county c in 1961 and convicted by age 35 by the total number of individuals born in county c in 1961.

For our Head Start analyses, we restrict the sample to individuals born between 1955 and 1968, allowing us to leverage the variation in Head Start availability that occurred up

[31] For the cohorts in our sample, this includes offenders aged 16 and older. Until a law change in 2017 (not fully implemented until 2019), 16-year-olds were charged as adults in North Carolina. No cohorts in our sample were affected by this law change.

to and including 1972 (as Head Start availability is measured four years after birth). For our Smart Start analyses, we focus on individuals born between 1980 and 1994, which allows us to observe criminal convictions through age 24 for all cohorts in the sample (we observe convictions through 2018).[32]

Summary statistics are contained in Table 1. Slightly less than 5 percent of individuals born between 1955 and 1968 were convicted of a serious crime by age 35; likewise, slightly more than 5 percent of individuals born between 1980 and 1994 were convicted of a serious crime by age 24. Our definition of a serious crime is based on the FBI's Part I offenses.[33]

While the data contain the universe of individuals convicted of a crime in North Carolina during this time period and allow us to link these individuals to their counties of birth, they are limited in that they do not allow us to observe convictions for individuals who are born in North Carolina and then leave the state. While most likely criminals remain in their state of birth (and the likelihood of criminal behavior is lower for those who leave), this may be a concern for the interpretation of our estimates if Head Start availability or Smart Start spending affects out-of-state migration, especially if it leads individuals with a higher-propensity for criminal behavior to leave the state.[34] We return to this concern below in Section 5.3, providing evidence that program availability and funding do not appear to

[32]We find similar effects when we focus our estimates on convictions through older ages, which necessitates a reduction in our identifying variation (e.g., through age 25 (using cohorts through 1993), and through age 26 (using cohorts through 1992)). See Appendix Table A12. We focus on convictions by age 24 because it provides a balance between variation in Smart Start funding exposure across birth cohorts and our capacity to observe these cohorts at ages with the highest rates of criminal behavior.

[33]We largely follow the convention of FBI's Uniform Crime Reporting Statistics for Part I offenses. Violent crimes are defined as offenses containing the words "murder", "assault", or "robbery". Property crimes are defined as offenses containing the words "burglary" or "larceny".

[34]Roughly 70% of individuals born in North Carolina during this period reside there between the ages of 18 and 35. This share is even higher (roughly 80%) for those with the highest rates of criminal behavior (between ages 18 and 24, non-white, or with less than a high-school degree). Authors' calculations using public Census and ACS data.

influence migration rates.[35,36]

3.2 Head Start Availability Measure

We follow Barr and Gibb's (2017) definition of Head Start availability, which relies on county-year data from the Community Action Programs (CAP) and Federal Outlay System (FOS) files obtained from the National Archives and Records Administration (NARA).[37] We construct an availability indicator equal to one if a county had Head Start expenditures per four-year old above the national tenth percentile.[38,39] We do not otherwise leverage data on appropriated dollar amounts due to concerns about the accuracy of the recorded funding amounts in the early years of the Head Start program as well as the potential endogeneity of funding levels.

[35]Specifically, for Head Start, we explore the relationship between measures of childhood Head Start availability (at the state of birth by birth cohort level) and the likelihood of living in one's state of birth. Across a variety of approaches and subsamples our estimates indicate a small and non-significant relationship between childhood Head Start availability and the likelihood of living in one's state of birth. Similar analyses support Ladd et al.'s (2014) conclusion that Smart Start funding did not result in differential migration out of North Carolina. We address concerns about differential migration further in Section 5.3

[36]We may also be missing individuals with one-time nonviolent convictions (at any age) or one-time drug convictions (under age 22) that hired a lawyer and had the record expunged. However, these possible missing observations are a threat to identification only if they occur differentially based on early childhood educational status.

[37]See Barr and Gibb's (2017) Data Appendix for details.

[38]The measure for a cohort is based on funding 4 and 5 years after birth, when the cohort was age 3-5 years old, following Barr and Gibb's (2017). The modal age of participation in the summer program was 5, but the modal age of participation in the full-year program was 4 between 1966 and 1968. Throughout this period 86 to 92 percent of participants were ages 3 to 5. We use the national tenth percentile threshold for consistency with Barr and Gibbs (2017), who find that using this threshold better predicts Head Start take-up, but neither the values of the availability indicator (in North Carolina) nor the main results are sensitive to moving this threshold. Indeed, in North Carolina nearly all positive cohort by county Head Start spending exceeds the tenth percentile. As a result, there is little concern regarding misclassification and the point estimates are insensitive to using thresholds between the 1st and 30th percentile.

[39]Kindergarten was not widely available in North Carolina in the 1960s. Even by 1973, only 16 percent of children were enrolled in kindergarten and it wasn't until 1978 that public kindergarten became available for all age-eligible children.

3.3 Smart Start Funding Exposure Measure

We follow Ladd et al.'s (2014) definition of Smart Start funding penetration, which relies on county-year data on Smart Start funding obtained from the Budget & Grants Compliance Officer at Smart Start. For each county-year, we compute the annual funding amount per child aged 0-5. We then sum this county-year funding penetration by cohort, generating the total amount of funding exposure a given cohort received from ages 0 to 5. Thus, our Smart Start funding measure tracks the cumulative exposure to Smart Start funding a given county birth-cohort received.

4 Estimation of Program Availability Effects

To estimate the effect of early childhood education on adult crime, we exploit within county variation in exposure generated by the roll out of each program. For Head Start, we leverage within county variation in the availability of Head Start generated by the initial roll-out of the program in the 1960s. For example, we utilize the fact that eligible four-year-olds in 55 out of North Carolina's 100 counties had access to Head Start in 1968 while no four-year-olds had access to Head Start prior to 1965 (panel (a) of Figure 1). For Smart Start, we leverage the funding expansion from zero counties in 1992 to all 100 counties by 2000, with substantial differences in per capita funding levels (panel (b) of Figure 1).

For both interventions, we estimate the following specification:

$$C_{ct} = \beta EC_{ct} + \gamma(X_c \times t) + \alpha_c + \sigma_t + \epsilon_{ct}, \tag{1}$$

where C_{ct} is the conviction rate for those born in county c in year t, EC_{ct} is a measure of county-year birth cohort exposure to the early childhood policy, and α_c and σ_t are birth county and birth year fixed effects. In robustness checks, we include $X_c \times t$, which are

controls for baseline birth county characteristics interacted with a time trend, to account for any differential trends by county characteristics.[40] For example, this would account for differences in crime trends between more and less rural counties.[41] Standard errors are clustered at the county of birth level.

The key identifying assumption is that exposure to each early childhood education policy is, conditional on birth county and birth year fixed effects, unrelated to the propensity to be convicted of a serious crime for some reason other than the policy. For example, if Head Start was rolled out at the same time that other programs targeted at children were adopted we could be mistakenly attributing the effects of this other program to Head Start.

The exogeneity of the national-level Head Start roll out has been supported at length in related work, with multiple studies referencing the quick and haphazard nature of initial Head Start grant-making as well as demonstrating a limited association between baseline county observables and the presence and timing of Head Start adoption once a county's poverty level is accounted for (see, for example, Barr and Gibbs (2017); Johnson and Jackson (2017); and Thompson (2017)).[42] We further bolster the exogeneity of the Head Start rollout in North Carolina with an event study (Figure 2), discussed in detail in Section 5.1, that finds no "effects" on the likelihood of conviction in the years Head Start's initial adoption in a county and a sharp jump immediately following the program's introduction.

[40] For Head Start the baseline characteristics are from 1960, whereas for Smart Start they are from 1980.

[41] Following Hoynes et al. (2016), the 1960 county characteristics include the percent of people living in families with less than $3,000 (1960 dollars), the percent living in urban areas, the percent black, the percent under 5 years old, the percent over 65 years old, the percent of land in farming, and the percent of employment in agriculture. Following Ladd et al. (2014), the 1980 county characteristics include the share of births to black mothers, the share of births to Hispanic mothers, the share of births to low education mothers, the share of the population using food stamps, the total number of births, the total population, and the median family income.

[42] We present analogous analyses specific to North Carolina in Appendix Tables A1 and A2. We find no statistically significant relationship between county characteristics in 1960 and the timing of Head Start availability within North Carolina. Appendix Figure A10 presents this relationship graphically and similarly suggests little relationship between baseline county characteristics and the timing of introduction. Appendix Table A1 indicates that more populous counties were more likely to get the program at all during this time period.

Just as in the case of Head Start, the nature of Smart Start's introduction suggests that program exposure was, conditional on birth county and birth year fixed effects, unrelated to an individual's propensity for criminal behavior. The pilot partnerships were chosen to be representative of North Carolina's diversity and geography, with one county selected from each congressional district. As noted by Ladd et al. (2014), experts used county-level data on poverty-rates, tax base, and number of young children in need, to rank North Carolina's 100 counties into one of four resource bands: high, medium high, medium low, and low. As required by the Smart Start legislation, the experts selected three counties within each resource band to receive initial funding. The program quickly expanded, reaching all counties by the 1998-99 school year. Consistent with the initial intent to provide balanced exposure to the program, there is little relationship between baseline county observables and the timing or magnitude of Smart Start funding (Appendix Figures A11 and A12, Appendix Table A3). We provide additional evidence for the exogeneity of the Smart Start rollout in North Carolina with another event study (Figure 3), discussed in detail in Section 5.1, that finds no "effects" on the likelihood of conviction prior to the initial funding of Smart Start in a county and a phasing in of effects following initial funding that is consistent with the targeting of funds to a broader age range.[43]

While both programs experienced a similarly rapid rollout across the state, they differed substantially in how funding was distributed to participating counties. As Appendix Figure A2 shows, Head Start funding was concentrated in higher poverty counties, while Smart Start funding was distributed relatively evenly across counties. This difference reflects the different mandates of the two programs. As part of the War on Poverty, the Head Start program focused on children from families at or below the federal poverty line; indeed, at least 90 percent of Head Start participants at each site were required to be from families below the poverty line. Consistent with this, Head Start funding per four-year old is three

[43]We discuss the expected and observed pattern of results further below.

to four times as high in high-poverty counties (Appendix Figure A3). By contrast, poverty was not an explicit focus of the Smart Start program. This yielded a more equal distribution of funds across high and low poverty counties. Nevertheless, differential effects of Smart Start by poverty status may still occur if families in poverty have fewer outside options for high-quality early childhood education. Accordingly, for both policies we conduct much of our analyses separately for high and low poverty counties, splitting counties at the median county poverty level for North Carolina in the baseline year for each program (i.e., 1960 for Head Start and 1980 for Smart Start).[44]

5 Main Results

Our primary interest is in the coefficient β in Equation 1, which represents the effect of early childhood education availability or funding on adult crime.[45] We find large effects on the likelihood of an adult criminal conviction for individuals born in high-poverty counties, with effect sizes around 20 percent (Table 2).[47]

Our main Head Start estimates indicate that program availability generated a 1.3 percentage point reduction in the likelihood of a criminal conviction by age 35 for high-poverty counties, but no measurable effect for low poverty counties (top row of Table 2).[48] While

[44]It is worth noting, however, that poverty rates improved dramatically across the state between the 1960s and 70s and 1980s and 90s.

[45]We observe county of birth and not county of residence throughout early childhood. If individuals move from their county of birth before age 5 this will introduce some classification error into our exposure measures. While prior evidence presented here and elsewhere suggests that this natural migration is unlikely to bias our estimates, the estimated effect of availability in one's county of birth will be attenuated somewhat from the effect of availability in one's county of residence. Given rates of moving during the late 1960s between birth and age 5 of around 0.15, the estimates should be scaled up by 17 to 18 percent for Head Start.[46] Overall lower rates of moving in the 1990s suggest this is an upper bound for the scaling of the Smart Start estimates.

[47]For Smart Start, the effect size is for average funding exposure, which is slightly less than $1,000 for cohorts born after Smart Start enters a county (these cohorts are affected throughout ages 0 to 5).

[48]Figure A4 presents coefficient estimates for the same specification by poverty quintiles. The most dramatic effect occurs in counties in the highest poverty quintile. We have also estimated specifications that interact the continuous poverty rate with an indicator for Head Start availability (Appendix Table A6). Using this approach, we estimate that the reduction in crime rate due to Head Start availability is

our main estimates are identified using the set of counties that ever received Head Start (by 1976), the results are similar when including the counties that did not receive Head Start during this period (Appendix Table A5).[49] Our reliance on later-treated counties to serve as controls for the earlier-treated counties and vice-versa has the potential to generate bias in the resulting estimate if the treatment causes a change in trends (for example, if the effect of treatment grows substantially over time) (Goodman-Bacon, 2018). Our dynamic estimates (presented in Figure 2), suggest at most a modest increase in treatment effects over time, implying that the magnitude of our overall DD estimate may be slightly attenuated. There may still be bias in the identification of our treatment effects if there is heterogeneity in the dynamic effects of treatment across counties that adopt Head Start at different times. In Appendix E, we use techniques in Sun (2020) and Chaisemartin (2018) to diagnose and directly address the potential for this type of "contamination bias" to influence our estimates. We find little evidence to suggest that contamination bias is a concern in our setting. Consistent with this, estimates that allow for dynamic heterogeneity are similar to our main estimates; if anything, they suggest that our standard estimates are modestly attenuated.

Because Smart Start funding is not targeted at a particular age group, we follow Ladd et al. (2014) in estimating the effect of the average funding provided per 0 to 5 year old over the first five years of life. In the bottom row of Table 2, we report estimates per $1,000, which is roughly equal to the average funding exposure for cohorts born after Smart Start enters a county during our sample period.[50] Our main estimates suggest a 0.6 percentage point (12 percent) reduction in the likelihood of a serious criminal conviction by age 24

0.2 percentage points larger for each 10 percentage point increase in the poverty rate. Consistent with our prior estimates, these estimates suggest that the effect of Head Start ranged from 0.0 percentage points in the county with the lowest poverty rate (23 percent) to roughly 1.4 percentage points in the county with the highest (74 percent).

[49] Appendix Figure A7 presents the corresponding event study, which similarly suggests a reduction in criminal behavior following Head Start availability.

[50] These cohorts are fully treated, unlike those born a 1 to 4 years before Smart Start entry, which experience partial treatment.

from an additional $1,000 in funding exposure (Table 2).[51] Due to the timing of the Smart Start variation, we cannot present effects on conviction by age 35. As with Head Start, the estimated effects are larger in high-poverty counties, a 1.2 percentage point (23 percent) reduction in the likelihood of a serious criminal conviction.[52] While SS did not explicitly target children in poverty, this pattern of results is perhaps explained by the lower quality counterfactual care options available to low-income families or in high-poverty areas. This would conform with recent evidence that suggests larger short-term effects of early childhood programming for those who would otherwise have relied on parental or relative care (Kline and Walters, 2016).

For both programs, the estimated effects in high poverty counties appear to increase with age by which conviction is measured. Appendix Table A11 shows our Head Start estimates for conviction by age 24, 30, and 35, while Appendix Table A12 shows Smart Start results for conviction by age 24, 25, and 26 (older ages are not feasible given the timing of the Smart Start variation). While the estimate of Head Start's effect on convictions by age 24 is not statistically significant, the percentage reduction in convictions remains similar to Smart Start for the same age. Furthermore, the corresponding event study (Appendix Figure A5) strongly suggests a reduction in the likelihood of conviction by age 24 following Head Start availability in a county.

While the rollout of the two programs occurred two to three decades apart, Head Start continued to operate in many North Carolina counties during the rollout of Smart Start. This provides an opportunity to investigate whether early childhood education funding yields

[51] For completeness, we have included Appendix Tables A7 and A8 that contain estimates from a binary specification (the analogs of Tables 2 and 3). Because, (1) this specification averages effects across age of exposure, and (2) the resulting DD estimate is downward biased in the presence of treatment effects that strengthen over time, the estimates are attenuated somewhat from those implied by the baseline exposure measure.

[52] As with Head Start, we have estimated specifications that interact the continuous poverty rate with our Smart Start funding measure (Appendix Table A9). Using this approach, we estimate that the reduction in crime rate due to a $1,000 of Smart Start funding is 0.5 percentage points larger for each 10 percentage point increase in the poverty rate.

increasing or diminishing returns. In Table 3, we examine whether Smart Start funding was more or less effective in the presence of Head Start. The estimates indicate that the effects of Smart Start are largest in counties that were not previously served by Head Start (columns (2) and (5)). In contrast, in counties previously served by Head Start, there were small effects of Smart Start that are not significantly different from zero (columns (3) and (6)), underscoring the importance of targeting funds to areas with the greatest need. Appendix Table A10 shows estimates of the full interaction of Smart Start funding and the presence of Head Start, illustrating that these differences are statistically different and underscoring the diminishing returns to early childhood education investments. In addition to confirming the results in Table 3, the full interaction results demonstrate that among both high and low poverty counties, counties without Head Start experienced larger effects of Smart Start funding than those with Head Start.

Our results appear consistent with the general pattern of effects in the literature, though it is difficult to draw strong conclusions from a small set of studies that identify heterogeneous parameters of different programs. For example, the fact that the treatment effects on criminal behavior are larger for Perry than Abecedarian could be explained by differences in the affluence of the local communities and the availability of counterfactual care options. The participants in Perry were drawn from Ypsilanti, a town with high levels of poverty and limited access to alternative care options. In contrast, the participants in Abecedarian were from Chapel Hill, a more affluent town with extensive alternative access to care, including Head Start. Indeed, 75 percent of control group members in Abecedarian attended an alternative center-based care. While it is more difficult to compare across the heterogeneous parameters estimated in the Head Start literature, the general pattern of larger effects in earlier years (when there were greater levels of disadvantage and less alternative care options) is consistent with this explanation.

5.1 Dynamics and Robustness

To understand the dynamics of how the program may have affected adult criminal outcomes and to test for pre-trends that may confound our baseline specification, we also present estimates from event study specifications. For example, for Head Start, we center counties around the first year that the program is available, and estimate the effect of leads and lags of program availability.[53]

Figure 2 indicates a flat trend in cohort conviction rates before Head Start rollout for both high and low poverty counties.[54] This provides evidence that our difference-in-differences estimates are not capturing differential pre-existing trends in the years prior to county's rollout of Head Start.[55] For cohorts exposed to Head Start, we see significant decreases in the conviction rate for the high poverty counties but continue to see no evidence of changes in the low poverty counties.[56] In the high poverty counties, the estimates of crime reduction

[53]We estimate the following specification separately for counties above and below the median poverty rate:

$$C_{ct} = \sum_{\tau=-6}^{7+} \beta_\tau 1(t = T_c + \tau) + \alpha_c + \alpha_t + \gamma(X_{c,60} \times t) + \epsilon_{ct} \tag{2}$$

. We are primarily interested in the coefficients on the indicators, $1(t = T_c + \tau)$, each of which indicates how many years cohort t in county c is removed from the first cohort in county c exposed to Head Start, T_c. The distant event-times are binned (i.e., 7+) and the corresponding estimates are not presented in the figure. We use the first year in which the treatment measure exceeds the national 10th percentile (i.e., \$22 per child per year) as the year of adoption within a county. Nearly all of the positive Head Start funding amounts in North Carolina exceed the national 10th percentile , so our results would be almost identical regardless of whether we set our Head Start availability variable equal to one for county-birth year cohorts that exceeded the 1st, 5th, 10th, ..., 30th national percentile.

[54]While the estimates in Figure 2 includes linear trends interacted with 1960 county characteristics, Figure A6 shows that the results are similar without inclusion of these linear trends.

[55]Two of the pre-period estimates for high poverty counties suggest small but borderline significant negative effects before the availability of Head Start in a county. This is driven by the choice of omitted time period ($t = -1$), which has slightly higher conviction rates than the other pre-period cohorts. If anything, the pattern of pre-period effects suggests that there may be a very slight upward trend in criminality across cohorts prior to the first cohort exposed to Head Start (suggesting an attenuation of our estimated effect), although we cannot distinguish this trend from zero.

[56]This figure also addresses concerns that there were subsequent changes in a county that affected crime rates, such as changes to its criminal justice system, that are correlated with but not caused by the timing of a county's Head Start adoption. For such a correlation to produce our event study results, the policy change would have to precisely target only cohorts exposed to Head Start availability and have no effect on cohorts born just a couple years earlier.

appear to grow somewhat as the program persists in a county. In particular, the impact of Head Start availability in the first year of the program is substantially smaller than in subsequent years. This may be due to centers improving (or increasing the size of) their Head Start programs during the first years of operation or as a result of peer effects.[57] Funding does appear to increase somewhat during the early years of program operation, consistent with program growth (Appendix Figure A3).

The targeting of the Smart Start intervention to a broad age range does not lend itself to a standard event study because cohorts experience different intensities of treatment based on their age relative to the maximum targeted age (i.e., 5) when the funds become available in a county. This is illustrated in the top panel of Figure 3, which demonstrates that our treatment measure rises for five to six years following adoption. The bottom panel of Figure 3 provides event study estimates of the effect of Smart Start availability on the likelihood of a criminal conviction. As with Head Start, our baseline dynamic estimates (bottom panel of Figure 3) indicate a flat trend in cohort conviction rates before Smart Start funding arrived in a county. This provides evidence that our estimates are not capturing differential pre-existing trends in the years prior to a county's adoption of Smart Start. Consistent with the phase in of our treatment measure, we observe a phase in of effects on the likelihood of a criminal conviction. The figure also shows that higher poverty counties enjoyed larger decreases in cohort conviction rates than low poverty counties, despite Smart Start funding being similar across county poverty levels.

The main estimates of the effects of both programs using Equation 1 are robust to the inclusion of pretreatment county characteristics interacted with time trends (see column (2) of Appendix Table A13). These controls limit concerns that that counties with different characteristics have differing trends in the likelihood of criminal behavior across cohorts

[57] If peer effects are an important factor in criminal behavior, we would expect smaller effects of the program in the first year as compared to subsequent years when older peers would have also experienced the program.

that are correlated with the timing of program rollout. In the case of Smart Start, the results are also robust to the inclusion of time-varying controls at the county by birth cohort level, which addresses concerns about changes in the composition or wealth of birth cohorts over time that could be driving the observed result (Appendix Table A14). Similarly, the Head Start estimates are robust to the inclusion of covariates indicating availability of other War on Poverty programs, such as Food Stamps, Medicaid, Community Health Centers, etc. (Appendix Table A15).[58,59]Finally, in Appendix Table A4 we explore whether the timing of adoption is correlated with either the predicted level or growth in crime for cohorts surrounding the rollout of each program. For both programs and samples, we construct predicted crime measures by regressing each county's baseline crime rate and the sample period growth in crime on the county's characteristics in the base year. We then regress the predicted crime measures against our timing and funding measures. For Smart Start, we find no evidence of earlier program availability or greater funding in NC counties that were predisposed to lower crime (or crime growth) based on their characteristics in the base year. In the case of Head Start, while the coefficients on funding suggest that program rollout was not endogenously determined by conviction levels and trends, the coefficients on timing suggest that early adopting counties experienced somewhat higher baseline conviction levels and predicted growth in convictions over our sample cohorts. While these associations are small, if anything, they suggest that our estimates may be *slightly* biased downward, since the baseline characteristics of early adopting counties predicted a growth in convictions over this time period, while we still estimate that program availability decreased convictions.

[58]We also test directly for relationships between these potential confounders and our measure of Head Start availability. Consistent with the limited effect of the War on Poverty controls on our estimates, we find no significant relationships between funding for various War on Poverty Programs and Head Start availability (Appendix Table A16)

[59]While the same time-varying county characteristics that we employ for our Smart Start robustness checks are not available for Head Start cohorts, we find no relationship between Head Start availability and measures of infant mortality. This suggests that the relationship between Head Start availability and later criminal behavior is not driven by broader improvements in infant health or medical treatment unrelated to Head Start.

While our baseline inference relies on standard errors clustered at the county of birth, we have also explored the robustness of our p-values to an even more conservative approach: randomization inference. Under this procedure, we randomly assign the rollout year or funding level of Head Start and Smart Start in each county and estimate our baseline specification. The distribution of these estimates over 1,000 iterations is contained in Appendix Figure A8 and A9. As can be seen in the figures, the estimates we observe in our baseline results are quite unlikely under random assignment. The implied two-tailed "p-values" we obtain from this randomization inference approach are similar to those obtained using our baseline approach with standard errors clustered on the county of birth.[60]

5.2 Heterogeneity of Effects

The most natural channel through which early childhood education may affect crime is by raising the return to work and thereby increasing the opportunity cost of participating in criminal behavior. This channel is supported by the observed increases in human capital demonstrated in recent evaluations of modern early childhood programs (for example, Fitzpatrick (2008) or Ladd et al., (2014)), although the existing evidence suggests that improvements in test scores may not translate into crime reduction in the absence of improvements in non-cognitive skills. Early childhood education may also affect the financial or psychological benefits of criminal behavior or it may alter preferences in other ways, either by directly influencing child development or by influencing parenting, perhaps via effects on parental labor supply (Fitzpatrick, 2010).

Heterogeneity in effects across crime types may provide a hint at how these early child-

[60]Implied "p-values" presented are the two-tailed statistics calculated as the share of coefficient estimates obtained under random assignment of Head Start timing or Smart Start funding that are larger in absolute magnitude than the estimate produced using the true timing of assignment and funding. We preserve the patterns of availability and funding and randomly reassign to counties to better approximate the structure of rollout. This results in more conservative p-values than a strictly random assignment of timing or funding and better mimics our assumption of conditionally random assignment.

hood programs are affecting an individual's adult propensity to commit crime. For example, changes in the opportunity cost of crime may be more likely to affect property offenses, whereas effects on child development may be more likely to affect violent offenses. In Appendix Table A17, we explore whether the effects of early childhood education on serious convictions differ by crime type. The coefficients for violent crimes are not statistically distinguishable from those for property crimes for Head Start or Smart Start. The point estimates indicate a 0.25 percentage point reduction in the likelihood of a property conviction and a 0.39 percentage point reduction in the likelihood of a violent conviction by age 24 (column 1, Panel B); as before, these effects are larger in high-poverty counties. The lack of clear differences in effects across crime types limits any insights into the channels through which early childhood education influences later criminal behavior.

Racial differences in the likelihood of being eligible for or affected by each program, along with racial differences in the counterfactual options for early childhood education, suggest that the effects of Head Start and Smart Start may differ by race. In Table 4, we explore these heterogeneous effects by presenting our main estimates from Table 2 separately by race.[61] We find similar effects of Head Start availability for whites and non-whites, though only the reductions in adult convictions for non-whites in high-poverty counties are different from zero at a 10 percent significance level.[62] The lack of significant differences across races is somewhat surprising given that black children were more likely to be eligible for and enroll in Head Start. In the case of Smart Start, we find that funding exposure had significantly larger effects for non-whites, and again, we find that these effects are concentrated in high-poverty counties. Smart Start reduced non-white convictions by 20 percent overall and 27 percent in high poverty counties (Table 4). This result is even more dramatic in counties

[61] In Appendix Table A19 we explore heterogeneous effects by sex by presenting our main estimates from Table 2 separately by sex. We cannot rule out that in both programs, early childhood availability in high poverty counties resulted in similar percentage decreases in criminal behavior for both sexes.

[62] During this period in North Carolina, the black population comprised more than 95% percent of the non-white population (1970 Census).

without Head Start by 1980: in high poverty counties without prior Head Start availability, Smart Start reduced non-white convictions by 35 percent (Appendix Table A18). These results are consistent with a substantial racial gap in counterfactual early childhood education opportunities in high-poverty counties in North Carolina in the 1990s.

Ladd et al.'s (2014) estimates similarly suggest larger effects of Smart Start funding on the test scores of black children, although the implied differences are modest and the effects are entirely driven by differences in maternal education.[63] This lesser degree of heterogeneity in test scores is consistent with differential effects of early childhood education on the accumulation of cognitive versus non-cognitive skills. For example, evaluations of Perry Preschool, which enrolled disadvantaged black children, did not produce long run gains in IQ (particularly for males), but did produce persistent improvements in behavior.

5.3 Concerns about Migration out of North Carolina

One potential threat to the validity of our estimates relates to the data's coverage of convictions. While the data contain the universe of individuals convicted of a crime in North Carolina during this time period and allow us to link these individuals to their counties of birth, they are limited in that they do not allow us to observe convictions for individuals who are born in North Carolina and then leave the state. Fortunately, most individuals born in North Carolina remain there during adulthood; roughly 70% of individuals born in North Carolina during our sample periods reside there between the ages of 18 and 35. This share is even higher (roughly 80%) for those with the highest rates of criminal behavior (between ages 18 and 24, non-white, or with less than a high-school degree).

[63] Ladd et al. (2014) find no evidence of differential effects of Smart Start funding on the test scores of black children once they condition on maternal education and its interaction with Smart Start funding. However, they also show that the effects on test scores are larger for mothers with less than a high-school degree. The lower average education levels among black mothers in North Carolina implies larger effects for black children when not separately conditioning on maternal education and its interaction with Smart Start funding levels, but the implied difference in the effect of Smart Start on test score outcomes is modest.

If Head Start availability or Smart Start funding has differential effects on migration out of the state, it will not affect our estimates of convictions in North Carolina, but it may limit our ability to interpret them as representing an overall reduction in criminal behavior. A specific concern is that Head Start availability or Smart Start funding did not reduce criminal behavior, but simply increased the likelihood of leaving the state, leading us to misattribute the estimated reduction in the likelihood of criminal convictions.

In Appendix Table A20, we explore the relationship between measures of childhood Head Start availability (at the state of birth by birth cohort level) and the likelihood of living in one's state of birth. Across a variety of approaches and subsamples our estimates indicate a small and non-significant relationship between childhood Head Start availability and the likelihood of living in one's state of birth. Assuming similar patterns of criminality among North Carolina leavers and stayers, our upper bound estimate of additional migration can explain at most 5% of our estimated effect.[64] Even this upper bound is likely an overestimate as the mean rate of criminal conviction for movers to North Carolina (i.e., the equivalent of state of birth leavers) is lower than the rate for those born in North Carolina in our data. We have conducted similar analyses at the county-year birth cohort level using restricted-use American Community Survey (ACS) data.[65] While these results are undergoing the disclosure process, they are consistent with the estimates in Appendix Table A20.

Smart Start funding similarly appears to have no effect on migration out of North Carolina, at least as measured by administrative schooling data through grade 3 (Ladd et al., 2014). While we have been unable to determine a way to use publicly available data to test for differential migration at older ages (18 to 24), we have conducted similar analyses (as discussed above) with restricted-use ACS data that contain county of birth. While the

[64]Even assuming the largest estimated effect on migration, it would have to be the case that 65% of the marginal migrants were criminals to account for our estimates.

[65]These restricted data are available in the Federal Statistical Research Data Center (FSRDC) and allow us to directly observe county of birth for a large sample.

results are undergoing the disclosure process, they are consistent with the conclusion of Ladd et al. (2014) regarding the effect of Smart Starting funding on migration.[66]

Given the modest level of migration out of one's state of birth (and the lack of evidence of any differential out-migration), we view our estimates as providing a lower bound for the overall effect on criminal behavior.

5.4 Magnitude of Effect on Criminal Behavior

Our estimates suggest that the availability and funding of early childhood education causes substantial reductions in adult criminal behavior. Our preferred estimates from studying both Head Start and Smart Start indicate reductions in the likelihood of any serious conviction of approximately 1 percentage point (approximately 20 percent) for the average program in high poverty counties. While both programs generate similar reductions in serious criminal convictions in high poverty counties, Head Start is somewhat more expensive in real terms.[67] Scaling by the average cost of Head Start in high poverty counties and adjusting everything to 2015 dollars, we find that Head Start availability reduced the likelihood of a serious criminal conviction by 0.621 percentage points per $1,000 (in 2015 dollars), while Smart Start reduced the likelihood by roughly 0.720 percentage points per $1,000 (in 2015 dollars).[68] Alternatively put, Head Start and Smart Start cost about $1,500 per percentage point reduction in serious criminal convictions.

While the specific criminal outcome measures differ, the implied cost per percentage point of crime reduction for Head Start and Smart Start spending are similar to but somewhat smaller than those implied by evaluations of the Perry Preschool program: $1,800 per per-

[66] If anything, exposure to additional Smart Start funding in early childhood appears to make individuals less likely to leave North Carolina, suggesting that the magnitude of our estimated effects on overall crime may be slightly attenuated.

[67] Around $2,000 per individual in the cohort for Head Start and around $1,600 per fully treated individual for Smart Start in high poverty counties (all in 2015 dollars).

[68] The Smart Start treatment is denoted in nominal dollars so we adjust to $1,000 in 2015 dollars.

centage point *arrest* reduction (authors' calculations from Heckman et al. (2010) estimates). When we scale arrests to convictions, we find that this cost-effectiveness advantage for Head Start and Smart Start increases.[69]

5.5 Quantifying the Benefits

How do the future benefits of crime reduction compare to the costs of these programs? To enable this comparison, we use the detailed conviction histories of offenders in North Carolina to construct back-of-the-envelope measures of the discounted social cost of their crimes and then estimate the effect of Head Start and Smart Start on these measures. We construct this measure by applying the social cost estimates for the given crime from McCollister et al (2010) and Miller et al. (1996) to each conviction by age 24 (Smart Start) or 35 (Head Start), and then discounting that cost from the age at conviction back to age four (for ease of comparison with the program costs).[70] We then aggregate this social cost measure to the birth cohort level. This measure serves as a lower bound for social cost of a cohort's crimes as it includes only crimes with a social cost estimate available and does not account for the fact that multiple offenses are often associated with a single conviction. To address the latter issue, we construct a second measure using a similar approach, but where we weight each conviction by the average number of offenses per conviction for each crime type.[71] For this measure we are limited to the set of crimes for which we can observe the number of offenses in FBI's Uniform Crime Reports (i.e. Part 1 crimes). In all cases, we limit our

[69] Another approach to comparing magnitudes is to calculate implied treatment-on-the-treated (TOT) effects of each intervention. While this isn't possible in the case of Smart Start, our Head Start estimates imply treatment-on-the-treated (TOT) effects of 6 to 9 percentage points, somewhat smaller than those reported in evaluations of the Perry Preschool program for somewhat similar measures (See Appendix D for additional discussion.)

[70] When cost estimates for the same crime are available in McCollister et al (2010) and Miller et al. (1996), we use the more recent estimates from McCollister et al (2010). Following the baseline scenario in Hendren et al. (2020), we use a 3 percent discount rate.

[71] Weights are calculated as the number of offenses reported in North Carolina from 1980-2010 (FBI Uniform Crime Reports) divided by the number of convictions in North Carolina from 1980-2010, for a given type of crime.

estimates to convictions observed within our data (i.e., by age 35 for Head Start cohorts and age 24 for Smart Start cohorts); expected benefits are likely substantially larger if we included forecasted commission of crime across ages.

Table 5 shows the Equation 1 estimates of the effects of early childhood education programs on these discounted social cost of crime measures, where each cell represents the result from a different regression. The estimates are imprecise and based on imperfect measures, but the magnitudes suggest that, particularly in high poverty counties, the crime benefits of these programs make up a large proportion of the cost of the programs when accounting for multiple offenses per conviction.[72] For example, Column 2, Row 2 of Table 5 shows that Head Start availability caused a $2,348 (16%) reduction in discounted social crime costs in high poverty counties. Comparing the magnitudes of the Head Start estimates in column (2) to the average cost of the program per child (both participants and non-participants) of $886 in all counties and $2,044 in high poverty counties suggests the crime benefits of the program account for 26% of the costs in all counties and 115% of the costs in high poverty counties.[73] Performing a similar comparison between the magnitudes of the Smart Start estimates in column (4) with the $1,000 cost per child suggests that crime benefits through age 24 account for 79% of the costs of the program in all counties and 93% in high poverty counties.[74] These results support the conclusion that individual investments in early childhood education are inefficiently low without subsidies, particularly given the large impacts of these programs on other outcomes with substantial social benefits (e.g. education).

[72]Effects are substantially larger when including costs to victims rather than only supervision costs. Supervision only costs are smaller for the Smart Start sample than the Head Start sample, reflecting both the differences in age ranges over which these costs are calculated and the shift from incarceration to probation for these cohorts.

[73]2015 dollars are used in all estimates/calculations.

[74]These estimates are based on convictions through age 24. If we scale by the ratio of Head Start benefits at age 35 versus 24 (1.5 to 2.5), crime benefits generally exceed the costs.

6 Conclusion

We contribute to the sparse literature on the developmental factors that influence an individual's likelihood of becoming a criminal by exploring the effect of early childhood education on criminal behavior. This relationship has become increasingly relevant given recent expansions in the share of children attending public preschools and widespread efforts to improve the quality of early childhood education. These expansions have been driven in large part by policymakers' belief that early childhood education has large impacts later in life. Crime reduction in particular is central to the widely publicized benefit-cost analyses of these programs (e.g., crime reduction accounts for 40-65% of the benefits estimated in the context of Perry Preschool). However, inconsistent findings across similar studies, limited statistical power with correspondingly high implied false discovery rates, and differences in contexts across decades challenge whether we can generalize the crime effect estimates from earlier studies to the present day.

We bring new evidence to this literature by using administrative crime data to investigate the effects of two large-scale early childhood education programs operating in two different time periods. This approach yields substantial external validity improvements over prior long-run crime effect estimates by relying on a much broader treated population and enabling a direct comparison of similar adult crime reductions from two different interventions implemented in very different contexts. Our administrative crime data grants us significant improvements in sample size (hundreds of thousands versus a hundred or thousands) and corresponding increases in precision. This precision reduces the likelihood that any statistically significant estimated effect is a false positive (false discovery rate), reduces the expected upward bias in reported effects, and analogously increases precision. This precision also allows us to explore heterogeneity in effects across areas with different levels of disadvantage and access to existing care, potentially informing the debate between universal and more

targeted early education policies. While the precision and five-decade-long time span of these data are substantial advantages, they come at the cost of limiting our analysis to the state of North Carolina.[75]

We show that early childhood education reduces adult criminal behavior across two different programs and time periods. We find that Head Start availability in the 1960s and 1970s reduces the likelihood of a serious conviction by age 35 by 1.3 percentage points in high-poverty counties, but has no measurable effect in low-poverty counties. Implemented two to three decades later, Smart Start generates similar reductions in adult criminal behavior, with effects similarly concentrated in high-poverty counties and among black children. Taken together, these results suggest the general capacity of early childhood education to reduce the propensity for criminal behavior in those areas and among those individuals with the fewest resources. We also find that Smart Start's effects are largest in counties without Head Start access, which suggests that there are diminishing returns to early childhood education funding. Both patterns of results are consistent with the general pattern of effects in the literature, though we cannot draw strong conclusions from a small set of studies that identify heterogeneous parameters of different programs.

Also consistent with prior literature on earlier cohorts, we find that the effects of a more modern program on cognitive skills are not predictive of effects on longer-run behavioral outcomes. Indeed, our estimated effects of Smart Start on criminal behavior are more than 7 times the effect implied by earlier estimates of the program's test score effects applied to the conditional correlation between test scores and crime. Future work might explore whether effects on non-cognitive measures may be more predictive of effects on this margin.

Finally, we take advantage of the greater sample sizes and detail contained in the data

[75] While North Carolina is the 9th most populous state and has a similar age distribution (and percentage of young children) as the rest of the country, the state has a much higher percentage of black individuals. This is advantageous for studying the effects of early childhood education on the criminality of black individuals (who are both more likely to be affected by these programs and more likely to be convicted of a crime) but may limit the generalizability of the resulting estimates to states with smaller minority populations.

to show that the discounted benefits generated by early childhood education's later crime reduction are substantial relative to the costs of the education provided. This is especially noteworthy considering that later crime reduction was not the stated objective of either program and that these benefits likely accrue in large part to those who did not experience the program themselves. The magnitude of these external benefits implies that the social benefits of early childhood education vastly exceed the private benefits. That said, these benefits accrue disproportionately to high-poverty areas and to those lacking access to other early childhood education subsidies. Taken together, our results provide evidence in support of recent efforts to expand and improve early childhood education, but point to large potential gains from targeting these efforts toward areas with the greatest need.

References

Anderson, Michael L, "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*, 2008, *103* (484), 1481–1495.

Bailey, Martha J and Andrew Goodman-Bacon, "The War on Poverty's experiment in public medicine: Community health centers and the mortality of older Americans," *American Economic Review*, 2015, *105* (3), 1067–1104.

Barr, Andrew and Alex Smith, "Fighting Crime in the Cradle: The Effects of Early Childhood Access to Nutritional Assistance," Technical Report 2018.

_ **and Chloe Gibbs**, "The Longer Long-term Impact of Head Start: Intergenerational Transmission of Program Effects," Working Paper 2017.

Bayer, Patrick, Randi Hjalmarsson, and David Pozen, "Building criminal capital behind bars: Peer effects in juvenile corrections," *The Quarterly Journal of Economics*, 2009, *124* (1), 105–147.

Campbell, Frances A, Craig T Ramey, Elizabeth Pungello, Joseph Sparling, and Shari Miller-Johnson, "Early childhood education: Young adult outcomes from the Abecedarian Project," *Applied developmental science*, 2002, *6* (1), 42–57.

_ **, Elizabeth P Pungello, Margaret Burchinal, Kirsten Kainz, Yi Pan, Barbara H Wasik, Oscar A Barbarin, Joseph J Sparling, and Craig T Ramey**, "Adult outcomes as a function of an early childhood educational program: an Abecedarian Project follow-up.," *Developmental psychology*, 2012, *48* (4), 1033.

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

Census Bureau, "Cartographic Boundary Files - Shapefile," 2021. Data retrieved from https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html in May of 2016.

Chaiken, Jan M and Marcia R Chaiken, *Varieties of criminal behavior*, Rand Corporation, 1982.

de Chaisemartin, Clément, Xavier D'Haultfoeuille, and Yannick Guyonvarch, "DID_MULTIPLEGT: Stata module to estimate sharp Difference-in-Difference designs with multiple groups and periods," 2018.

Deming, David, "Early childhood intervention and life-cycle skill development: Evidence from Head Start," *American Economic Journal: Applied Economics*, 2009, pp. 111–134.

_ , "Better schools, less crime?," *The Quarterly Journal of Economics*, 2011, *126* (4), 2063–2115.

Doyle, Joseph J, "Child protection and child outcomes: Measuring the effects of foster care," *American Economic Review*, 2007, *97* (5), 1583–1610.

_ , "Child protection and adult crime: Using investigator assignment to estimate causal effects of foster care," *Journal of political Economy*, 2008, *116* (4), 746–770.

Farrington, David P, "Methodological quality and the evaluation of anti-crime programs," *Journal of Experimental Criminology*, 2006, *2* (3), 329–337.

Fitzpatrick, Maria D, "Starting school at four: The effect of universal pre-kindergarten on children's academic achievement," *The BE Journal of Economic Analysis & Policy*, 2008, *8* (1).

Fitzpatrick, Maria Donovan, "Preschoolers enrolled and mothers at work? The effects of universal prekindergarten," *Journal of Labor Economics*, 2010, *28* (1), 51–85.

Garces, Eliana, Duncan Thomas, and Janet Currie, "Longer-term effects of Head Start," *American economic review*, 2002, *92* (4), 999–1012.

García, Jorge Luis, James J Heckman, and Anna L Ziff, "Early childhood education and crime," *Infant mental health journal*, 2019, *40* (1), 141–151.

Gibbs, Chloe, Jens Ludwig, and Douglas L Miller, "Does Head Start do any lasting good?," Working Paper, National Bureau of Economic Research 2011.

Goodman-Bacon, Andrew, "Difference-in-differences with variation in treatment timing," Technical Report, National Bureau of Economic Research 2018.

Head Start Funding Data, "Head Start Appropriations," 2021. Data accessed from https://catalog.archives.gov/ using National Archives identifiers 604417 and 599052.

Heckman, James J, Jora Stixrud, and Sergio Urzua, "The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior," *Journal of Labor economics*, 2006, *24* (3), 411–482.

_ , Rodrigo Pinto, and Peter Savelyev, "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes," *American Economic Review*, 2013, *103* (6), 2052–86.

_ , _ , Azeem M Shaikh, and Adam Yavitz, "Inference with imperfect randomization: The case of the Perry Preschool Program," Technical Report, National Bureau of Economic Research 2011.

_ , **Seong Hyeok Moon, Rodrigo Pinto, Peter A Savelyev, and Adam Yavitz**, "The rate of return to the HighScope Perry Preschool Program," *Journal of public Economics*, 2010, *94* (1), 114–128.

Hendren, Nathaniel and Ben Sprung-Keyser, "A unified welfare analysis of government policies," *The Quarterly Journal of Economics*, 2020, *135* (3), 1209–1318.

Hill, Patrick L, Brent W Roberts, Jeffrey T Grogger, Jonathan Guryan, and Karen Sixkiller, "Decreasing delinquency, criminal behavior, and recidivism by intervening on psychological factors other than cognitive ability: A review of the intervention literature," Technical Report, National Bureau of Economic Research 2011.

Hindelang, Michael J, Travis Hirschi, and Joseph G Weis, *Measuring delinquency*, Sage Publications Beverly Hills, 1981.

Hoynes, Hilary, Diane Whitmore Schanzenbach, and Douglas Almond, "Long-run impacts of childhood access to the safety net," *American Economic Review*, 2016, *106* (4), 903–34.

Hoynes, Hilary W. and Diane Whitmore Schanzenbach, "Consumption Responses to In-Kind Transfers: Evidence from the Introduction of the Food Stamp Program," *American Economic Journal: Applied Economics*, 2009, *1* (4), 109–39.

Ioannidis, John, T Stanley, and Chris Doucouliagos, "The Power of Bias in Economics Research," *Economic Journal*, 2017, *127* (605), F236–F265.

Jackson, C Kirabo, "What do test scores miss? The importance of teacher effects on non–test score outcomes," *Journal of Political Economy*, 2018, *126* (5), 2072–2107.

Johnson, Rucker C and C Kirabo Jackson, "Reducing inequality through dynamic complementarity: Evidence from Head Start and public school spending," Technical Report, National Bureau of Economic Research 2017.

Katz, Lawrence F, Jeffrey R Kling, and Jeffrey B Liebman, "Moving to opportunity in Boston: Early results of a randomized mobility experiment," *The Quarterly Journal of Economics*, 2001, *116* (2), 607–654.

Kids Counts Database, "Kids Count Data," 2021. Data retrieved from https://datacenter.kidscount.org/data in June of 2019.

Kline, Patrick and Christopher R. Walters, "Evaluating public programs with close substitutes: The case of Head Start," *Quarterly Journal of Economics*, 2016, *131(4).*

Kling, Jeffrey R, Jens Ludwig, and Lawrence F Katz, "Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment," *The Quarterly Journal of Economics*, 2005, *120* (1), 87–130.

Ladd, Helen F, Clara G Muschkin, and Kenneth A Dodge, "From birth to school: Early childhood initiatives and third-grade outcomes in North Carolina," *Journal of Policy Analysis and Management*, 2014, *33* (1), 162–187.

List, John A, Fatemeh Momeni, and Yves Zenou, "Are estimates of early education programs too pessimistic? Evidence from a large-scale field experiment that causally measures neighbor effects," 2019.

Lochner, Lance and Enrico Moretti, "The effect of education on crime: Evidence from prison inmates, arrests, and self-reports," *The American Economic Review*, 2004, *94* (1), 155–189.

Ludwig, Jens and Douglas Miller, "Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design," *The Quarterly Journal of Economics*, 2007, *122* (1), 159–208.

_ **and Jeffrey R Kling**, "Is crime contagious?," *The Journal of Law and Economics*, 2007, *50* (3), 491–518.

McCollister, Kathryn E, Michael T French, and Hai Fang, "The cost of crime to society: New crime-specific estimates for policy and program evaluation," *Drug & Alcohol Dependence*, 2010, *108* (1), 98–109.

Miller, Douglas L, Na'ama Shenhav, and Michel Z Grosz, "Selection into identification in fixed effects models, with application to Head Start," Technical Report, National Bureau of Economic Research 2019.

North Carolina Convictions Data, "North Carolina Department of Public Safety," 2020. Data accessed from https://webapps.doc.state.nc.us/opi/downloads.do?method=view in June of 2020.

North Carolina Demographic Data, "North Carolina Office of State Budget and Management," 2020. Data accessed from https://www.osbm.nc.gov/content/july-1-2010-county-total-age-groups-standard in June of 2019.

North Carolina Partnership for Children, "Smart Start Funding." Data shared by North Carolina Partnership for Children in April of 2019.

Office of Child Development, "Project Head Start 1965-1967: A descriptive report of programs and participants," Report, Department of Health, Education, and Welfare, Washington, DC 1968.

_ , "Project Head Start 1968: A descriptive report of programs and participants," Report, Department of Health, Education, and Welfare, Washington, DC 1970.

Olds, David, Charles R Henderson Jr, Robert Cole, John Eckenrode, Harriet Kitzman, Dennis Luckey, Lisa Pettitt, Kimberly Sidora, Pamela Morris, and Jane Powers, "Long-term effects of nurse home visitation on children's criminal and antisocial behavior: 15-year follow-up of a randomized controlled trial," *Jama*, 1998, *280* (14), 1238–1244.

Olds, David L, Lois Sadler, and Harriet Kitzman, "Programs for parents of infants and toddlers: recent evidence from randomized trials," *Journal of child psychology and psychiatry*, 2007, *48* (3-4), 355–391.

Ruggles, Steven, Sarah Flood, Ronald Goeken, Jose Pacas, Megan Schouweiler, and Matthew Sobek, "Integrated Public Use Microdata Series, Decennial Census," 2021. Data retrieved from https://doi.org/10.18128/D010.V11.0 in June of 2019.

Sanbonmatsu, Lisa, Lawrence F Katz, Jens Ludwig, Lisa A Gennetian, Greg J Duncan, Ronald C Kessler, Emma K Adam, Thomas McDade, and Stacy T Lindau, "Moving to opportunity for fair housing demonstration program: Final impacts evaluation," 2011.

Stuart, Elizabeth A, "Matching methods for causal inference: A review and a look forward," *Statistical science: a review journal of the Institute of Mathematical Statistics*, 2010, *25* (1), 1.

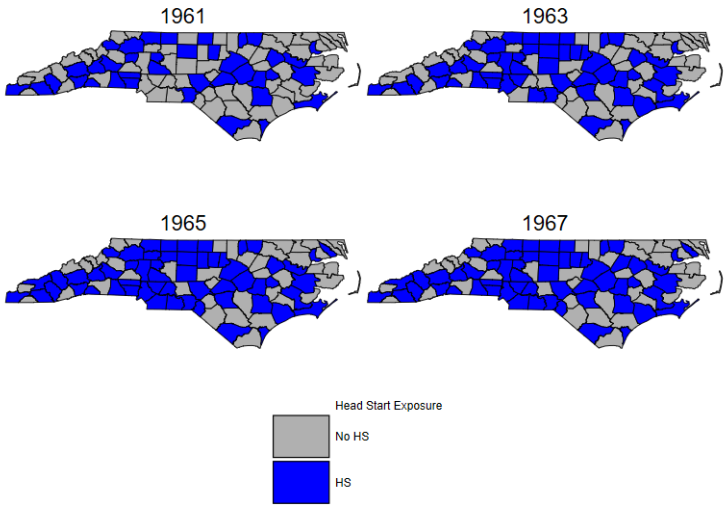
Sun, Liyang, "EVENTSTUDYWEIGHTS: Stata module to estimate the implied weights on the cohort-specific average treatment effects on the treated (CATTs)(event study specifications)," 2020.

_ **and Sarah Abraham**, "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*, 2020.

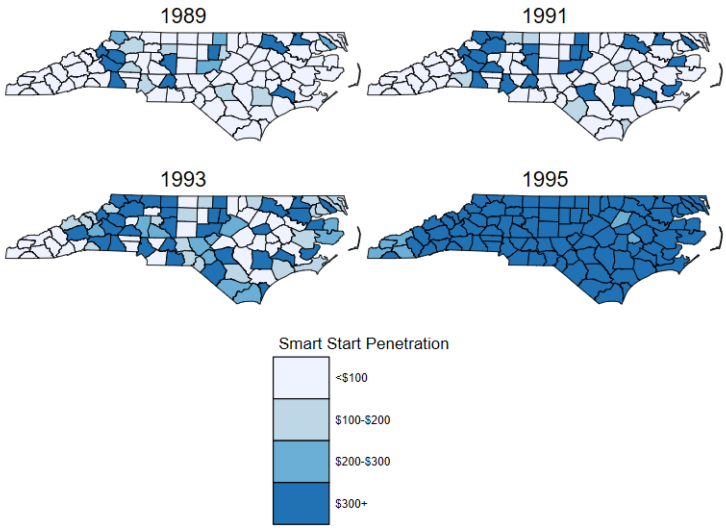
Thompson, Owen, "Head Start's Long-Run Impact: Evidence from the Program's Introduction," *Journal of Human Resources*, 2017, pp. 0216–7735r1.

United States Department of Agriculture, "SNAP Data Tables," 2021. Data retrieved from https://www.fns.usda.gov/pd/supplemental-nutrition-assistance-program-snap in June of 2019.

Figure 1: County by Birth Cohort Early Childhood Program Rollout in North Carolina



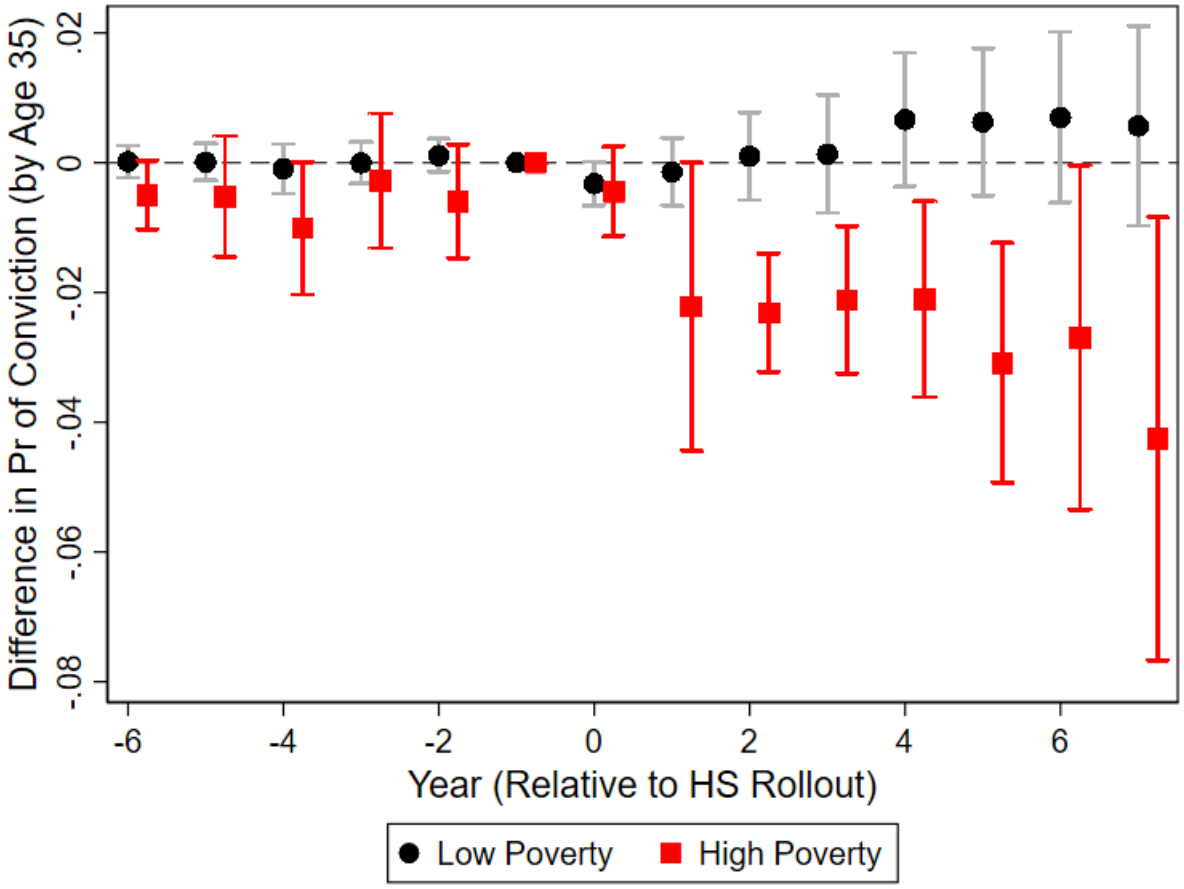
(a) Head Start Rollout



(b) Smart Start Rollout

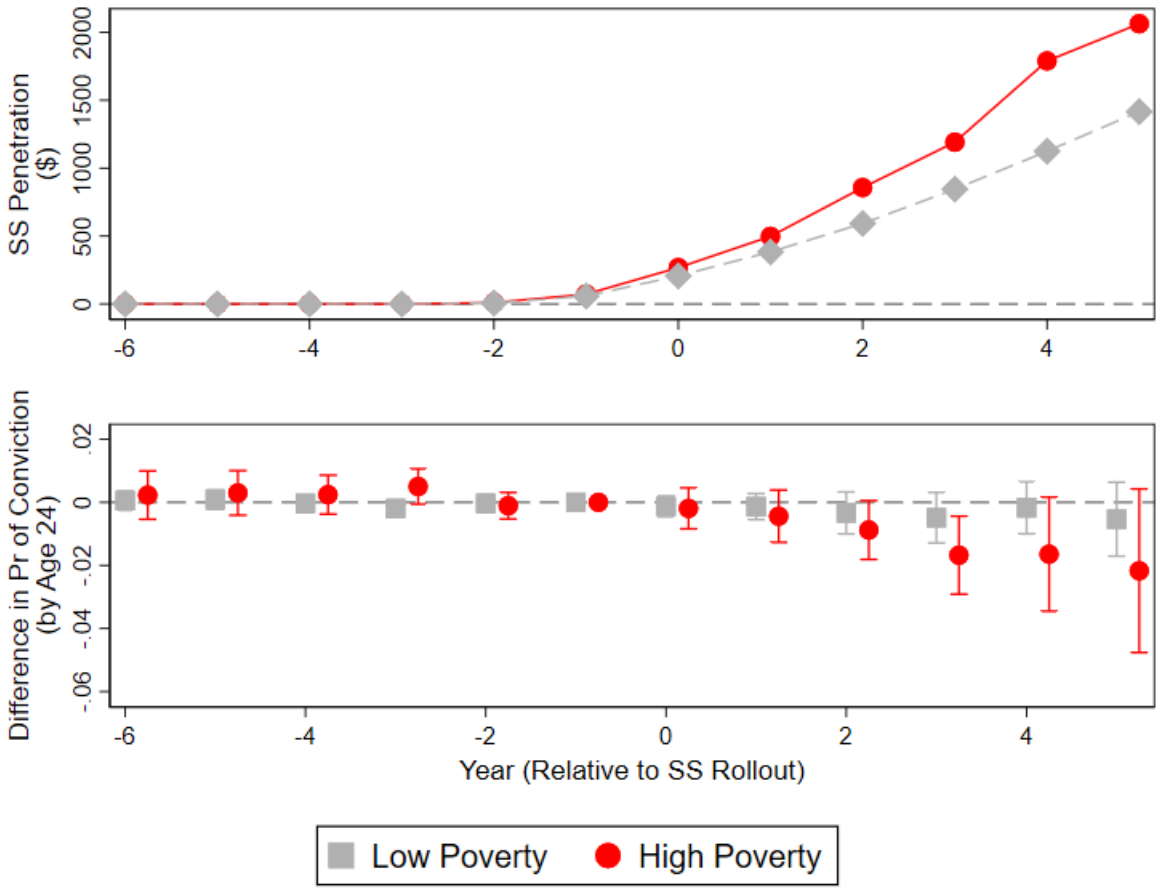
Note: Figure shows which birth-cohorts born to which counties had Head Start and Smart Start available to them in North Carolina. Panel (a): Prior to the 1961 birth cohort no counties had Head Start available. Head Start availability is identified from county by year level Head Start funding data following Barr and Gibbs (2017). Head Start funding levels are obtained from Head Start Historical Records. Panel (b): Prior to the 1989 birth cohort no counties had Smart Start funding penetration. Smart Start penetration is defined following Ladd et al. (2014) and depicted in nominal terms. Smart Start funding levels are obtained from the Smart Start organization. See the text or additional details.

Figure 2: Event Study of Head Start's Impact on Criminal Conviction



Note: Figure shows the coefficient estimates and 95% confidence interval from estimating Equation 2 separately for high and low poverty counties. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1955. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 35. UCR Part 1 crimes include violent crimes (those in which the description of the offense contains the words "murder", "assault", or "robbery" (rape not being included), and property crimes (those in which the description of the offense contains the words "burglary" or "larceny"). All specifications include birth county and birth-cohort fixed effects as well as 1960 county characteristics interacted with a time trend in birth cohort. 1960 county characteristics include: percent of land in farming, percent of people living in families with less than $3,000, percent of population in urban area, percent black, percent less than age 5, percent greater than age 65, and percent of employment in agriculture. Those counties whose poverty rate in 1960 was above the median in North Carolina (40.2% poverty) are called "High Poverty", while those below the median are called "Low Poverty". The sample is restricted to counties that ever received Head Start between 1965 and 1976. The sample is further restricted to cohorts who were born between 1955 and 1968.

Figure 3: Event Study of Smart Start's Impact on Criminal Conviction



Note: In the top panel, figure shows trends in Smart Start funding penetration in nominal dollars defined following Ladd et al. (2014), separately for high and low poverty counties. In the bottom panel, figure shows the coefficient estimates and 95% confidence interval from estimating Equation 2 separately for high and low poverty counties. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1980. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 24. UCR Part 1 crimes include violent crimes (those in which the description of the offense contains the words "murder", "assault", or "robbery" (rape not being included), and property crimes (those in which the description of the offense contains the words "burglary" or "larceny"). All specifications include birth county and birth-cohort fixed effects. Binary Smart Start availability, the independent variable of interest, is defined as Smart Start penetrationabove the 25th percentile of penetration. Those counties whose poverty rate in 1980 was above the median in North Carolina (17.3% poverty) are called "High Poverty", while those below the median are called "Low Poverty". The sample is restricted to cohorts who were born between 1980 and 1994.

Table 1: Descriptive Statistics

	All	High Poverty	Low Poverty
Panel A: Head Start Sample			
First Cohort With Head Start	1962.3	1962.3	1962.3
HS Funding (2015$ per 4 year old)	893.8	2061.1	605.7
Criminal Conviction	0.0469	0.0462	0.0471
Observations (cells)	882	308	574
Individuals Represented	1487225	444848	1042377
Panel B: Smart Start Sample			
First Calendar Year of Smart Start	1995.5	1996.7	1995.0
SS Penetration (2015$)	818.9	750.0	838.3
Criminal Conviction	0.0516	0.0512	0.0517
Observations (cells)	1500	750	750
Individuals Represented	1407042	666073	740969

Note: Panel A contains summary statistics of crime outcome variables for the sample of birth cohorts born from 1955 to 1968, which is used in the Head Start analysis. Each observation (cell) is at the county birth-cohort level. Head Start availability and funding are reported in the first two rows. Funding levels are given in 2015 dollars and averaged over exposed county-cohorts only, so that only non-zero values are included. Criminal conviction by adulthood is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 35. UCR Part 1 crimes include violent crimes (those in which the description of the offense contains the words "murder", "assault", or "robbery" (rape not being included), and property crimes (those in which the description of the offense contains the words "burglary" or "larceny"). All variables are further broken down by county level poverty status. Those counties whose poverty rate in 1960 was above the median in North Carolina (40.2% poverty) are called "High Poverty", while those below the median are called "Low Poverty". The sample is restricted to counties that ever received Head Start between 1965 and 1976. Panel B contains summary statistics of crime outcome variables for the sample of birth cohorts born from 1980 to 1994, which is used in the Smart Start analysis. Each observation (cell) is at the county birth-cohort level. Smart Start availability and funding penetration are reported in the first two rows of Panel B. Penetration measures are defined following Ladd et al. (2014) reported in 2015 dollars and averaged over exposed county-cohorts only, so that only non-zero values are included. Criminal conviction by adulthood is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 24. All variables are further broken down by county level poverty status. Those counties whose poverty rate in 1980 was above the median in North Carolina (17.3% poverty) are called "High Poverty", while those below the median are called "Low Poverty". Data sources are the NC Department of Corrections, the NC Department of Corrections, Head Start Historical Records, and Smart Start Records.

Table 2: Effect of Early Childhood Education on Criminal Conviction

	(1)	(2)	(3)
	All	High Poverty	Low Poverty
Panel A: Head Start			
Head Start Availability	-0.0017	-0.0128**	0.0026
	(0.0031)	(0.0058)	(0.0033)
Observations	882	308	574
Mean	0.0469	0.0462	0.0471
Panel B: Smart Start			
SS ($1000s)	-0.0064**	-0.0118**	-0.0030
	(0.0029)	(0.0051)	(0.0035)
Observations	1500	750	750
Mean	0.0516	0.0512	0.0517

Note: Each cell reports a separate OLS regression with standard errors clustered at the birth county level and reported in parentheses. All specifications include birth county and birth-cohort fixed effects. Panel A reports results using the Head Start sample. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1955. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 35. The reported variable of interest is an indicator for whether Head Start was available to a given county birth cohort. Panel B reports results using the Smart Start sample. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1980. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 24. The reported variable of interest is a measure of Smart Start funding penetration for a given county birth cohort, constructed following Ladd et al. (2014). See the notes to Table 1 for additional sample restrictions and definitions. Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$).

Table 3: Effect of Smart Start Funding on Criminal Conviction - By Presence of Head Start

	(1)	(2)	(3)	(4)	(5)	(6)
	All			High Poverty		
SS ($1000s)	-0.0064**	-0.0146***	-0.0040	-0.0118**	-0.0156***	-0.0015
	(0.0029)	(0.0050)	(0.0032)	(0.0051)	(0.0055)	(0.0033)
Observations	1500	555	945	750	435	315
Mean	0.0516	0.0528	0.0513	0.0512	0.0532	0.0494
Head Start	All	No Head Start	Head Start	All	No Head Start	Head Start

Note: Each column reports a separate OLS regression with standard errors clustered at the birth county level and reported in parentheses. All specifications include birth county and birth-cohort fixed effects. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1980. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of either UCR Part 1 property crimes or Part 1 violent crimes in North Carolina by age 24. The reported variable of interest is a measure of Smart Start funding penetration for a given county birth cohort, constructed following Ladd et al. (2014). See the notes to Table 1 for additional sample restrictions and definitions. Columns (2) and (5) further restrict the sample to counties without a Head Start program by 1980, while columns (3) and (6) restrict to counties with a Head Start program by 1980. Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$).

Table 4: Effect of Early Childhood Education on Criminal Conviction- By Race

	(1)	(2)	(3)
	All	High Poverty	Low Poverty
Panel A: Head Start			
White			
Head Start Availability	-0.0038	-0.0105	0.0001
	(0.0039)	(0.0069)	(0.0040)
Observations	667	252	415
Mean	0.0267	0.0254	0.0271
Non-White			
Head Start Availability	0.0014	-0.0119*	0.0051
	(0.0067)	(0.0062)	(0.0106)
Observations	667	252	415
Mean	0.1008	0.0737	0.1091
Panel B: Smart Start			
White			
SS ($1000s)	-0.0026	-0.0029	-0.0012
	(0.0020)	(0.0053)	(0.0022)
Observations	1329	674	655
Mean	0.0315	0.0272	0.0328
Non-White			
SS ($1000s)	-0.0191***	-0.0216***	-0.0128
	(0.0064)	(0.0057)	(0.0122)
Observations	1329	674	655
Mean	0.0948	0.0805	0.1061

Note: Each cell reports a separate OLS regression with standard errors clustered at the birth county level and reported in parentheses. All specifications include birth county and birth cohort fixed effects. Panel A reports results using the Head Start sample for white cohorts and non-white cohorts separately. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1955. The dependent variable is the fraction of individuals of a given race in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 35. (Sample sizes are smaller for these specifications because the natality files for 25% of counties in North Carolina do not have race breakdowns before 1969 and we do not know the race of approximately 13% of births in our sample.) The reported variable of interest is an indicator for whether Head Start was available to a given county birth cohort. Panel B reports results using the Smart Start sample for white cohorts and non-white cohorts separately. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1980. The dependent variable is the fraction of individuals of a given race in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 24. The reported variable of interest is a measure of Smart Start funding penetration for a given county birth cohort, constructed following Ladd et al. (2014). (Sample sizes are smaller for these specifications because from 1989 to 1993 the natality files for 25% of counties in North Carolina do not have race breakdowns. For these years, race is available only for counties in which 1980 populations for the non-white group formed at least 10 percent of the total population or numbered at least 10,000.) See the notes to Table 1 for additional sample restrictions and definitions. Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$).

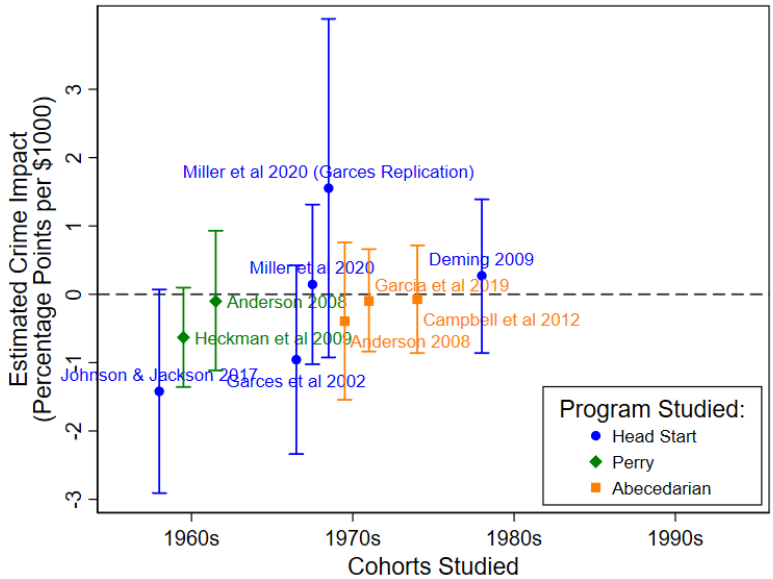
Table 5: Social Benefits of Crime Reduction from Early Childhood Education

	HS Availability		SS ($1,000)	
	All Crimes	Part 1 (Offense-Weighted)	All Crimes	Part 1 (Offense-Weighted)
	(1)	(2)	(3)	(4)
All Counties	-129.836	-226.994	-170.869**	-785.579
	(174.046)	(993.404)	(80.533)	(550.538)
Mean	3128.42	13295.42	2050.44	11521.31
High Poverty Counties	-537.455	-2348.500	-133.408	-930.291
	(365.998)	(2311.682)	(103.431)	(920.457)
Mean	3243.10	14787.51	1899.64	11135.20
Low Poverty Counties	55.537	790.803	-141.017	-609.216
	(170.637)	(874.318)	(103.420)	(707.764)
Mean	3096.28	12877.19	2100.52	11649.52

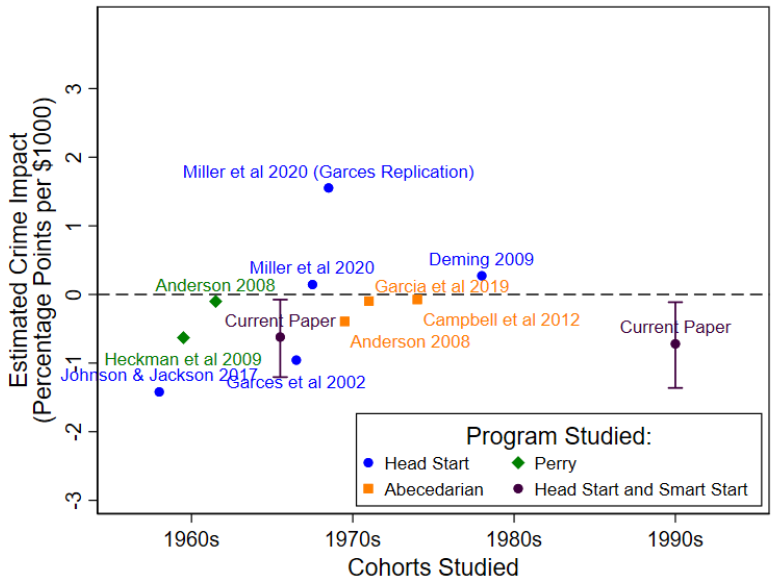
Note: Each cell reports a separate OLS regression with standard errors clustered at the birth county level and reported in parentheses. All specifications include birth county and birth cohort fixed effects. In columns 1 and 3, the dependent variable is a measure of discounted social cost of crime constructed by applying the social cost estimates for each crime from McCollister et al (2010) and Miller et al. (1996) to each conviction by age 24 (Smart Start) or 35 (Head Start), and then discounting that cost from the age at conviction back to age four using a 3 percent discount rate. In columns 2 and 4, the dependent variable is a similar measure constructed for UCR Part 1 crimes only that adjusts for the average number of offenses per conviction (1980-2010) using counts from North Carolina UCR known offenses and NC Department of Public Safety data. (for ease of comparison with the program costs). Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1955 (Head Start) or 1980 (Smart Start). All values are converted to 2015 dollars. For Head Start, the reported variable of interest is an indicator for whether the program was available to a given county birth cohort. For Smart Start, the reported variable of interest is a measure of Smart Start funding penetration for a given county birth cohort, constructed following Ladd et al. (2014). Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$).

Appendix A: Supplementary Figures

Figure A1: Effect Size Comparison



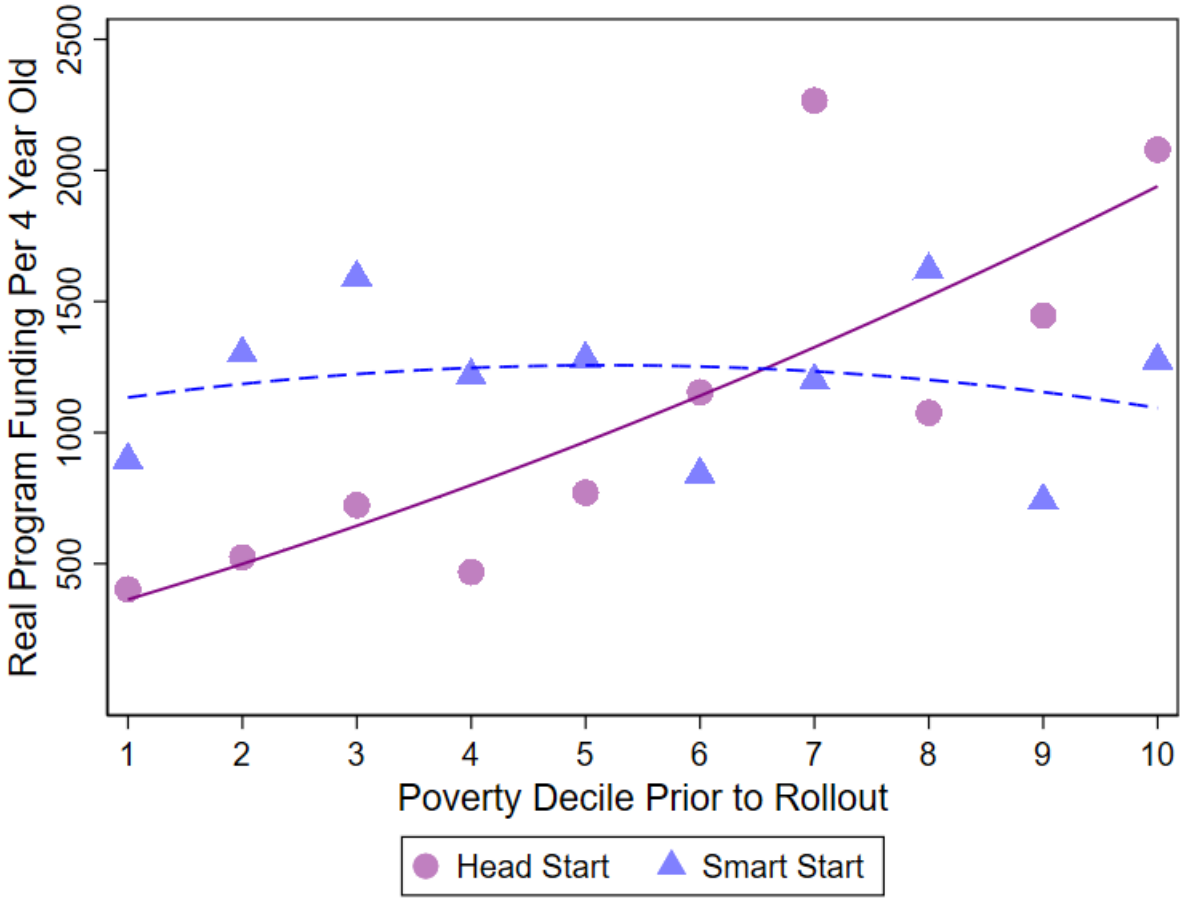
(a) Existing Evidence



(b) New Evidence

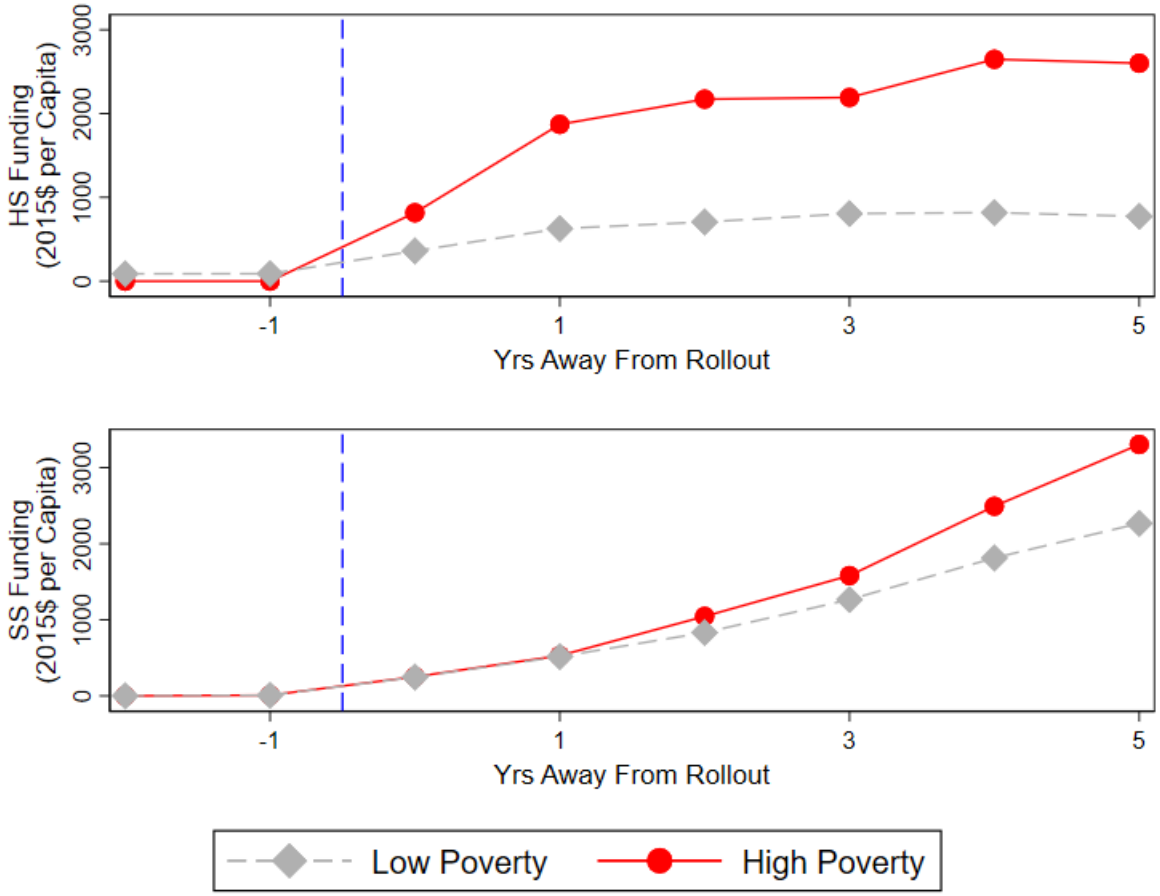
Note: Figure shows estimated crime impact per $1,000 (2015 dollars) and associated 95 percent confidence intervals. Program costs for Perry and Abecedarian are taken from Heckman et al (2009) and Garces et al (2002), respectively, and are $20,648 and $22,687 respectively. Because program funding for Head Start varies across cohorts and different papers study different cohorts, we use the funding reported by each study. For Deming (2009), Garces et al. (2002) and Johnson and Jackson (2017) those numbers are, respectively, $6,981, $5,540, and $6,027. The selected effect estimates are provided in column (4) of Table B2. See Appendix B for additional details on construction.

Figure A2: Targeting of Head Start and Smart Start at Poverty



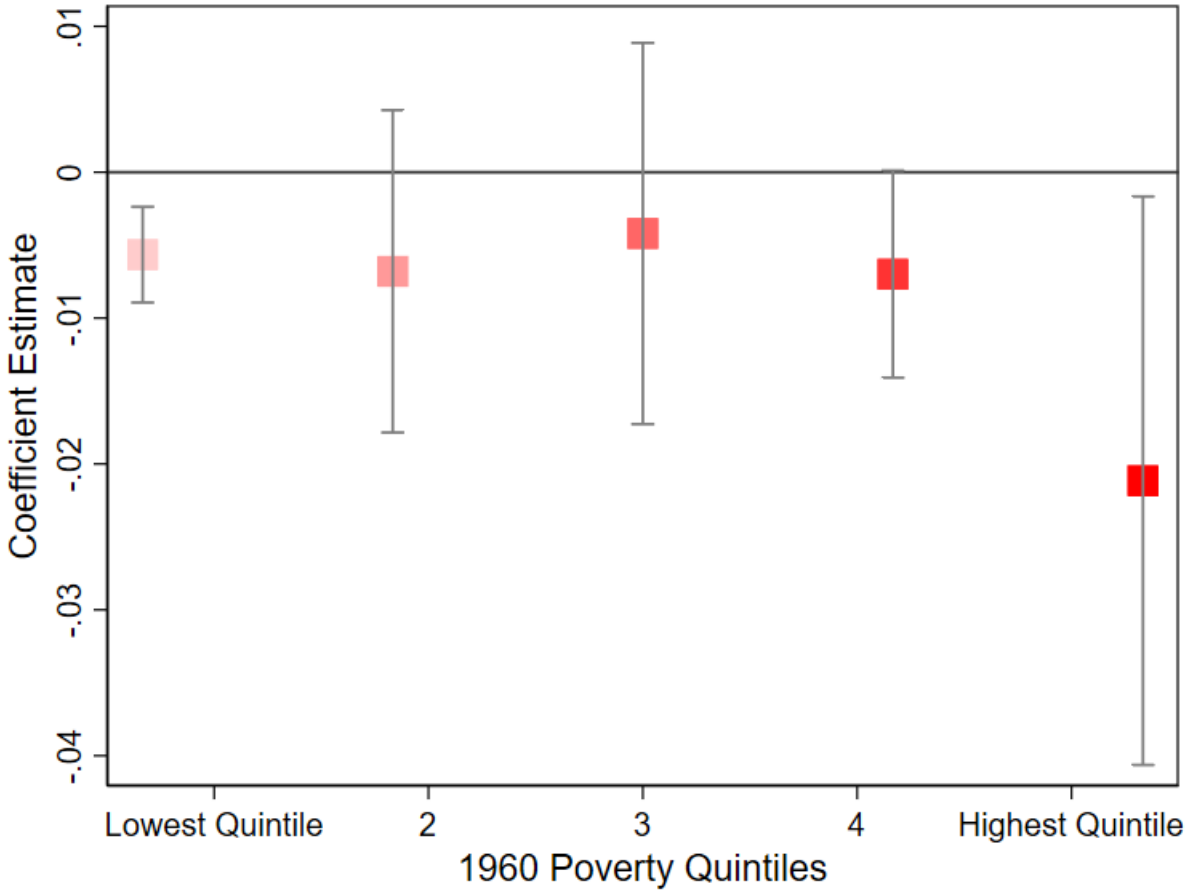
Note: Figure shows per capita county level Head Start and Smart Start funding (given in $ per 4 year olds) by county poverty deciles before program rollout. For Head Start 1960 poverty deciles are used, while for Smart Start 1980 poverty deciles are used. All values are in 2009 dollars. The sample is restricted to counties with nonzero funding amounts.

Figure A3: Funding By County Poverty Level



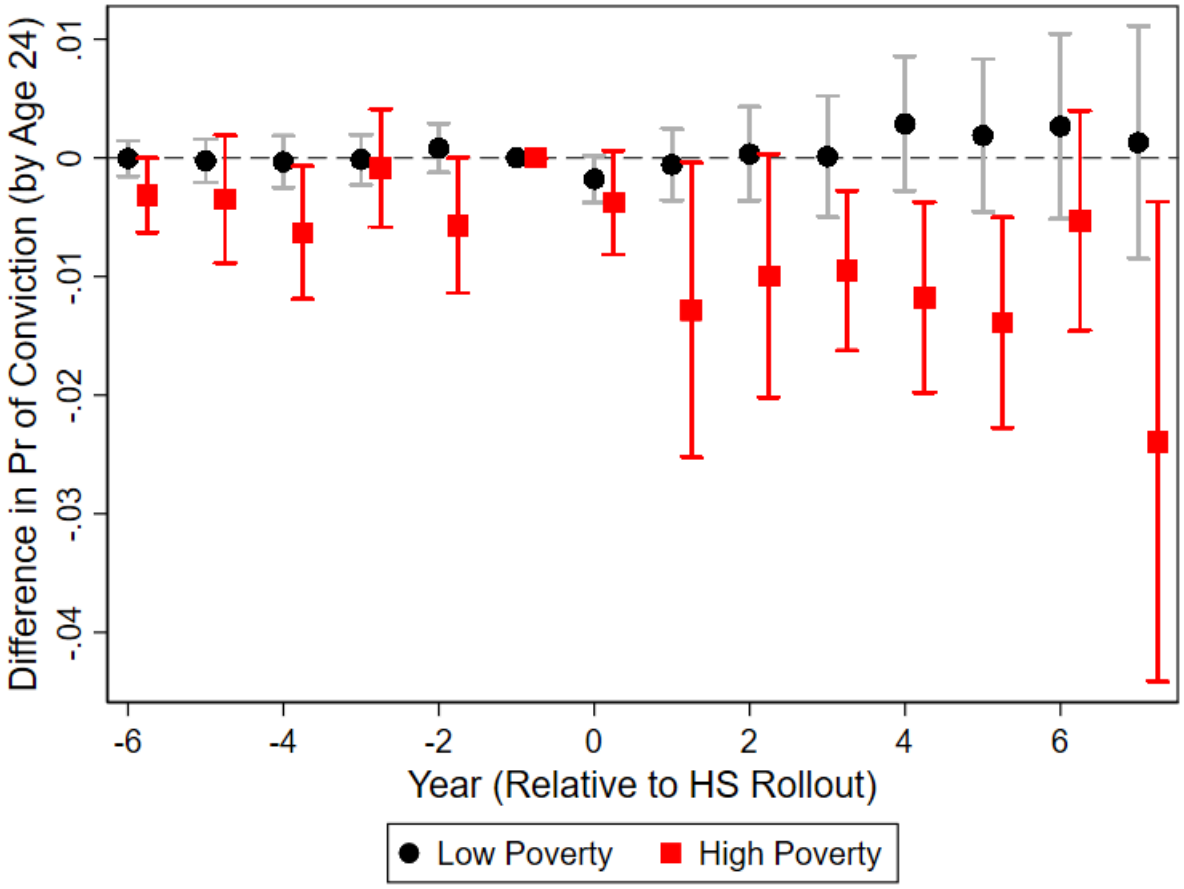
Note: Figure shows per capita county level Head Start and Smart Start funding (given in $ per 4 year olds) separately for high and low poverty counties. In the upper panel which shows Head Start funding, high poverty counties are those counties with a 1960 poverty rate above the median in North Carolina (40.2% poverty), while low poverty are those with a below median 1960 poverty rate. In the lower panel which shows Smart Start funding, high poverty counties are those counties whose poverty rate in 1980 was above the median in North Carolina (17.3% poverty) , while low poverty are those below the median. (There exist small, non-zero funding levels in the year prior to Head Start rollout for two reasons: first, following Barr and Gibbs (2017), county birth cohorts with very low funding levels are treated as not having Head Start availability, and, second, we do not count 1965 as the first year of availability since the Head Start program was introduced only as a pilot program over the Summer in that year.)

Figure A4: Head Start Estimates by Quintiles



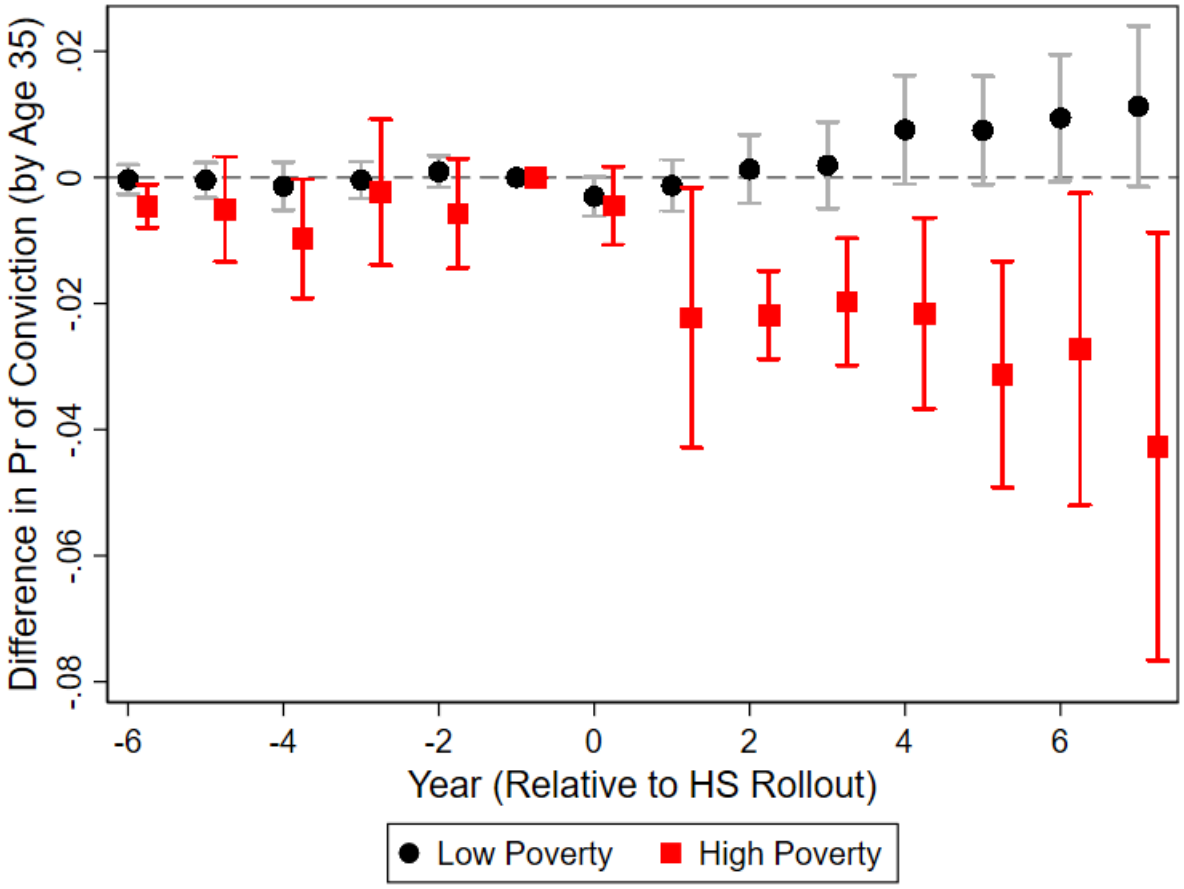
Note: Figure shows the coefficient estimates and 95% confidence intervals from estimating our basic difference-in-differences specification separately for counties in each quintile of the 1960 North Carolina poverty rate. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1955. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 35. UCR Part 1 crimes include violent crimes (those in which the description of the offense contains the words "murder", "assault", or "robbery" (rape not being included), and property crimes (those in which the description of the offense contains the words "burglary" or "larceny"). All specifications include birth county and birth-cohort fixed effects as well as 1960 county characteristics interacted with a time trend in birth cohort. 1960 county characteristics include: percent of land in farming, percent of people living in families with less than $3,000, percent of population in urban area, percent black, percent less than age 5, percent greater than age 65, and percent of employment in agriculture. The sample is restricted to counties that ever received Head Start between 1965 and 1976. The sample is further restricted to cohorts who were born between 1955 and 1968.

Figure A5: Event Study of Head Start's Impact on Criminal Conviction (by age 24)



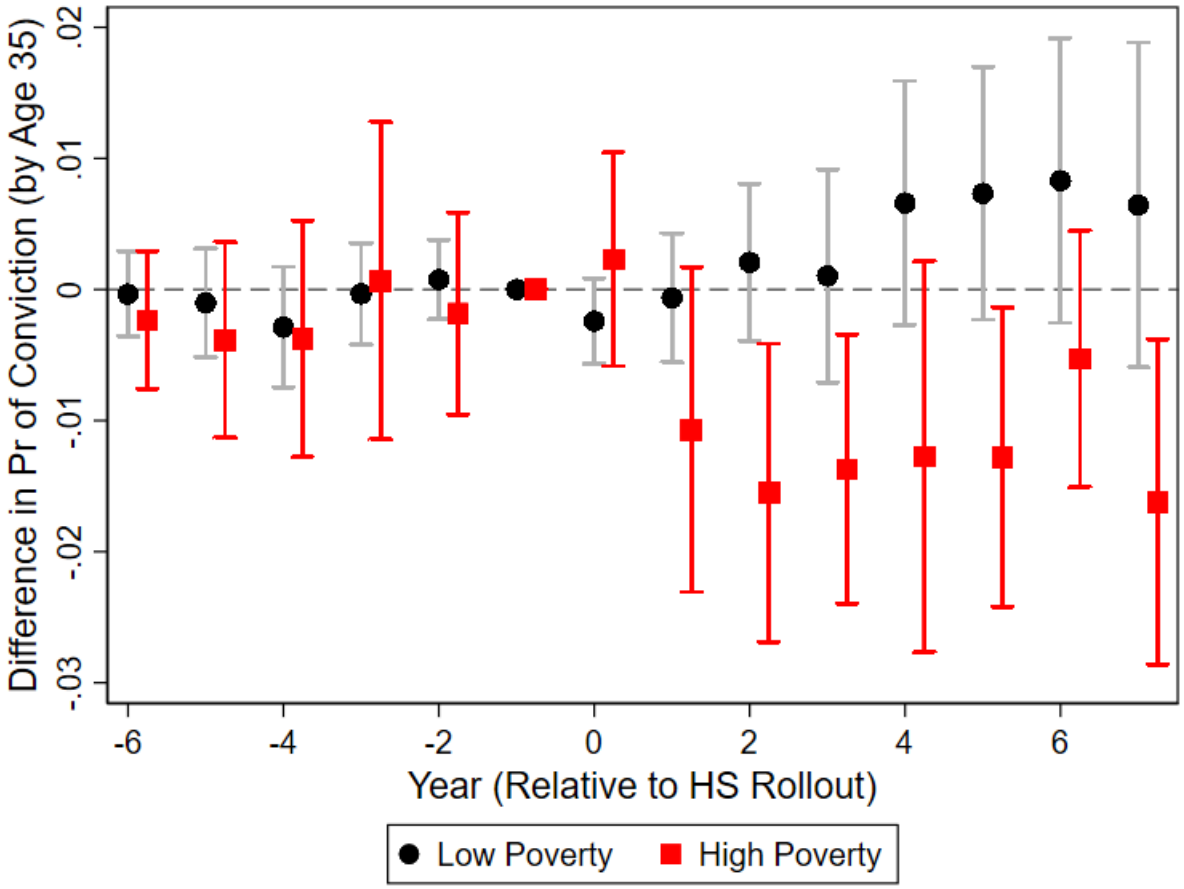
Note: Figure shows the coefficient estimates and 95% confidence interval from estimating Equation 2 separately for high and low poverty counties. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1955. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 24. UCR Part 1 crimes include violent crimes (those in which the description of the offense contains the words "murder", "assault", or "robbery" (rape not being included), and property crimes (those in which the description of the offense contains the words "burglary" or "larceny"). All specifications include birth county and birth-cohort fixed effects as well as 1960 county characteristics interacted with a time trend in birth cohort. 1960 county characteristics include: percent of land in farming, percent of people living in families with less than $3,000, percent of population in urban area, percent black, percent less than age 5, percent greater than age 65, and percent of employment in agriculture. Those counties whose poverty rate in 1960 was above the median in North Carolina (40.2% poverty) are called "High Poverty", while those below the median are called "Low Poverty". The sample is restricted to counties that ever received Head Start between 1965 and 1976. The sample is further restricted to cohorts who were born between 1955 and 1968.

Figure A6: Event Study of Head Start's Impact on Criminal Conviction (without trends)



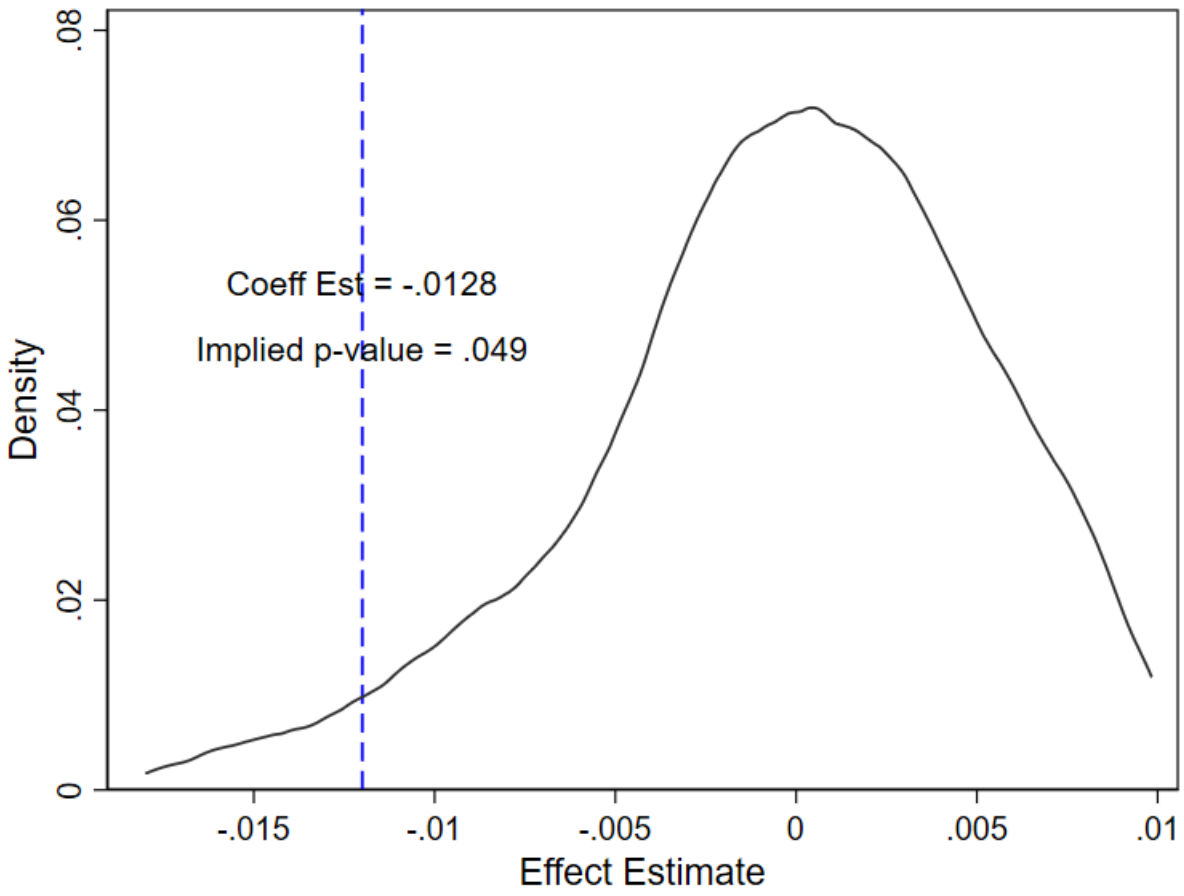
Note: Figure shows the coefficient estimates and 95% confidence interval from estimating Equation 2 separately for high and low poverty counties. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1955. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 35. UCR Part 1 crimes include violent crimes (those in which the description of the offense contains the words "murder", "assault", or "robbery" (rape not being included), and property crimes (those in which the description of the offense contains the words "burglary" or "larceny"). All specifications include birth county and birth-cohort fixed effects, but, unlike Figure 2, they do not include 1960 county characteristics interacted with a time trend in birth cohort. Those counties whose poverty rate in 1960 was above the median in North Carolina (40.2% poverty) are called "High Poverty", while those below the median are called "Low Poverty". The sample is restricted to counties that ever received Head Start between 1965 and 1976. The sample is further restricted to cohorts who were born between 1955 and 1968.

Figure A7: Event Study of Head Start's Impact on Criminal Conviction – Robustness to Inclusion of All Counties



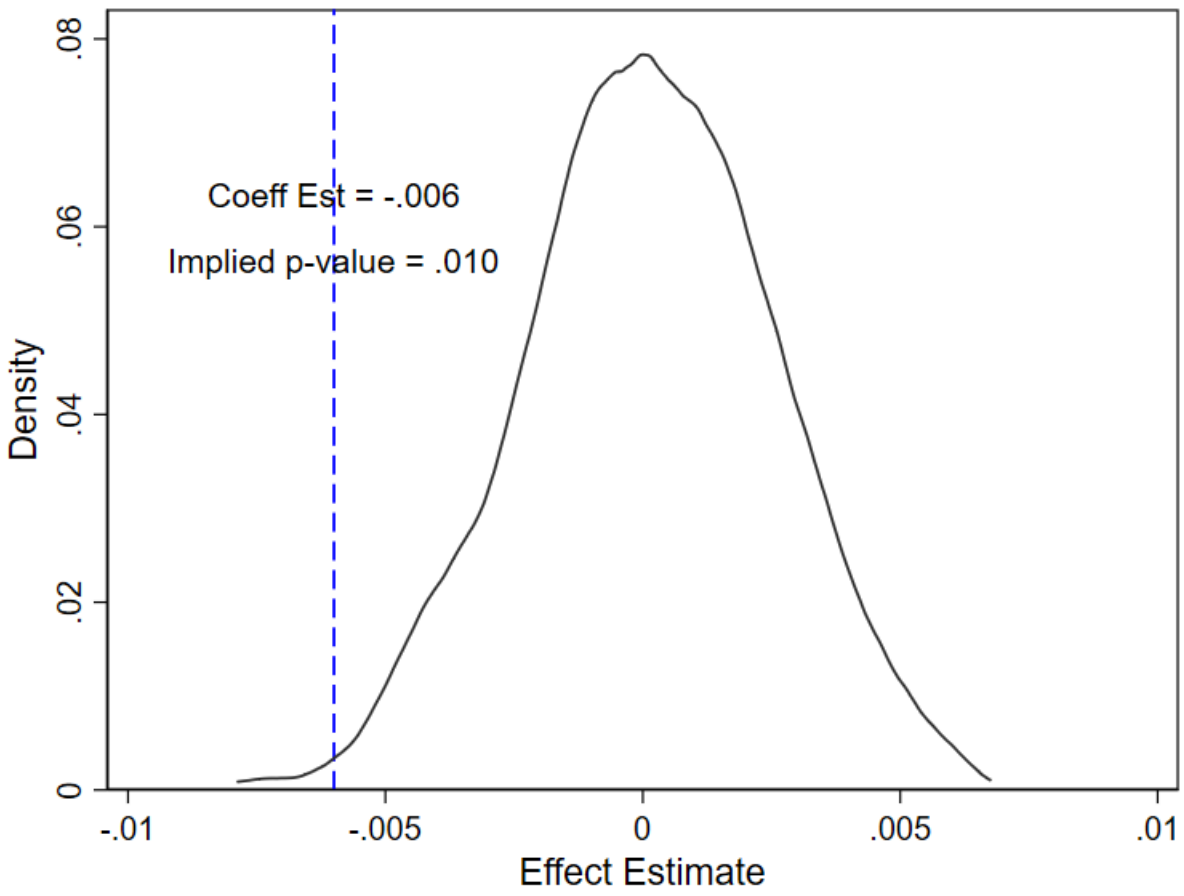
Note: Figure shows the coefficient estimates and 95% confidence interval from estimating Equation 2 separately for high and low poverty counties. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1955 multiplied by the inverse propensity to receive Head Start as defined by Stuart (2010). The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 35. UCR Part 1 crimes include violent crimes (those in which the description of the offense contains the words "murder", "assault", or "robbery" (rape not being included), and property crimes (those in which the description of the offense contains the words "burglary" or "larceny"). All specifications include birth county and birth-cohort fixed effects as well as 1960 county characteristics interacted with a time trend in birth cohort. 1960 county characteristics include: percent of land in farming, percent of people living in families with less than $3,000, percent of population in urban area, percent black, percent less than age 5, percent greater than age 65, and percent of employment in agriculture. Those counties whose poverty rate in 1960 was above the median in North Carolina (40.2% poverty) are called "High Poverty", while those below the median are called "Low Poverty". The sample is further restricted to cohorts who were born between 1955 and 1968.

Figure A8: Head Start Randomization Inference (High-Poverty Counties)



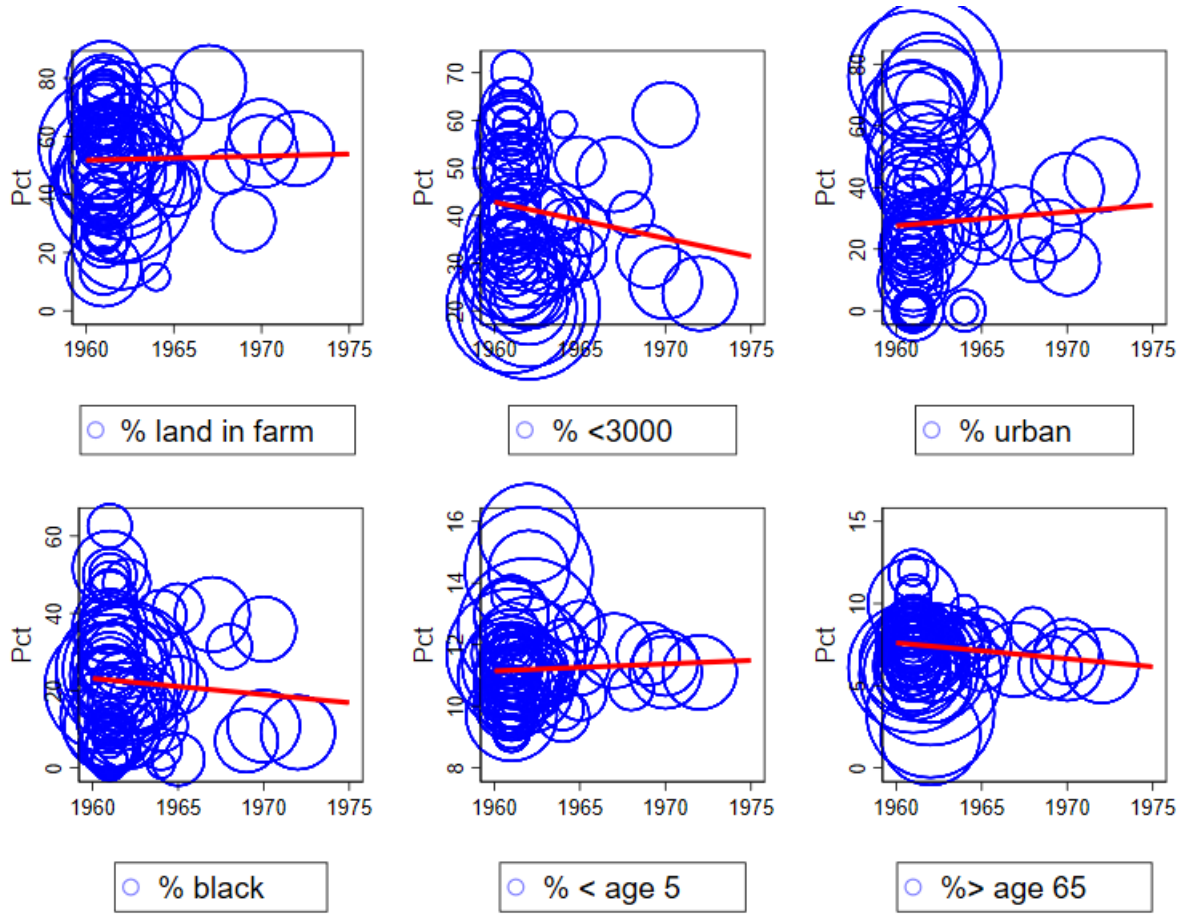
Note: Figure shows the kernel density of coefficient estimates under random assignment of Head Start availability to high poverty counties. 1000 repetitions were performed. The vertical line indicates the coefficient estimate obtained using the actual rollout of Head Start (See Table 2). A two-tailed test statistic is calculated as the share of estimates whose absolute value is greater than or equal to the estimate obtained using the actual rollout. Calculating this statistic gives an implied p-value of .049 as compared with the p-value of .040 given by the standard errors clustered at the county level.

Figure A9: Smart Start Randomization Inference



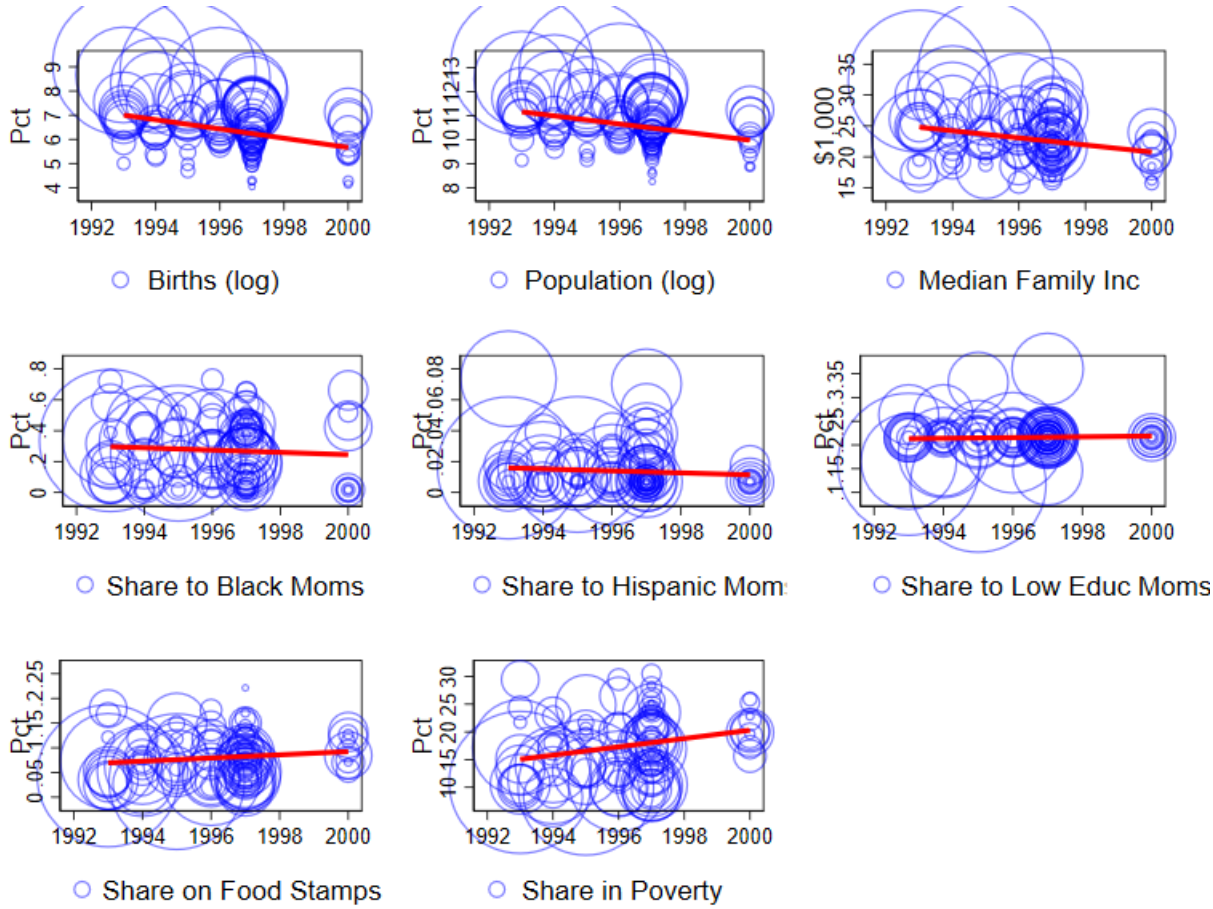
Note: Figure shows the kernel density of coefficient estimates under random assignment of Smart Start availability. 1000 repetitions were performed. The vertical line indicates the coefficient estimate obtained using the actual rollout of Smart Start (See Table 2). A two-tailed test statistic is calculated as the share of estimates whose absolute value is greater than or equal to the estimate obtained using the actual rollout. Calculating this statistic gives an implied p-value of .01 as compared with the p-value of .032 given by the standard errors clustered at the county level.

Figure A10: Relationship between Year of Initial Head Start Funding and Baseline County Characteristics



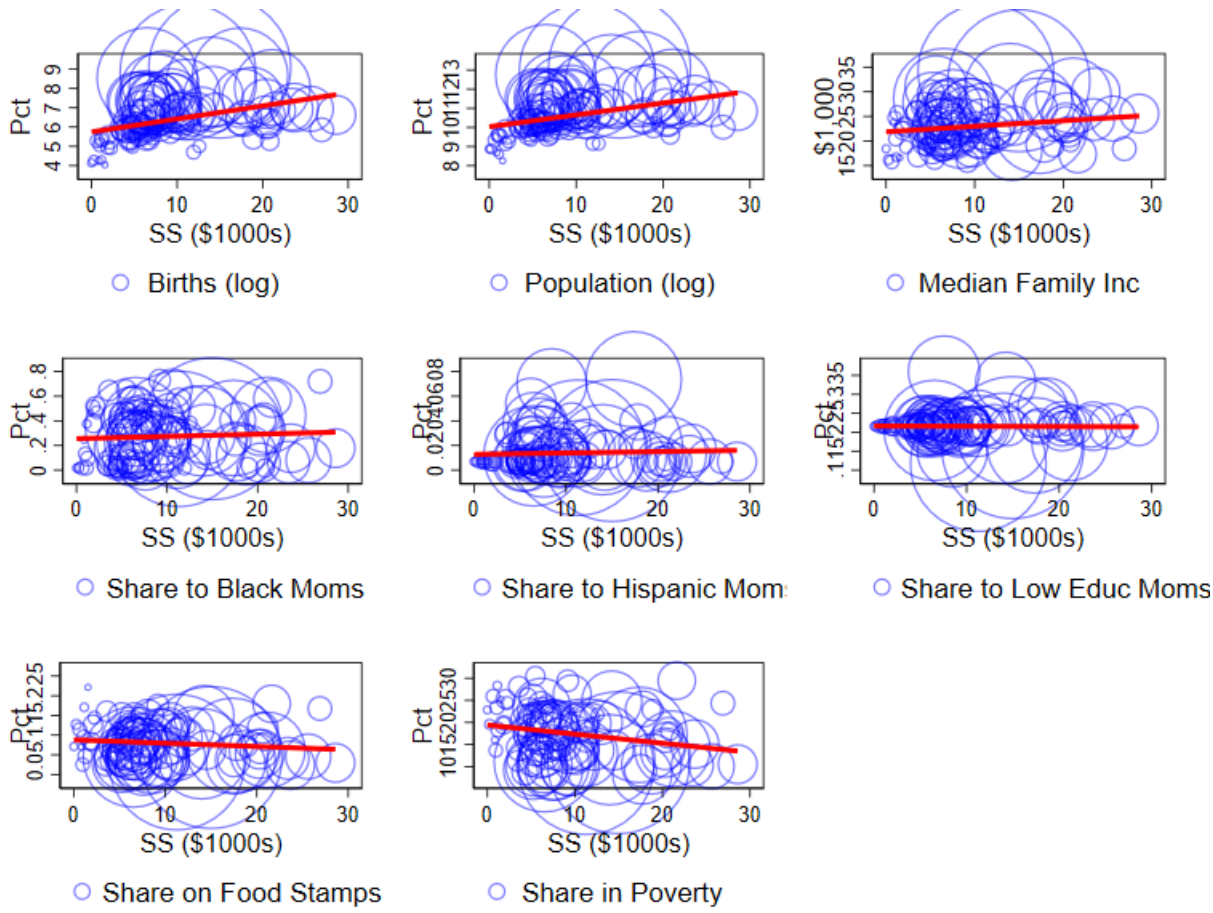
Note: Figure shows population weighted scatterplots of county characteristics against the year in which Head Start first became available in that county. Data are at the county level and weights are defined using 1955 births (represented by circle radius). A flat, horizontal fitted line suggests that the values of a given county characteristic are not systematically connected to the timing of Head Start availability.

Figure A11: Relationship between Year of Initial Smart Start Funding and Baseline County Characteristics



Note: Figure shows population weighted scatterplots of county characteristics against the year in which Smart Start first became available in that county. Data are at the county level and weights are defined using 1980 births (represented by circle radius). A flat, horizontal fitted line suggests that the values of a given county characteristic are not systematically connected to the timing of Smart Start availability.

Figure A12: Relationship between Total Smart Start Funding Levels and Baseline County Characteristics



Note: Figure shows population weighted scatterplots of county characteristics against the total amount of Smart Start funding received in a county over the sample period. Data are at the county level and weights are defined using 1980 births (represented by circle radius). A flat, horizontal fitted line suggests that the values of a given county characteristic are not systematically connected to total Smart Start funding.

Appendix: Supplementary Tables

Table A1: Relationship Between Head Start Ever Available in Sample Period and Baseline County Characteristics

	All	High Poverty	Low Poverty
	(1)	(2)	(3)
Head Start Ever Available In County			
1960 CCDB: % of land in farming	0.00353	0.00782	0.00110
	(0.0177)	(0.0290)	(0.0265)
1960 CCDB: % of people living in families with $\leq$ \$3000	-0.0503	0.0939	-0.381***
	(0.0369)	(0.0945)	(0.123)
1960 CCDB: % of population urban	-0.0297	-0.00259	-0.0712
	(0.0283)	(0.0427)	(0.0621)
1960 CCDB: % of people black	0.00244	0.0233	0.00175
	(0.0259)	(0.0272)	(0.0472)
1960 CCDB: % of people $\leq$ age 5	-0.364	-0.766	0.696
	(0.404)	(0.488)	(0.678)
1960 CCDB: % of people $\geq$ age 65	-0.112	-0.568	1.137
	(0.337)	(0.423)	(1.014)
1960 CCDB: % of employment in agriculture	-9.870	-25.45	16.11
	(14.23)	(20.82)	(20.80)
1960 CCBD: log population	1.720**	0.988	4.548*
	(0.717)	(0.791)	(2.478)
Observations	100	50	50
Mean	0.630	0.440	0.820

Note: Each column reports a separate logistic regression of an indicator for whether a county ever got Head Start by 1976 against the eight county level characteristics recommended in Hoynes and Schanzenbach (2009) and drawn from the 1960 City and County Data Books (CCDB). Observations are at the county level. Those counties whose poverty rate in 1960 was above the median in North Carolina (40.2% poverty) are called "High Poverty", while those below the median are called "Low Poverty". Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$)

Table A2: Relationship Between Year of Initial Head Start Availability and Baseline County Characteristics

	All	High Poverty	Low Poverty
	(1)	(2)	(3)
First Birth Cohort in County To Have Head Start			
1960 CCDB: % of land in farming	0.00735	-0.0396	0.00827
	(0.0234)	(0.0658)	(0.0330)
1960 CCDB: % of people living in families with $\leq$ \$3000	-0.0375	0.0158	-0.138
	(0.0473)	(0.168)	(0.105)
1960 CCDB: % of population urban	-0.0122	-0.00119	-0.000130
	(0.0235)	(0.0358)	(0.0607)
1960 CCDB: % of people black	0.00251	0.0215	-0.0117
	(0.0328)	(0.0756)	(0.111)
1960 CCDB: % of people $\leq$age 5	-0.198	-0.187	0.115
	(0.387)	(1.374)	(0.582)
1960 CCDB: % of people $\geq$age 65	-0.418	-0.358	-0.350
	(0.259)	(0.898)	(0.266)
1960 CCDB: % of employment in agriculture	-4.219	-7.168	1.668
	(13.35)	(34.97)	(19.64)
1960 CCBD: log population	-0.397	1.003	-1.444
	(0.749)	(1.483)	(1.224)
Observations	63	22	41
Mean	0.381	0.0455	0.561

Note: Each column reports a separate OLS regression of the birth year (normalized to 1962) of the first birth cohort in a given county to which Head Start was available against the eight county level characteristics recommended in Hoynes and Schanzenbach (2009) and drawn from the 1960 City and County Data Books (CCDB). Observations are at the county level. Those counties whose poverty rate in 1960 was above the median in North Carolina (40.2% poverty) are called "High Poverty", while those below the median are called "Low Poverty". Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$)

Table A3: Relationship Between Smart Start Penetration in Sample Period and Baseline County Characteristics

	All	High Poverty	Low Poverty
	(1)	(2)	(3)
Total SS Funding ($1000s)			
Share to Hispanic Moms	-0.1437	0.0220	-0.1243
	(0.3496)	(0.3795)	(0.4813)
Share to Black Moms	-0.0118	-0.0246	0.0270
	(0.0280)	(0.0396)	(0.0616)
Share to Low Educ Moms	-0.0417	-0.1474	-0.0221
	(0.0936)	(0.2466)	(0.1180)
Births (log)	1.0846	0.2368	0.1913
	(3.5013)	(3.7876)	(5.7702)
Median Family Inc	-0.0081	-0.1464	-0.1077
	(0.1531)	(0.2513)	(0.2399)
Population (log)	-0.9304	-0.2198	-0.1293
	(3.6591)	(3.7706)	(5.9627)
Share on Food Stamps	0.0369	0.3222	-0.1260
	(0.1644)	(0.2478)	(0.2445)
Observations	100	50	50
Mean	2.0458	1.2738	2.2835

Note: Each column reports a separate OLS regression of total county-level Smart Start penetration funding against selected 1980 county characteristics. Following Ladd et al. (2014), the 1980 county characteristics include the share of births to black mothers, the share of births to Hispanic mothers, the share of births to low education mothers, the share of the population using food stamps, the total number of births, the total population, and the median family income. Observations are at the birth-county level. Those counties whose poverty rate in 1980 was above the median in North Carolina (17.3% poverty) are called "High Poverty", while those below the median are called "Low Poverty". Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$)

Table A4: Relationship Between Baseline and Predicted Crime Changes and Timing and Funding of Program

Crime Index		Program Availability	
	Mean	Coefficient	% of Mean
(1)	(2)	(3)	(4)
Panel A: Head Start			
Head Start Timing			
Adulthood Conviction Rate (Born 1955))	2.7	-0.0004**	1.4
		(0.0002)	
Δ Conviction Rate (1961-1968))	2.1	-0.0005**	2.6
		(0.0003)	
Head Start Funding			
Adulthood Conviction Rate (Born 1955))	2.7	-0.0005	1.7
		(0.0007)	
Δ Conviction Rate (1961-1968))	2.1	0.0008	3.6
		(0.0009)	
Panel B: Smart Start			
Smart Start Timing			
Adulthood Conviction Rate (Born 1980))	6.4	-0.0020	3.1
		(0.0014)	
Δ Conviction Rate (1989-1994))	-1.6	0.0005	2.9
		(0.0003)	
Smart Start Funding			
Adulthood Conviction Rate (Born 1980))	6.4	0.0002	0.3
		(0.0009)	
Δ Conviction Rate (1989-1994))	-1.6	0.0000	0.1
		(0.0003)	

Note: Estimates show the relationship between baseline county characteristics and the program timing and funding, respectively. Each row represents a separate OLS regression, weighted by number of births in the baseline year. An index measuring the conviction rate is the dependent variable in odd rows while an index measuring the trend in the conviction rate is the dependent variable in even rows. In rows 1 and 2 the birth year of the cohort first exposed to Head Start (normed to 1962) is the sole independent variable, while in rows 5 and 6 the cohort first exposed to Smart Start (normed to 1997) is the sole independent variable. In rows 3-4 and 7-8, the county total of Head Start or Smart Start funding per pupil is the sole independent variable. The data are at the county-level and contain the 63 ever exposed counties in rows 1-4, and all 100 counties in rows 5-8. The indexes are constructed by regressing the crime measure on baseline county characteristics and using those coefficient estimates to predict the crime measure for each county. Robust standard errors are in parentheses in column (3). Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$)

Table A5: Effect of Head Start Availability on Criminal Conviction - Robustness of High Poverty Estimates to Inclusion of Counties that Did Not Receive Head Start

	High Poverty			
	Baseline		All Counties	
	(1)	(2)	(3)	(4)
HS Availability	-0.0128**	-0.0128**	-0.0105**	-0.0092**
	(0.0058)	(0.0061)	(0.0040)	(0.0039)
Observations	308	308	700	700
Mean	0.0462	0.0462	0.0429	0.0429
Baseline Chars X Trend		X		X

Note: Each column reports a separate OLS regression with standard errors clustered at the birth county level and reported in parentheses. Observations are at the birth county by birth year level and are weighted by number of births in each county in 1955 multiplied by the inverse propensity to receive Head Start as defined by Stuart (2010). The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 35. The reported variable of interest is an indicator for whether Head Start was available to a given county birth cohort. All specifications include birth county and birth-cohort fixed effects, and, where indicated, 1960 county characteristics interacted with a time trend in birth cohort. 1960 county characteristics include: percent of land in farming, percent of people living in families with less than $3,000, percent of population in urban area, percent black, percent less than age 5, percent greater than age 65, and percent of employment in agriculture. The sample is restricted to those counties whose poverty rate in 1960 was above the median in North Carolina (40.2% poverty), the "High Poverty" counties. In the first two columns, the sample is restricted to counties that ever received Head Start between 1965 and 1976. In the second two columns, the sample includes counties that never received Head Start between 1965 and 1976. See the notes to Table 1 for additional sample restrictions and definitions. Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$).

Table A6: Head Start Availability and Criminal Conviction - Continuous Measure of Poverty Estimates

	All	
	(1)	(2)
HS Availability	0.0059	0.0045
	(0.0059)	(0.0043)
HS Availability X Poverty	-0.0197*	-0.0188**
	(0.0111)	(0.0082)
Observations	882	882
Mean	0.0469	0.0469
Baseline Chars X Trend		X

Note: Each column reports a separate OLS regression with standard errors clustered at the birth county level and reported in parentheses. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1955. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 35. The reported variable of interest is an indicator for whether Head Start was available to a given county birth cohort interacted with the county poverty rate in 1960. (The reported estimates are also scaled up by a factor of 100.). All specifications include birth county and birth-cohort fixed effects, and, where indicated, 1960 county characteristics interacted with a time trend in birth cohort. 1960 county characteristics include: percent of land in farming, percent of people living in families with less than $3,000, percent of population in urban area, percent black, percent less than age 5, percent greater than age 65, and percent of employment in agriculture. These regressions do not restrict the sample based on the county poverty rate in 1960. See the notes to Table 1 for additional sample restrictions and definitions. Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$).

Table A7: Smart Start Availability and Criminal Conviction – Binary Availability Measure

	(1)	(2)	(3)
	All	High Poverty	Low Poverty
SS Availability (Binary)	-0.0032*	-0.0067*	-0.0013
	(0.0018)	(0.0038)	(0.0021)
Observations	1500	750	750
Mean	0.0516	0.0512	0.0517

Note: Each cell reports a separate OLS regression with standard errors clustered at the birth county level and reported in parentheses. All specifications include birth county and birth-cohort fixed effects. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1980. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 24. Binary Smart Start availability, the independent variable of interest, is defined as Smart Start penetration level above the 25th percentile of penetration. See the notes to Table 1 for additional sample restrictions and definitions. Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$).

Table A8: Effect of Smart Start Availability on Criminal Conviction - By Presence of Head Start, Binary Measure

	(1)	(2)	(3)	(4)
	All		High Poverty	
Panel A: Binary Measure				
SS Availability (Binary)	-0.0032*	-0.0099***	-0.0067*	-0.0098**
	(0.0018)	(0.0033)	(0.0038)	(0.0037)
Observations	1500	555	750	435
Mean	0.0516	0.0528	0.0512	0.0532
Head Start	All	No Head Start	All	No Head Start

Note: Each column reports a separate OLS regression with standard errors clustered at the birth county level and reported in parentheses. All specifications include birth county and birth-cohort fixed effects. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1980. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of either UCR Part 1 property crimes or Part 1 violent crimes in North Carolina by age 24. Binary Smart Start availability, the independent variable of interest, is defined as Smart Start penetration level above the 25th percentile of penetration. See the notes to Table 1 for additional sample restrictions and definitions. Columns (2) and (4) further restrict the sample to counties without a Head Start program by 1980. Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$).

Table A9: Smart Start Funding and Criminal Conviction - Continuous Measure of Poverty Estimates

	All	
	(1)	(2)
SS ($1000s)	0.0015	0.0053
	(0.0045)	(0.0040)
SS ($1000s) X Poverty	-0.0508*	-0.0721***
	(0.0266)	(0.0255)
Observations	1500	1500
Mean	0.0516	0.0516
Baseline Chars X Trend		X

Note: Each column reports a separate OLS regression with standard errors clustered at the birth county level and reported in parentheses. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1980. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 24. The reported variable of interest is an indicator for whether Smart Start was available to a given county birth cohort interacted with the county poverty rate in 1980. (The reported estimates are also scaled up by a factor of 100.). All specifications include birth county and birth-cohort fixed effects, and, where indicated, 1980 county characteristics interacted with a time trend in birth cohort. Following Ladd et al. (2014), the 1980 county characteristics include the share of births to black mothers, the share of births to Hispanic mothers, the share of births to low education mothers, the share of the population using food stamps, the total number of births, the total population, and the median family income. These regressions do not restrict the sample based on the county poverty rate in 1980. See the notes to Table 1 for additional sample restrictions and definitions. Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$).

Table A10: Smart Start Funding and Criminal Conviction – Full Interaction with Presence of Head Start

	(1)	(2)	(3)
	All	High Poverty	Low Poverty
SS ($1000s)	-0.0049	-0.0036	-0.0028
	(0.0030)	(0.0028)	(0.0035)
No Head Start X SS ($1000s)	-0.0062	-0.0101**	-0.0073*
	(0.0046)	(0.0048)	(0.0041)
Observations	1500	750	750
Mean	0.0516	0.0512	0.0517

Note: Each column reports a separate OLS regression with standard errors clustered at the birth county level and reported in parentheses. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1980. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 24. The reported variable of interest is a measure of Smart Start funding penetration for a given county birth cohort, constructed following Ladd et al. (2014). In the second row, the reported variable of interest is the same measure of Smart Start funding penetration interacted with an indicator variable for whether the county was served by Head Start by 1980. All specifications include birth county and birth-cohort fixed effects. See the notes to Table 1 for additional sample restrictions and definitions. Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$).

Table A11: Head Start Availability and Criminal Conviction – Effects by Conviction Age

	(1)	(2)	(3)
	All	High Poverty	Low Poverty
Conviction By Age 24			
Head Start Availability	-0.0005	-0.0044	0.0009
	(0.0015)	(0.0038)	(0.0015)
Mean	0.0234	0.0229	0.0236
Conviction By Age 30			
Head Start Availability	-0.0017	-0.0101*	0.0016
	(0.0024)	(0.0050)	(0.0025)
Mean	0.0376	0.0373	0.0377
Conviction By Age 35			
Head Start Availability	-0.0017	-0.0128**	0.0026
	(0.0031)	(0.0058)	(0.0033)
Mean	0.0469	0.0462	0.0471
Observations	882	308	574

Note: Each column reports a separate OLS regression with standard errors clustered at the birth county level and reported in parentheses. All specifications include birth county and birth-cohort fixed effects. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1955. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 24, age 30 and age 35, respectively. The reported variable of interest is an indicator for whether Head Start was available to a given county birth cohort. See the notes to Table 1 for additional sample restrictions and definitions. Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$).

Table A12: Smart Start Availability and Criminal Conviction – Effects by Conviction Age

	(1)	(2)	(3)
	All	High Poverty	Low Poverty
Conviction By Age 24			
SS ($1000s)	-0.0064**	-0.0118**	-0.0030
	(0.0029)	(0.0051)	(0.0035)
Observations	1500	750	750
Mean	0.0516	0.0512	0.0517
Conviction By Age 25			
SS ($1000s)	-0.0059*	-0.0119**	-0.0022
	(0.0032)	(0.0051)	(0.0038)
Observations	1400	700	700
Mean	0.0571	0.0561	0.0575
Conviction By Age 26			
SS ($1000s)	-0.0068*	-0.0138**	-0.0024
	(0.0039)	(0.0057)	(0.0047)
Observations	1300	650	650
Mean	0.0616	0.0603	0.0621

Note: Each cell reports a separate OLS regression with standard errors clustered at the birth county level and reported in parentheses. All specifications include birth county and birth-cohort fixed effects. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1980. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 24, age 25 or age 26, respectively. The reported variable of interest is a measure of Smart Start funding penetration for a given county birth cohort, constructed following Ladd et al. (2014). See the notes to Table 1 for additional sample restrictions and definitions. Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$).

Table A13: Effect of Early Childhood Education on Criminal Conviction – With Trends

	(1)	(2)	(3)	(4)	(5)	(6)
	All		High		Low	
Panel A: Head Start						
Head Start Availability	-0.0017	-0.0028	-0.0128**	-0.0128**	0.0026	0.0013
	(0.0031)	(0.0031)	(0.0058)	(0.0061)	(0.0033)	(0.0040)
Observation	882	882	308	308	574	574
Mean	0.0469	0.0469	0.0462	0.0462	0.0471	0.0471
Panel B: Smart Start						
SS ($1000s)	-0.0064**	-0.0059**	-0.0118**	-0.0110***	-0.0030	-0.0027
	(0.0029)	(0.0025)	(0.0051)	(0.0040)	(0.0035)	(0.0025)
Observations	1500	1500	750	750	750	750
Mean	0.0516	0.0516	0.0512	0.0512	0.0517	0.0517
Baseline Chars X Trend		X		X		X

Note: Each cell reports a separate OLS regression with standard errors clustered at the birth county level and reported in parentheses. In Panel A, observations are at the birth county by birth year level and are weighted by the number of births in each county in 1955. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 35. The reported variable of interest is an indicator for whether Head Start was available to a given county birth cohort. All specifications include birth county and birth-cohort fixed effects, and, where indicated, 1960 county characteristics interacted with a time trend in birth cohort. 1960 county characteristics include: percent of land in farming, percent of people living in families with less than $3,000, percent of population in urban area, percent black, percent less than age 5, percent greater than age 65, and percent of employment in agriculture. In Panel B, observations are at the birth county by birth year level and are weighted by the number of births in each county in 1980. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 24. The reported variable of interest is a measure of Smart Start funding penetration for a given county birth cohort, constructed following Ladd et al. (2014) All specifications include birth county and birth-cohort fixed effects, and, where indicated, 1980 county characteristics interacted with a time trend in birth cohort. Following Ladd et al. (2014), the 1980 county characteristics include the share of births to black mothers, the share of births to Hispanic mothers, the share of births to low education mothers, the share of the population using food stamps, the total number of births, the total population, and the median family income. See the notes to Table 1 for additional sample restrictions and definitions. Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$).

Table A14: Effect of Early Childhood Education on Criminal Conviction - Robustness to Inclusion of Time-Varying Covariates

	(1)	(2)	(3)
	All	High Poverty	Low Poverty
SS ($1000s)	-0.0064**	-0.0114**	-0.0029
	(0.0029)	(0.0048)	(0.0032)
Observations	1500	750	750
Mean	0.0516	0.0512	0.0517
Covariates	Yes	Yes	Yes

Note: Each column reports a separate OLS regression with standard errors clustered at the birth county level and reported in parentheses. All specifications include birth county and birth-cohort fixed effects, as well as a set of county-year level covariates. Following Ladd et al. (2014), the county-year level characteristics include the share of births to black mothers, the share of births to Hispanic mothers, the share of births to low education mothers, the share of the population using food stamps, the total number of births, the total population, and the median family income. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1980. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 24. The reported variable of interest is a measure of Smart Start funding penetration for a given county birth cohort, constructed following Ladd et al. (2014) See the notes to Table 1 for additional sample restrictions and definitions. Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$).

Table A15: Other War On Poverty Programs and Head Start - High Poverty Counties

	High Poverty				
	(1)	(2)	(3)	(4)	(5)
Head Start Availability	-0.0128**	-0.0123*	-0.0122*	-0.0155**	-0.0149**
	(0.0058)	(0.0060)	(0.0062)	(0.0072)	(0.0065)
Observations	308	308	308	308	308
Mean	0.0462	0.0462	0.0462	0.0462	0.0462
Baseline Chars X Trend	X		X		X
WOP Controls	None	FS	FS	FS + Other WOP	FS + Other WOP

Note: Each column reports a separate OLS regression with standard errors clustered at the birth county level and reported in parentheses. All specifications include birth county and birth-cohort fixed effects. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1955. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 35. The reported variable of interest is an indicator for whether Head Start was available to a given county birth cohort. All specifications include birth county and birth-cohort fixed effects, and, where indicated, 1960 county characteristics interacted with a time trend in birth cohort. 1960 county characteristics include: percent of land in farming, percent of people living in families with less than $3,000, percent of population in urban area, percent black, percent less than age 5, percent greater than age 65, and percent of employment in agriculture. In these specifications, controls for exposure to various War on Poverty programs, including the Food Stamp Program (FS) are also included. "Other War on Poverty Programs" are those recommended by Bailey and Goodman-Bacon (2015) and include per capita expenditures on Public Assistance Transfers, Medicaid expenditures, Community Health Centers and Community Action Agencies. See the notes to Table 1 for additional sample restrictions and definitions. Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$)

Table A16: Relationship between Head Start Availability and Possible Confounders

	All		High Poverty		Low Poverty	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: War on Povery Programs						
0-5 Food Stamp Exposure	-0.0429	-0.0429	0.0666	0.0666	-0.0639	-0.0639
	(0.0802)	(0.0802)	(0.1593)	(0.1593)	(0.0950)	(0.0950)
Public Assistance Transfers	6.9467	6.9467	0.7701	0.7701	11.5153**	11.5153**
	(5.7513)	(5.7513)	(6.7353)	(6.7353)	(4.5238)	(4.5238)
Medicaid	9.2430	9.2430	2.2806	2.2806	12.8272**	12.8272**
	(5.7438)	(5.7438)	(5.9320)	(5.9320)	(5.9322)	(5.9322)
Community Health Center Funds	681.9056	681.9056	-82.5776	-82.5776	916.6069	916.6069
	(521.5891)	(521.5891)	(502.3024)	(502.3024)	(720.6964)	(720.6964)
CAP Seniors Program Grant	0.0356	0.0356	0.0297	0.0297	0.0302	0.0302
	(0.0521)	(0.0521)	(0.0407)	(0.0407)	(0.0734)	(0.0734)
Legal Services Program Grant	0.0460	0.0460	-0.0013	-0.0013	0.0557	0.0557
	(0.0333)	(0.0333)	(0.0080)	(0.0080)	(0.0434)	(0.0434)
Panel B: Health						
Adjusted Mortality Rate, All Ages	0.8698	0.8698	19.5563	19.5563	-5.3087	-5.3087
	(9.0852)	(9.0852)	(17.5338)	(17.5338)	(9.3156)	(9.3156)
White, Infant Mortality Rate	0.4678	0.4678	0.2843	0.2843	0.1544	0.1544
	(0.8910)	(0.8910)	(1.3270)	(1.3270)	(1.4135)	(1.4135)
Nonwhite Infant Mortality Rate	-2.2424	-2.2424	-3.6512	-3.6512	-1.4567	-1.4567
	(2.3849)	(2.3849)	(3.8859)	(3.8859)	(3.2687)	(3.2687)
Infant Mortality Rate	-1.1731	-1.1731	-1.1761	-1.1761	-1.5512	-1.5512
	(0.9761)	(0.9761)	(1.1773)	(1.1773)	(1.0473)	(1.0473)
Neonatal Infant Mortality Rate	0.2524	0.2524	0.1171	0.1171	0.1792	0.1792
	(0.8686)	(0.8686)	(1.0237)	(1.0237)	(1.1750)	(1.1750)
Postneonatal Infant Mortality Rate	-1.4255*	-1.4255*	-1.2931	-1.2931	-1.7304**	-1.7304**
	(0.7816)	(0.7816)	(1.3091)	(1.3091)	(0.6643)	(0.6643)
Observations	882	882	308	308	574	574
Baseline Chars X Trend		X		X		X

Note: Each cell reports a separate OLS regression with standard errors clustered at the birth county level and reported in parentheses. All specifications include birth county and birth-cohort fixed effects. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1955. In each row the dependent variable is a county-year measure of spending or infant health that could potentially confound our estimates of the impact of Head Start. All dependent variables are taken from Bailey et al (2015). The reported variable of interest is an indicator for whether Head Start was available to a given county birth cohort. All specifications include birth county and birth-cohort fixed effects, and, where indicated, 1960 county characteristics interacted with a time trend in birth cohort. 1960 county characteristics include: percent of land in farming, percent of people living in families with less than $3,000, percent of population in urban area, percent black, percent less than age 5, percent greater than age 65, and percent of employment in agriculture. See the notes to Table 1 for additional sample restrictions and definitions. Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$).

Table A17: Effect of Early Childhood Education on Criminal Conviction - By Crime Type

	(1)	(2)	(3)
	All	High Poverty	Low Poverty
Panel A: Head Start			
Property Crime			
Head Start Availability	-0.0024	-0.0084***	-0.0000
	(0.0016)	(0.0028)	(0.0016)
Mean	0.0256	0.0255	0.0256
Violent Crime			
Head Start Availability	0.0007	-0.0045	0.0027
	(0.0017)	(0.0032)	(0.0018)
Mean	0.0213	0.0207	0.0215
Observations	882	308	574
Panel B: Smart Start			
Property Crime			
SS ($1000s)	-0.0025*	-0.0048*	-0.0006
	(0.0013)	(0.0024)	(0.0015)
Mean	0.0274	0.0265	0.0278
Violent Crime			
SS ($1000s)	-0.0039**	-0.0070**	-0.0024
	(0.0019)	(0.0030)	(0.0023)
Mean	0.0242	0.0247	0.0240
Observations	1500	750	750

Note: Each cell reports a separate OLS regression with standard errors clustered at the birth county level and reported in parentheses. All specifications include birth county and birth-cohort fixed effects. Panel A reports results using the Head Start sample. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1955. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of either a UCR Part 1 property crime or a Part 1 violent crime in North Carolina by age 35. Panel B reports results using the Smart Start sample. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1980. The dependent variable is the fraction of individuals in a given birth county and birth year cohort that are later convicted of either a Part 1 property crime or a Part 1 violent crime in North Carolina by age 24. The reported variable of interest is a measure of Smart Start funding penetration for a given county birth cohort, constructed following Ladd et al. (2014). See the notes to Table 1 for additional sample restrictions and definitions. Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$).

Table A18: Effect of Smart Start Funding on Criminal Conviction - By Race, By Presence of Head Start

	(1)	(2)	(3)	(4)
	All		High Poverty	
White				
SS ($1000s)	-0.0026	-0.0067	-0.0029	-0.0066
	(0.0020)	(0.0048)	(0.0053)	(0.0061)
Observations	1329	470	674	372
Mean	0.0315	0.0300	0.0272	0.0267
Non-White				
SS ($1000s)	-0.0191***	-0.0270***	-0.0216***	-0.0287***
	(0.0064)	(0.0051)	(0.0057)	(0.0055)
Observations	1313	454	662	360
Mean	0.0948	0.0820	0.0805	0.0804
Head Start	All	No Head Start	All	No Head Start

Note: Each cell reports a separate OLS regression with standard errors clustered at the birth county level and reported in parentheses. All specifications include birth county and birth-cohort fixed effects. Results are reported separately for white cohorts and non-white cohorts. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1980. The dependent variable is the fraction of individuals of a given race in a given birth county and birth year cohort that are later convicted of either UCR Part 1 crimes in North Carolina by age 24. The reported variable of interest is a measure of Smart Start funding penetration for a given county birth cohort, constructed following Ladd et al. (2014). (Sample sizes are smaller for these specifications because from 1989 to 1993 the natality files for 25% of counties in North Carolina do not have race breakdowns. For these years, race is available only for counties in which 1980 populations for the non-white group formed at least 10 percent of the total population or numbered at least 10,000.) See the notes to Table 1 for additional sample restrictions and definitions. Columns (2) and (4) further restrict the sample to counties without a Head Start program by 1980. Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$).

Table A19: Effect of Early Childhood Education on Criminal Conviction- By Sex

	(1)	(2)	(3)
	All	High Poverty	Low Poverty
Panel A: Head Start			
Male			
Head Start Availability	-0.0039	-0.0225*	0.0034
	(0.0053)	(0.0111)	(0.0053)
Observations	882	308	574
Mean	0.0794	0.0790	0.0796
Female			
Head Start Availability	0.0004	-0.0037**	0.0019
	(0.0014)	(0.0017)	(0.0016)
Observations	882	308	574
Mean	0.0145	0.0139	0.0147
Panel B: Smart Start			
Male			
SS ($1000s)	-0.0105**	-0.0163**	-0.0057
	(0.0045)	(0.0070)	(0.0057)
Observations	1500	750	750
Mean	0.0795	0.0783	0.0801
Female			
SS ($1000s)	-0.0020	-0.0077**	0.0003
	(0.0019)	(0.0035)	(0.0017)
Observations	1500	750	750
Mean	0.0230	0.0237	0.0228

Note: Each cell reports a separate OLS regression with standard errors clustered at the birth county level and reported in parentheses. All specifications include birth county and birth cohort fixed effects. Panel A reports results using the Head Start sample for male and female cohorts separately. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1955. The dependent variable is the fraction of individuals of a given sex in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 35. The reported variable of interest is an indicator for whether Head Start was available to a given county birth cohort. Panel B reports results using the Smart Start sample for male and female cohorts separately. Observations are at the birth county by birth year level and are weighted by the number of births in each county in 1980. The dependent variable is the fraction of individuals of a given sex in a given birth county and birth year cohort that are later convicted of a UCR Part 1 crime in North Carolina by age 24. The reported variable of interest is a measure of Smart Start funding penetration for a given county birth cohort, constructed following Ladd et al. (2014). (Sample sizes are smaller for these specifications because from 1989 to 1993 the natality files for 25% of counties in North Carolina do not have race breakdowns. For these years, race is available only for counties in which 1980 populations for the non-white group formed at least 10 percent of the total population or numbered at least 10,000.) See the notes to Table 1 for additional sample restrictions and definitions. Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$).

Table A20: Head Start and Likelihood of Residing in One's State of Birth (Census)

	National				South			
	All		Men Only		All		Men Only	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Fraction with HS Avail.	-0.021	0.004	-0.017	0.009	-0.018	-0.009	-0.013	-0.005
	(0.016)	(0.008)	(0.017)	(0.008)	(0.029)	(0.020)	(0.029)	(0.021)
Obs	3,150,292	3,150,292	1,546,355	1,546,355	1,002,875	1,002,875	487,059	487,059
Mean	0.66	0.66	0.66	0.66	0.68	0.68	0.68	0.68
State Linear Trend		X		X		X		X

Note: Each cell represents a separate OLS regression with standard errors clustered at the state of birth level (in parentheses). Observations are at the individual level from the 1990 and 2000 Census. The dependent variable is whether an individual is currently living in his or her state of birth. The key explanatory variables are measures of Head Start availability for a birth cohort in a particular state. This is the weighted average of the Head Start availability variable across counties in a state, where the weights are the number of births in each county in 1960. All specifications include birth state and birth year fixed effects as well as indicators for race, age, and sex. Sample restricted to ages 18-35. Significance levels indicated by: *($p < 0.10$), **($p < 0.05$), ***($p < 0.01$).

Appendix B: Effect Size Comparison Explanation and Backup

Figure A1 consolidates information on effect size and estimate precision from a comprehensive review of studies that contain estimates of the causal effect of an early childhood education program on the likelihood that an individual will become a criminal. These studies are included in Appendix Table B1 and B2 and focus on the evaluation of three separate early childhood education programs: Head Start, Perry Preschool, and the Abecedarian Project.

Perry Preschool and the Abecedarian Project were pilot interventions and enrolled cohorts from the later 1960s and 1970s, respectively. Head Start, the nation's largest early childhood education program, has been in operation since mid-1960s. Two studies that include criminal justice outcomes focus on cohorts from the 1960s and 1970s, while a third focuses on cohorts born in the early 1980s. The period of focus of each study is illustrated by its location on the x-axis in Figure A1.

A common theme among these studies is the relatively small sample size available to the researchers (column (5) of Table B1). In the case of Perry Preschool and the Abecedarian Project, the small-scale single site nature of the interventions resulted in samples of around 100 individuals in each case. In the case of the Head Start evaluations, while the number of program participants was extremely large, researchers were limited by the sample sizes available to them in survey data such as the PSID and CNLSY.

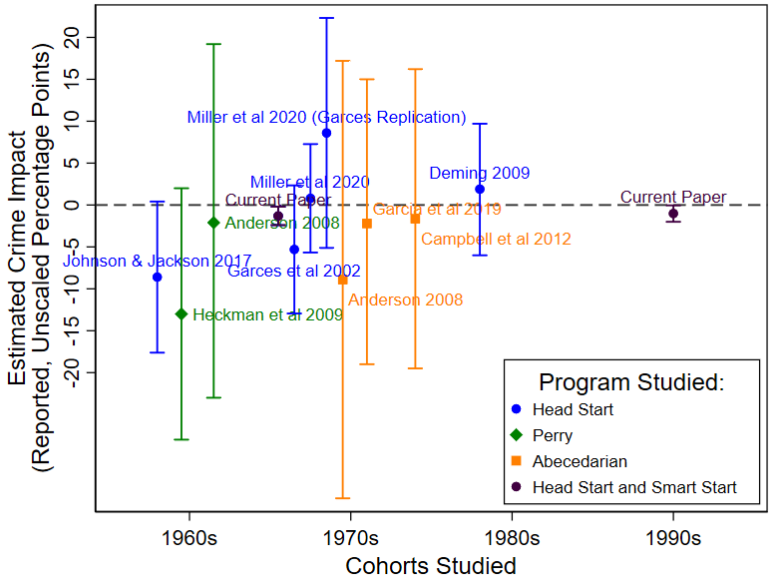
The estimates underlying Figure A1 are displayed in Figure B1 and contained in column (4) of Table B2. Whenever there are several outcomes considered in one study, we display the estimates for outcome measures which are closest to the measures of criminality we study (namely, likelihood of criminal conviction by adulthood). Having said that, it is important to note that some of the reported estimates measure criminality by the likelihood of arrest, some by the likelihood of conviction, and some by the likelihood of incarceration.

In each case we display the estimated effect likelihood that an individual will become a criminal per $1,000 (2015 dollars).[76] For example, in the case of Johnson and Jackson (2017), we take the estimates from Row 7, Col 10 of Table 2 and scale by the average Head Start funding levels in thousands of 2015 dollars (i.e., 6.027). (The raw reported and unscaled estimates are displayed in Figure B1 below.) The 95 percent confidence intervals are produced similarly and are contained in the final column of the table.

In Table B2 we provide additional details about the estimates used in Figure A1 and Table B2, including the raw estimate, the measure of criminality, the associated mean, whether the criminal behavior was self-reported, and characteristics of the sample (age at observation and subgroup). We also provide additional information on alternative estimates for other criminal measures or subgroups contained in each study.

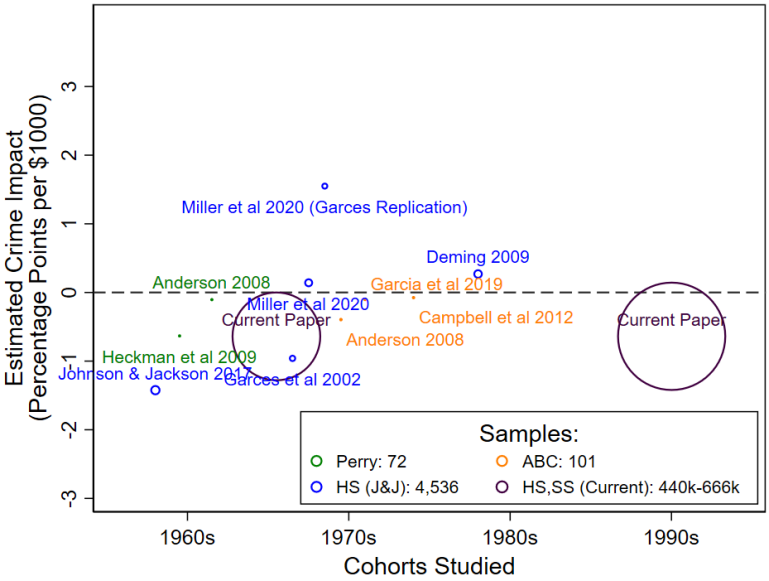
[76] Program costs for Perry and Abecedarian are taken from Heckman et al (2009) and Garces et al (2002), respectively, and are $20,648 and $22,687 respectively. Because program funding for Head Start varies across cohorts and different papers study different cohorts, we use the funding reported by each study. For Deming (2009), Garces et al (2002) and Johnson and Jackson (2017) those numbers are, respectively, $6,981, $5,540, and $6,027.

Figure B1: Effect Size Comparison: Reported, Unscaled Estimates



Note: Figure shows estimated crime impact and associated 95 percent confidence intervals. In this figure, as opposed to Figure A1, the estimates displayed are simply the estimates reported in each study with no effort to scale by cost.

Figure B2: Effect Size Comparison: Sample Sizes



Note: Figure shows estimated crime impact per $1,000 (2015 dollars). See Figure A1a and Appendix B for additional details on construction. Bubbles indicate number of individuals represented in each study. Figures presented for Head Start and Smart Start are for high-poverty counties only. Total numbers of individuals are presented in Table 1 and are 1,487,225 (Head Start) and 1,407,042 (Smart Start). See column (4) of Table B1 for exact sample sizes of other studies.

Table B1: Estimates and confidence intervals are reported per $2015 dollars.

Study	Strategy	Program	Cohorts	N	Estimate	CI
Deming 2009	Family FEs	Head Start	1980-1986	3698	.27	[-.85-1.38]
Garces et al 2002	Family FEs	Head Start	1966-1977	1,742	-.96	[-2.34-.42]
Miller et al 2020	Family FEs	Head Start	1966-1977	1535	1.55	[-.92-4.03]
Miller et al 2020	Reweighted FFE	Head Start	1966-1977	3206	.14	[-1.02-1.31]
Johnson and Jackson 2017	DD	Head Start	1950-1976	4536	-1.42	[-2.91-.07]
Precision weighted avg.	-	Head Start	-	-	-.21	[-.82-2.06]
Anderson 2008	RCT	Perry	1962-1967	72	-.10	[-1.11-.92]
Heckman et al 2009	RCT	Perry	1962-1967	72	-.62	[-1.35-.09]
Anderson 2008	RCT	Abecedarian	1972-1977	111	-.39	[-1.54-.75]
Garcia et al 2019	RCT	Abecedarian	1972-1980	143	-.04	[-.79-.66]
Campbell et al 2012	RCT	Abecedarian	1972-1980	101	-.072	[-.85-.71]

Note: Program costs for Perry and Abecedarian are taken from Heckman et al (2009) and Garces et al (2002), respectively, and are $20,648 and $22,687 respectively. Because program funding for Head Start varies across cohorts and different papers study different cohorts, we use the funding reported by each study. For Deming (2009), Garces et al. (2002) and Johnson and Jackson (2017) those numbers are, respectively, $6,981, $5,540, and $6,027. The selected effect estimates are provided in column (4) of Table B2.

Table B2: Estimates and Confidence Intervals: Detailed Breakdowns

Study	Program	Direct Reference	Reported Estimate	Mean	Measure	(Sub)Group	Age	Other Measures	Self Reported?
Current Paper	Head Start	Table 2 Row 1 Col 1	1.3 (.57) pp	4.7	Conviction	Overall	By Age 35	1.3 (.68) pp for Non-White	Administrative
Current Paper	Smart Start	Table 2 Row 2 Col 2	.65 (.30) pp	5.1	Conviction	Overall	By Age 24	2.1 (.59) pp for Non-White	Administrative
Deming 2009	Head Start	Table 5 Row 7 Col 1	1.9 (4.0) pp	NA*	Conviction, Probation, Sentence or Incarceration	Overall	Young Adulthood (Avg. Age: 23)	.2 (5.7) for Women 5.1 (5.0) for Black	Self Reported
Garces et al 2002	Head Start	Table 2 Row 12 Col 4	-5.3 (3.9) pp	10.0	Charge or Conviction	Overall	Young Adulthood (Avg. Age: 24)	-11.7 (5.9) for Black -10.3 (7.0) for Black and No High School	Self Reported
Miller et al 2020	Head Start	Table C7 Row D Col 3	.8 (3.3) pp	10.6	Charge or Conviction	Overall	By 1995 PSID Interview (Avg. Age 23)		Self Reported
" " (Garces Replication)	Head Start	Table D3 Row D Col 4	8.6 (7.0) pp	10.6	Charge or Conviction	Overall	By 1995 PSID Interview (Avg. Age 23)		
Johnson and Jackson 2017	Head Start	Table 2 Row 7 Col 10	-8.56 (4.57) pp	8	Incarceration	Overall	By Last PSID Interview (Up to Age 50)	-2.54 (1.48) pp Without Using IV	Reported by Family Member*
Anderson 2008	Perry	Table 10 Row 5 Col 9	-2.1 (10.9) pp	71.8	Record	Men	By Age 27	-14.6 (12.5) for Women -2.31 (1.50) for Nmb Male Lifetime Arrests	Admin (limitations)
Heckman et al 2009	Perry	Table 2 Row 20 Col 5-6	13 (7.7) pp	95	Arrest	Men	By Age 40	9 (14) pp for Women	Admin (limitations)
Anderson 2008	Abecedarian	Table 10 Row 1 Col 9	-8.9 (13.3) pp	34.8	Conviction	Men	By Age 21	-10.1 (7.9) for Women -11.3 (11.7) for Male Felony Convictions	Self Reported
Campbell et al 2012	Abecedarian	Table 5 Row 1	-1.64 (9.0) pp (*)	28.6	Conviction	Overall	By Age 30		Self Reported
Garcia et al 2019	Abecedarian	Table 1	-2.27 (8.41) pp (*)	51.3 (*)	Arrest or Sentencing	Overall	By Age 34		Administrative

Note: Raw estimates used to construct the estimates listed above in Table B2 and depicted in Figures A1. For each study the estimate referenced in column (3) is reported in column (4) as a percentage point impact with standard errors in parentheses. In column (5) control group means are reported in percentage point terms. * In the case of Deming (2009), the mean was not reported or easily inferred. In Johnson and Jackson (2019), the "ever jail" variable appears to have been constructed using "non-response" reports of incarceration from family members that indicate whether a family member reported another remember being in a prison or jail at the point of the sample interview (which occurred annually until 1999 and roughly every other year thereafter), perhaps combined with retrospective self-reports in a 1995 criminal history survey. In the case of Garcia (2019), authors inferred the estimates and standard errors using reported counts by treatment and outcome status and performing a basic t-test. In the case of Campbell (2012), authors inferred the estimates from reported odds ratios and reported counts of convictions or reported odds ratios to percentage point terms. For both Garcia (2019) and Campbell (2012), standard errors are derived from basic t-tests. In column (9), alternative measures studied in the paper are reported in the same format as in column (4). Column (10) lists whether or not the criminal measure in the study was self-reported by a survey respondent or obtained through administrative data. In both Anderson (2008), Heckman et al (2009), and Garcia et al (2019), limitations on the success of the administrative data linkages are noted. In Garces et al (2002), we use the family fixed effect estimate of Head Start, rather than the difference between Head Start and other preschool attendance, to ease comparison across studies. Lastly, for Miller et al (2020), we report both the authors replication of Garces et al (2002), as well as the authors own estimates using an empirical strategy that reweights the FFE estimate to recover the ATE in the same dataset. In each row, the estimate reported in the figure and listed in this Appendix's Table B2 is calculated by scaling the estimate reported in column (4) of this table by program costs in 2015 dollars.

Appendix C: Simulation Exercises

One of the key advantages of our administrative data relative to commonly used survey data sets (i.e. PSID and NLSY79) is the larger number of individuals supporting the resulting estimates (Appendix Figure B2). In particular, the larger number of individuals within counties (an average of 5000 versus an average of 4-9) generates a significant increase in our ability to identify true effects on criminal behavior as well as the precision of the resulting estimates. We demonstrate the advantages of increasing within county sample size via a series of simulation exercises based on random draws from an augmented version of our actual North Carolina data set.

We begin by taking our actual North Carolina data set of convicted individuals and adding in a row for each individual born in each county and year who was not convicted (backed out of the natality data). We focus on the sample of high-poverty counties, augmenting this data set by adding 50 copies such that the resulting data set contains 1,122 counties, a rough estimate of the number of counties available in the PSID (Johnson and Jackson 2019). We think of the resulting data set as representing the full population of individuals born in these counties in the relevant birth years, with a separate row for each individual indicating their year and county of birth as well as whether or not they were convicted.

To approximate the sampling designs of the PSID and NLSY we then repeatedly draw stratified random samples of size $n \in \mathcal{N} = \{2700, 4536, 10000, 50000\}$ from this data set, drawing from each county of birth in the same proportion as the county populations. The first two elements in $\mathcal{N}$ correspond to the sample sizes used in recent studies of Head Start that rely on rollout designs in the NLSY and PSID, respectively; the larger sample sizes were chosen arbitrarily. The chosen samples resulted in corresponding average sample sizes within county of 2.41, 4.04, 8.91, and 44.56, with significant variation across counties depending on the underlying population. For each sample size we drew 200 stratified random samples from the augmented data set and ran our primary specification using the true assignment of Head Start entry. The resulting distribution of estimates is displayed in red in Figure C1, with the vertical red line to the left of zero indicating the effect size in the population (the augmented data set). At the sample sizes available in the PSID (~ 4536), the distribution of estimated coefficients remains very wide, with reductions in crime of as large as 5.8 percentage points as well as increases in crime of 3.0 percentage points falling within the range of 95% of the estimates.

While the confidence intervals are quite wide for the smaller samples, it is perhaps informative to compare them to the simulated distribution of estimates when the expected effect is zero. To do this, we randomly assigned year of Head Start entry 100 times, drawing 20 stratified samples for each, and estimated our main specification. The resulting distribution of estimates is plotted in grey (Figure C1). As expected, it is centered around zero. While the distribution of true estimates is shifted somewhat to the left of the placebo estimates, it is clear from the figure that for many negative and reasonably sized point estimates our likelihood of obtaining them in small samples is not that much greater when there is no effect as when there is a substantial one.

Figures C2 and C3 perhaps make this point more clearly in terms of obtaining a sta-

tistically significant effect. As expected, the distribution of placebo p-values in Figure C2 is relatively flat under the assignment of placebo entry years, with only 5 percent of the estimates having a p-value of 0.05 or less across sample sizes (as expected). While the the mass of low p-values is larger under the true assignment (when there is an effect), it is only modestly larger for the smaller samples available in the PSID and NLSY79. Indeed, Figure C3 demonstrates that at these low sample sizes and with the true assignment (when there is an effect) we can only reject the null of no effect at the 5 percent level roughly 8 percent of the time.[77] If we assume that the ex-ante odds of there being a true effect or not being an effect are equivalent (i.e., balanced priors), we will falsely claim that a significant effect is a true effect $5/(5+8) \approx 38$ percent of the time. Figure 3 illustrates how this false discovery rate falls as we increase the sample size (and thus the number of individuals within each county).

A second and perhaps obvious point is that in the smaller samples (where the probability of a false positive report is higher), we will be more likely to achieve statistical significance when the estimated effect is large, suggesting an upward bias in the magnitude of *published* effects when there is a preference for statistically significant results.[78] As an illustration of this point, we have estimated our baseline specification in the National Longitudinal Survey of Youth 1979 (NLSY79). We focus on "ever in jail" because conviction information is not available in the data. In Table C1, we present the resulting estimates for the full sample and the set of high poverty counties. The point estimates in the full set of counties are actually positive. Despite containing over 1,500 counties of birth, the standard error of the estimate for the full set of counties here is three times the standard error for the estimates from the North Carolina data (Columns (1) and (2)). The estimate for high poverty counties is negative, but again the standard error is more than four times the standard error for the estimates from the North Carolina data (Columns (3) and (4)).

The key takeaway then is that the use of significantly larger sample sizes (such as those available in our administrative, North Carolina data) reduces the likelihood that a statistically significant estimated effect is a false positive, reduces the expected upward bias in *published* effects, and increases precision. The increase in precision is particularly important in estimating heterogeneity across types of counties. In our context, we can be more confident that early childhood education reduces criminal behavior, this effect is likely considerably smaller than some prior estimates suggest, and the effect appears to be concentrated in counties with high poverty levels.

In contrast, one of the advantages of the smaller survey data sets is additional background information on individuals. This information allows researchers to focus in on subpopulations where we might expect the effect to be larger due to an increased likelihood of program eligibility or participation. However, this focus comes at a cost as analytic samples get even smaller. In Table C2, we reproduce the main estimates for various subgroups with greater eligibility for Head Start.

[77] In other words, statistical power is 0.08.

[78] It is worth noting that the type of potential bias we refer to is likely to occur when there is a general preference among referees and editors for statistically significant effects. It suggests nothing about the internal validity of the underlying research strategies.

Again, the estimates are mixed and imprecise. In no case can we reject the null of no effect. Nor can we reject the effect sizes implied by our estimates using the North Carolina data. The results are largely uninformative as the study is simply underpowered to detect effect sizes of modest magnitude.

Our data have advantages not only over similar rollout approaches in survey data sets, but also over evaluations of small scale experiments and family fixed effect designs. As depicted in Figure A1, we have precision advantages over the small-scale experiments and family fixed effects approaches as well. All of the same relative advantages regarding false positives, an upward bias in reported effects, and precision apply here as well. In Appendix Table C3 we summarize the statistical power of each study to identify the average published effect size, following the guidance provided in the meta-analysis literature.[79] For example, to identify the simple mean effect size per $1,000 in the literature, the statistical power of the Head Start studies ranges from 0.09 to 0.13. Under balanced priors on the likelihood of an effect this implies false discovery rates of 42 to 54 percent. These types of concerns regarding effects on criminality are not limited to the Head Start literature; the randomized evaluations of the more intensive interventions are only modestly better powered. While we cannot provide a concrete example in the context of the small-scale experiments, it is illustrative that different evaluations of the *same* experiment have reached different conclusions regarding the effects on crime (Heckman et al., 2010; Anderson, 2008). Furthermore, the lack of statistical power perhaps explains the apparent lack of crime effects attributed to the Abecedarian intervention.

In the case of family fixed effects, the confidence intervals in Figure A1 illustrate the lack of statistical power of this approach. This problem is emphasized by Miller et al.'s (2020) inability to replicate Garces et al.'s (2002) crime effect estimates for Head Start using the same family fixed effects strategy and same data source. We have estimated similar models in the NLSY79 and again find positive estimates of Head Start participation on the likelihood of incarceration (Table C4).[80] While the outcomes and ages are somewhat different than those used in prior family fixed effect designs, this is quite surprising given prior estimates from the same general time period.[81] However, the inconsistency is resolved when one considers the large confidence intervals associated with this approach in these samples. In Figure C4 we demonstrate visually the relative advantage of our data and strategy over implementing the family fixed effects (FFE) and rollout approach in the NLSY79.

As a concluding point, it is worth emphasizing that all of these studies made large leaps forward in our understanding of the effects of early childhood education, making dramatic improvements in design over prior studies and providing innovative answers to previously unanswered or poorly answered questions. The above discussion is not intended to suggest

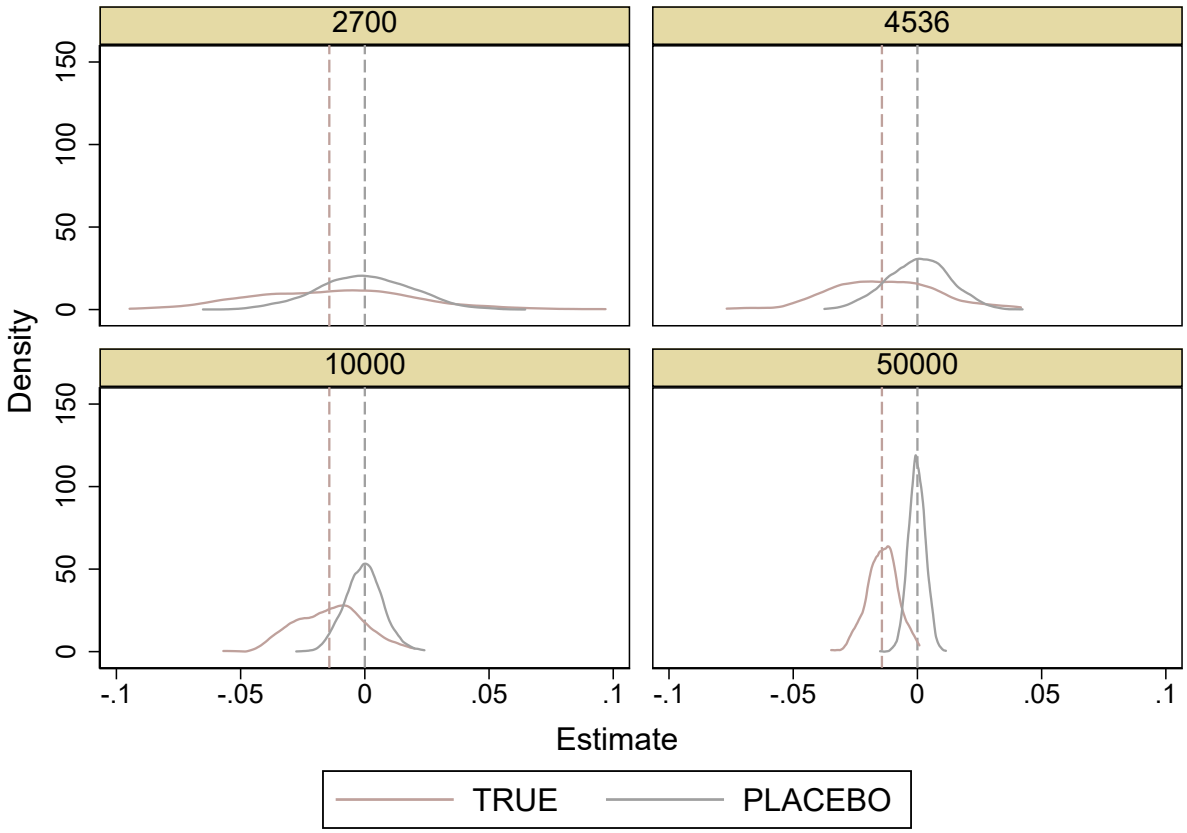
[79] This will be an overestimate of the true average effect size in the presence of a publication preference for statistically significant effects.

[80] We present estimates for blacks separately as earlier evidence suggested larger effects for this subgroup.

[81] Earlier studies use some combination of charges, arrests, and convictions to construct their binary criminal involvement variables. We have run similar specifications with an "ever charged" specification in the NLSY79 and find similar positive coefficients, although this question is only asked to individuals in 1980, when most are relatively young.

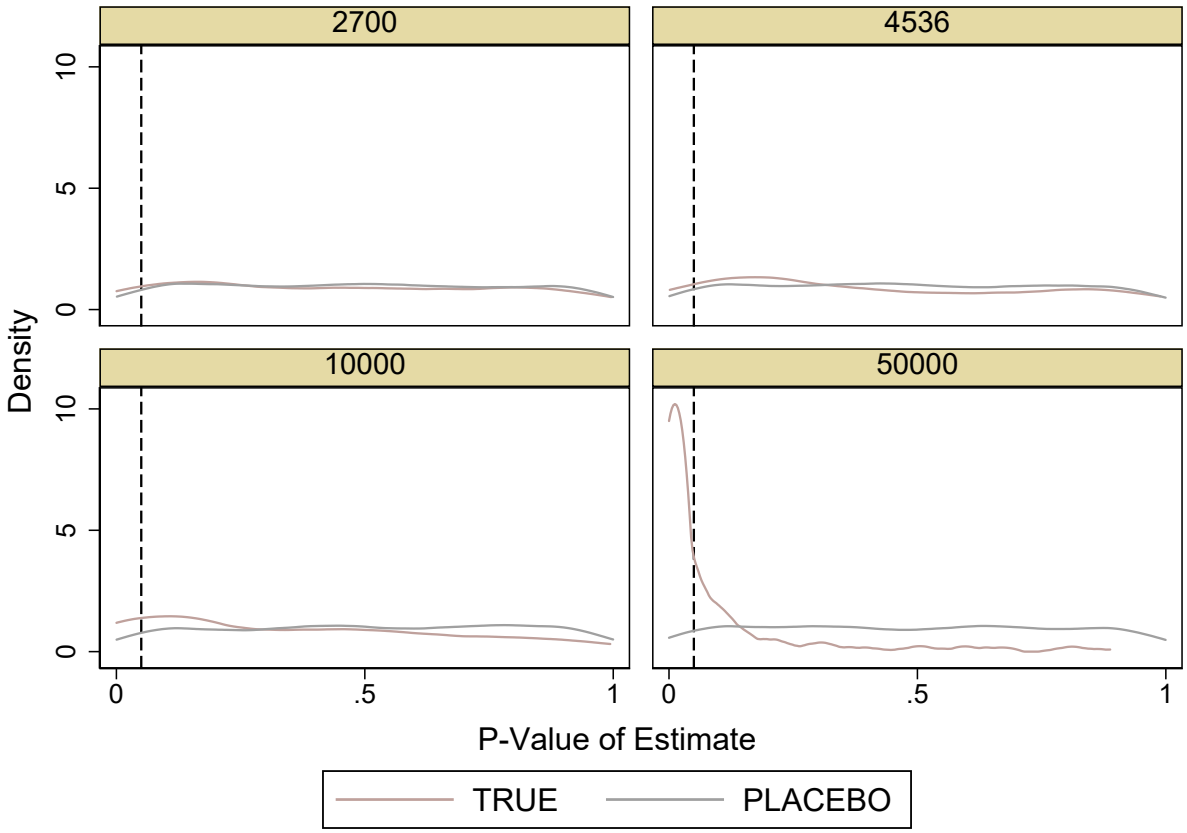
that estimates from prior path-breaking studies are wrong, only to highlight that limitations of the data used in these studies leave substantial remaining uncertainty regarding the true effects of these early childhood education programs on adult criminal behavior.

Figure C1: Distribution of Simulated Estimates by Sample Size



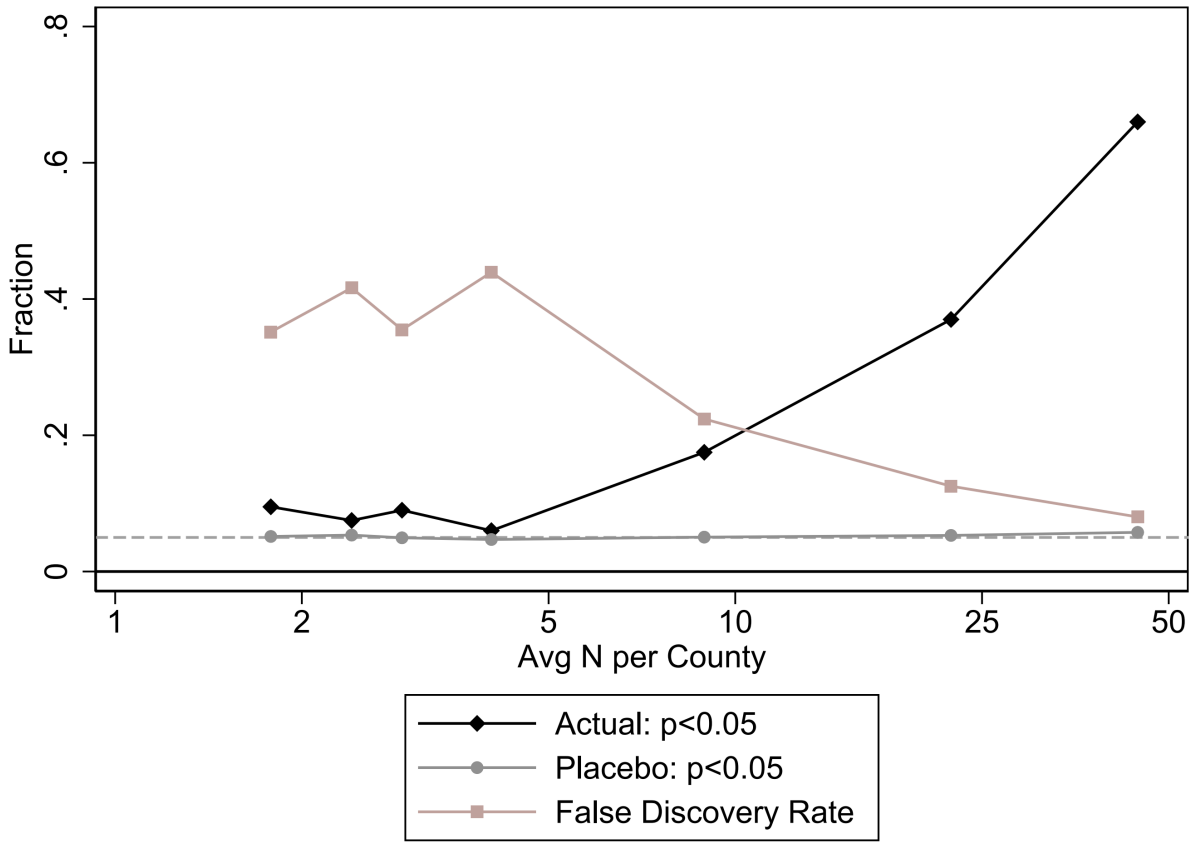
Note: Figure shows a distribution of estimates for our primary specification obtained for the true assignment of Head Start rollout (red) and for placebo assignments of Head Start rollout (grey) from drawing stratified random samples of various sizes from an augmented version of our North Carolina data set. For the true rollout assignment we draw 200 stratified random samples and for the placebo assignments we draw 20 stratified random samples for each of 100 random placebo rollout assignments. The augmented data set includes all convicted individualssin the administrative North Carolina crime dataset born in high poverty counties, uses natality files to add observations for non-convicted births in those counties, and then appends 50 additional copies of the data to increase the number of counties to reflect national survey data. To approximate the sampling designs of the PSID and NLSY, we repeatedly draw stratified random samples of size $n \in \mathcal{N} = \{2700, 4536, 10000, 50000\}$ from this data set, drawing from each county of birth in the same proportion as the county populations. The first two elements in $\mathcal{N}$ correspond to the sample sizes used in recent studies of Head Start that rely on rollout designs in the NLSY and PSID, respectively; the larger sample sizes were chosen arbitrarily. The chosen samples resulted in corresponding average sample sizes within county of 2.41, 4.04, 8.91, and 44.56. Each panel header identifies the overall size of the stratified random sample. Vertical dashed lines indicate the estimated effect for the full populationof the augmented data set.

Figure C2: Distribution of P-values by Sample Size



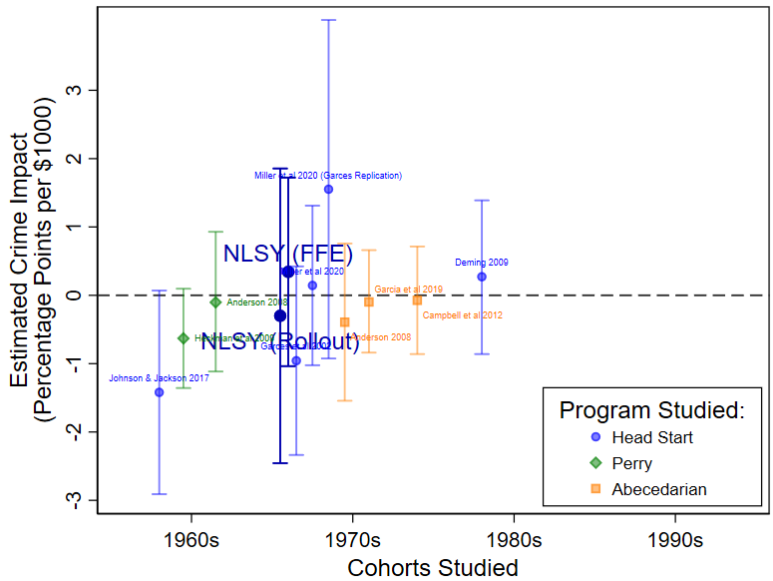
Note: The Figure depicts the distribution of p-values from the estimates in Figure C1. Dashed line shows p-value of .05. See the note to Figure C1 for more details.

Figure C3: Rejection Rates and Implied False Discovery Rate

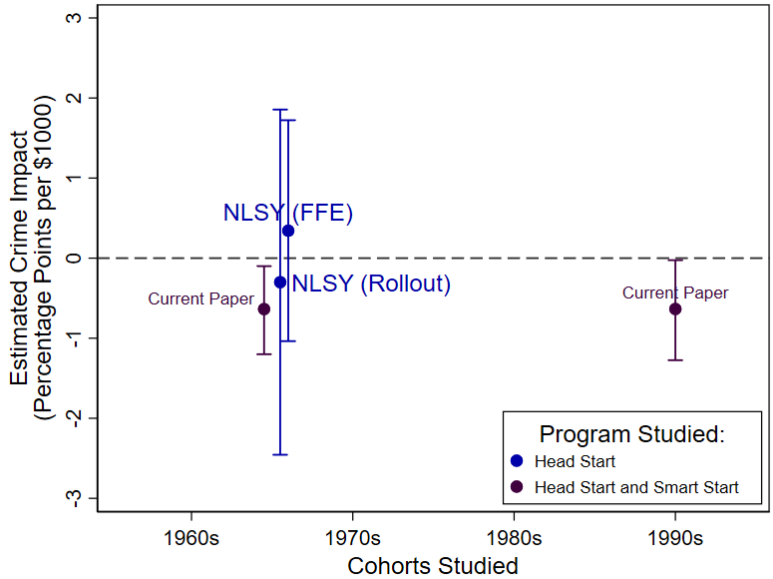


Note:
The Figure depicts the fraction of estimates of our primary specification from a series of random stratified samples of various sizes which are significant at the 5 percent level under the true assignment of Head Start rollout (black), under a random (placebo) assignment of Head Start rollout (grey), as well as the resulting rate of false discovery (red). These fractions are graphed on the y-axis against the average sample size per county for the sample draws that generated the fraction on the x-axis. The false discovery rate is defined as the number of false discoveries divided by the sum of false discoveries and true discoveries. This false discovery rate assumes that the ex-ante odds of there being a true effect and there being no effect are the same (i.e. balanced priors).

Figure C4: Effect Size Comparison: NLSY Comparison



(a) NLSY and Existing Literature



(b) NLSY and Current Paper

Note: Figure repeats Figure A1, this time adding two estimates obtained from using two different identification strategies in the NLSY79 dataset. The estimate labeled "NLSY (FFE)" replicates the family fixed effect strategy in Garces et al 2002 and Miller et al 2020. The estimate labeled "NLSY (Rollout)" replicates the strategy employed in the main analysis of this paper which identifies effects based on when Head Start funding became available to a county birth cohort.

Table C1: Head Start and Ever in Jail (NLSY)

VARIABLES	(1) All	(2) All	(3) High Poverty	(4) High Poverty
HS in County	0.006	0.004	-0.006	-0.013
	(0.009)	(0.009)	(0.022)	(0.024)
Covariates		X		X
Observations	9,617	9,617	1,500	1,500
Mean	0.0597	0.0597	0.0633	0.0633

Note: Table reports estimates of our baseline specification (reported above in main Table 2) using data from the National Longitudinal Survey of Youth 1979 (NLSY79). The outcome variable is an indicator variable for whether the respondent was "ever in jail". (Conviction information is not available in the NLSY79.) We present estimates on the full sample of counties in Columns (1) and (2) and the high poverty counties in Columns (3) and (4). See the note to main Table 2 for more details

Table C2: Head Start and Ever in Jail (NLSY)

VARIABLES	(1) Poverty (79)	(2) Poverty (79)	(3) Mom $\leq$ HS	(4) Mom $\leq$ HS	(5) Mom $<$ HS	(6) Mom $<$ HS
HS in County	-0.001	-0.015	0.004	-0.0001	0.028	0.019
	(0.024)	(0.027)	(0.011)	(0.011)	(0.018)	(0.019)
Covariates		X		X		X
Observations	1,539	1,539	5,086	5,086	2,670	2,670
Mean	0.111	0.111	0.0701	0.0701	0.0855	0.0855

Note: Table reports estimates with the same specification as in Table C1 above, but now broken down for various subgroups known to have greater Head Start eligibility. See the note the Table C1 for more details.

Table C3: Statistical Power and Implied False Discovery Rates to Identify Mean Effect Size

Study	Power	FPR
Deming 2009	0.087	0.54
Garces et al 2002	0.13	0.43
Johnson and Jackson 2017	0.095	0.51
Anderson 2008	0.15	0.40
Heckman et al 2009	0.26	0.28
Anderson 2008	0.13	0.43
Garcia et al 2019	0.20	0.33
Campbell et al 2012	0.26	0.28

Note: Power is calculated to identify the mean effect per $1,000 (2015 dollars) across studies (0.49 percentage points per $1,000). The false discovery rate (FDR) is calculated under balanced priors of the likelihood of an effect.

Table C4: Head Start and Ever in Jail (NLSY)

VARIABLES	(1) All	(2) All	(3) Black	(4) Black
Self Reported HS Participation	0.021	0.019	0.052	0.057
	(0.038)	(0.039)	(0.055)	(0.059)
Self Reported Preschool Participation	0.001	0.002	-0.021	-0.003
	(0.026)	(0.027)	(0.052)	(0.056)
Covariates		X		X
Observations	9,349	9,349	2,614	2,614
Mean	0.0597	0.0597	0.115	0.115

Note: Table reports estimates obtained using a family fixed effect strategy as in Garces et al (2002) and replicated in Miller et al (2020), but in the NLSY79. Each column reports estimates from a separate OLS regression. The outcome variable is an indicator variable for whether the respondent was "ever in jail". The explanatory variables of interest are indicators for whether the respondent reported attending Head Start ("Self-reported HS Participation") and whether the respondent reported attending some other preschool ("Self Reported Preschool Participation"). Estimates are performed separately for all respondents in columns (1) and (2), and for Black respondents in columns (3) and (4).

Appendix D: Implied Treatment on the Treated

These TOT estimates are based on estimated Head Start participation rates in high poverty counties of 15 to 21 percent. The lower bound is based on OEO statistics on state-level North Carolina Head Start enrollment in 1966 and the upper bound is based on author's calculations assuming the national per participant funding level is fixed across North Carolina counties. Our implied TOT effects are between half and two-thirds of the size of effects on somewhat similar measures reported in evaluations of the Perry Preschool program (11 to 12 percentage points on any arrest (or any charges) by age 40).[82,83] As in the Perry evaluation, we find larger effects on property crimes; Head Start access reduces the likelihood of a serious property conviction by 0.8 percentage points, a TOT effect of 4 to 5 percentage points in high-poverty counties (Appendix Table A17). While there is no significant effect on serious violent convictions, the point estimate (0.0046) implies a TOT of 2 to 3 percentage points. In comparison, Schweinhart et al. (2005) find a 16 percentage point reduction in violent arrests by age 40 (32 versus 48 percent) and a 22 percentage point reduction in property arrests by age 40 (36 versus 58 percent) in their evaluation of Perry preschool, four to five times the size of our effects.[84] Perry Preschool enrolled a very particular type of student: extremely disadvantaged, black children in Ypsilanti, Michigan. If we split our property crime estimates by race, we find similar effects for whites and non-whites.

If there are important spillover effects of program availability it is not reasonable to interpret these scaled estimates as TOT effects. In this case, improving the behavioral trajectories of a significant share of a group results in improvements for the group as a whole that are substantially larger than what we might expect to see if an individual was treated in isolation. Especially in high-poverty areas, a substantial fraction of children enrolled in Head Start. As these children interacted with others in their cohort, effects of the program might have spilled over to the other children in a way that would have been unlikely with the smaller treatment and control groups in experimental evaluations of small-scale programs. Particularly when it comes to criminal behavior, it is likely that these spillovers, operating through peer effects, are substantial. We therefore focus our discussion on the estimated effects of Head Start availability rather than participation.

[82] The treatment effect of Perry Preschool on any felony arrest, the definition of which overlaps substantially with Part 1 crimes, is even larger (15 percentage points), but is reported only for males (Heckman et al. 2009).

[83] Our TOT estimates are less than half of the effects estimated for the Nurse-Family Partnership by age 19 (16 percentage points on likelihood of conviction or arrest) (Olds et al. 1998, 2007). Our effect sizes are similar to recent estimates of the effects of early childhood Food Stamp access (Barr and Smith 2018). In contrast to both of these health interventions, which found strong effects on violent criminal behavior, the effects of Head Start access appear to be somewhat stronger on property crimes.

[84] Although we note that these are effects on *any* arrest and thus may not be directly comparable to convictions for a serious violent or property crime. Treatment estimates of Perry Preschool on the *number* of felony arrests indicates no significant difference in the number of serious violent crimes and a 90 % reduction in the number of felony property arrests (0.31 versus 2.91 per individual).

Appendix E: Event Study Heterogeneity

Recent work shows that when there is a staggered rollout of treatment, standard two-way fixed effects regressions provide event study estimates that are a weighted average of the county rollout-cohort specific effects (i.e., the effect from the set of programs adopted at time $t = 1965$, at $t = 1966$, etc.). In the presence of heterogeneous effects, this can lead to "contamination bias", where the coefficient on a given lead or lag can be contaminated by effects from other periods (Sun and Abraham 2020). This can lead to incorrect estimates of the average effect of a program as well as pre-treatment indicators that appear to be zero (i.e., no pre-trends) that are actually non-zero once the contamination of effects from other periods has been accounted for. Guided by the recent econometric and applied literature, we do three things to address the concern of contamination bias in our setting:

1. We implement a specification with a simple and intuitive control group that is not contaminated (i.e., the counties serving as the control have not been treated at the time they are used as control units). This allows us to estimate a treatment contrast without any county rollout-cohort heterogeneity and eliminates concerns about contamination bias.

2. We implement Sun (2020)'s techniques to build "interaction-weights" (IW) by event study indicator (i.e., each lead and lag) and county rollout-cohort groups. We then examine these weights for evidence of heterogeneity following prescriptions in Sun et al (2020). In the absence of heterogeneity, one need not be concerned about contamination bias influencing the estimates produced using the standard two-way fixed effects design.

3. We implement Chaisemartin (2018)'s techniques to produce dynamic treatment effect estimates which accommodate and account for treatment heterogeneity.

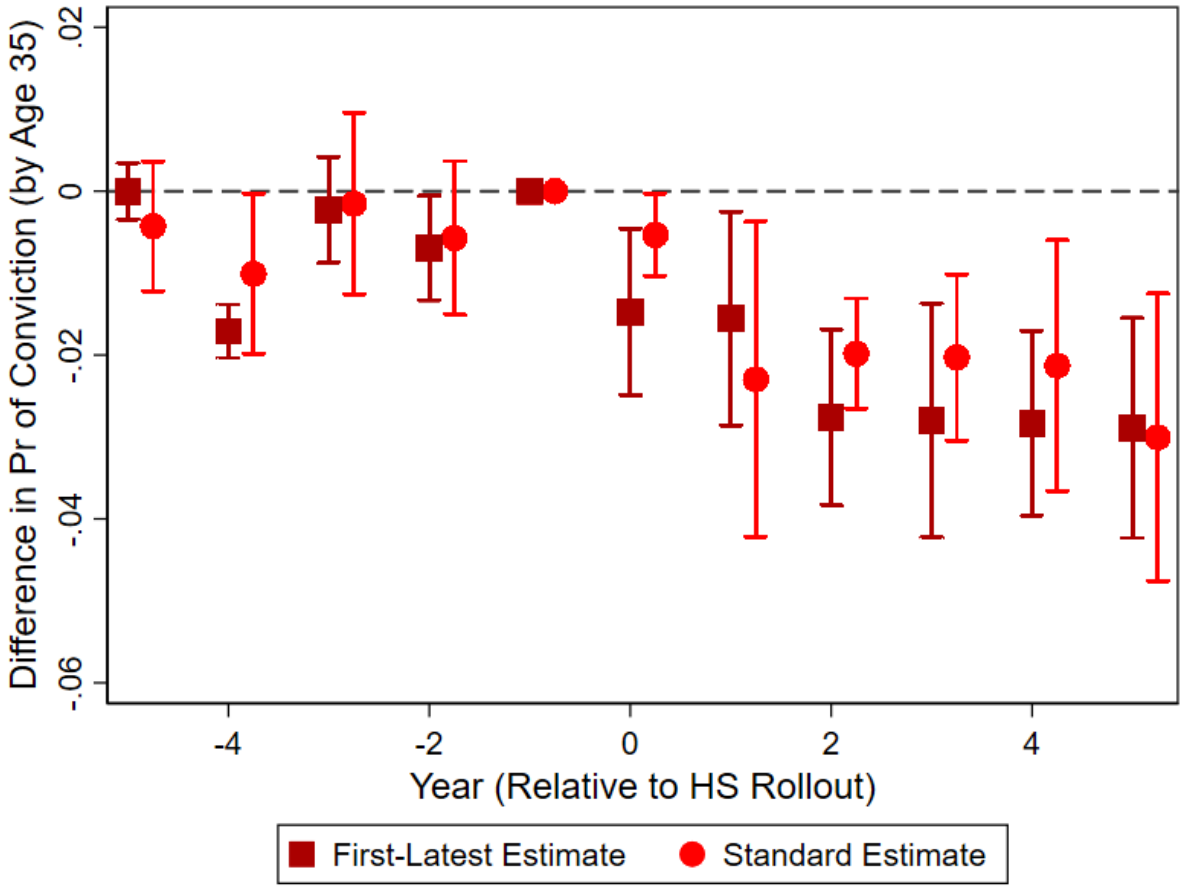
Overall, we find no evidence of contamination bias; if anything, heterogeneity in the county rollout-cohorts is slightly attenuating our standard estimates towards zero.

1. Simple, "Clean" Control Groups

Guided by the recent econometric and applied literature, we begin by presenting estimates from a simple and intuitive control group, which allows us to isolate a treatment contrast using a clean control group and intuitively test whether there are heterogeneous effects across county rollout-cohorts. Specifically, we focus on the first treated rollout-cohort (counties in which the first birth cohort exposed to Head Start was born in 1961) and include as a control group only the latest rollout-cohort (counties in which the first birth cohort exposed to Head Start was born in 1967 or 1970). To avoid contamination bias, we drop birth cohorts born after 1966 to ensure that the control county birth-cohorts are never treated in the estimation sample.

In other words, in the "first-latest" specification, the control group remains untreated throughout the sample time period. As Figure E1 shows, the results are similar to our main

Figure E1: Event Study of Head Start's Impact on Criminal Conviction – "First-Latest" Estimates



Note: "First-Latest Estimates" are the same as our standard estimates in Figure 2 except that they feature a control group consisting of the latest two county rollout-cohorts (1970 and 1967). Thus the control group consists of county-birth-cohort which are never treated in the sample period (but are treated just after the sample ends).

results. While the -4 estimate is statistically significant in this "first-latest" specification, the overall post-period results are slightly larger and suggest a sharp level drop. Overall, the results suggest little difference in the estimated effects when focusing on a clean difference-in-difference estimate of the effect of Head Start for the first county rollout-cohort, which accounts for much of our identifying variation. If anything, the estimates suggest a somewhat larger effect of treatment for the first county rollout-cohort, which would attenuate our main results towards zero.

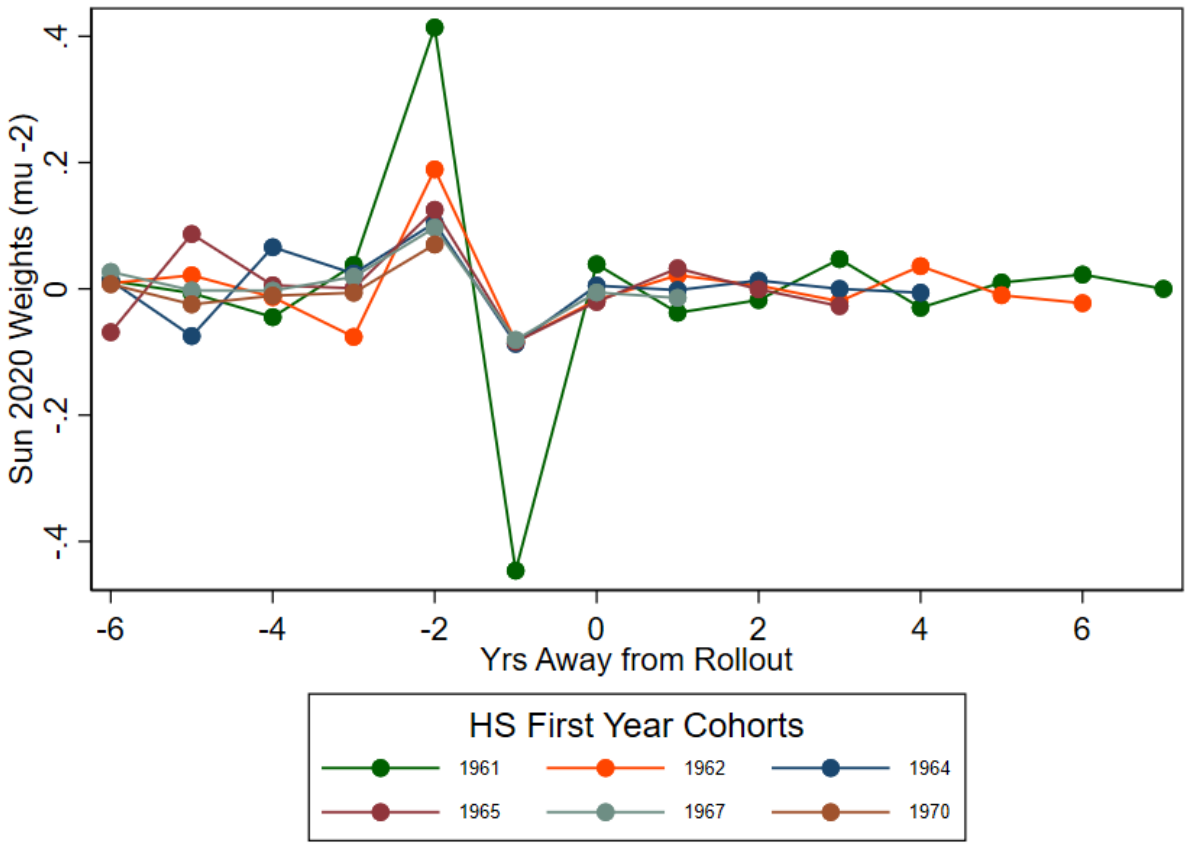
2. Diagnosing Contamination Bias using IW weights

An alternative approach to diagnosing the importance of treatment group heterogeneity is to evaluate the contribution of different county rollout-cohort groups and lead and lag effects to the estimate of each lead and lag coefficient (Sun 2020). Intuitively, if lead and lag effects are contributing in meaningful ways to the identification of other leads and lags, it implies that contamination bias is meaningful. We follow Sun (2020) and produce IW weights for the -2 estimator (two periods before treatment) to investigate the presence of "contamination bias" (Figure E2 below). We then confirm that, for the -2 estimator, the sum of the IW weights over treatment lags is zero for each rollout-cohort (Table E1 below).

The details are as follows: we implement Sun (2020)'s techniques to build "interaction-weights" (IW) by event study indicator and county rollout-cohort. An IW is a measure of a county rollout-cohort's contribution to the estimation of a given coefficient in the standard event study specification. For example, for the -2 estimator (two periods before treatment), we compute an IW weight for the first county rollout-cohort for each event study time period (-6, -5, etc.) and then for the second county rollout-cohort and so on. IW weights could show, e.g., that, in a given event study year, the 1st county rollout-cohort is contributing more to the -2 estimate than the 2nd county rollout-cohort; further, it could show, for example, that event study time periods other than -2 are contributing importantly to the identification of time period -2. We use these IW weights to test whether contamination bias is occurring in our data. If, for example, we see significant weight placed on event study time period 3, 4, or 5 (in identifying the estimate at -2), it would suggest contamination bias.

Figure E2 shows the IW weights estimated using methods in Sun (2020). Following our main specification, we compute the weights for our main sample (which are the ever-treated, high-poverty counties for birth cohorts 1955-1968). In our main specification we pooled all 6 county rollout-cohorts together, but for IW weights we break them apart. Following Sun (2020), we report the weights for the estimate of the indicator variable for two periods before treatment (-2 in our event study time) across rollout-cohorts. First, we note that the weights for each rollout-cohort are largest at event time=-2 with the first rollout-cohort contributing the most weight. Since these are weights for the -2 estimator, we expect each cohort to contribute a positive weight to the estimates of the -2 coefficient, and we expect the first rollout-cohort to have the most weight since more of the treated sample experienced treatment in the first rollout than in any other rollout-cohort. Secondly, we also expect the weights at -1 to be negative, since this is the excluded period in this specification. Finally, we note that the weights on the post-treatment indicators hover around zero and the sum of the post-treatment weights is very nearly zero (i.e. the sum of the IWs from our event study time =0 to =6). As Sun and Abraham (2020) notes, if the "weights are non-negative for lags of treatment", then this might suggest that the standard event study estimate for two periods before treatment is "sensitive to estimates of the dynamics effects ... and does not isolate the pre-trends" (p.33). Fortunately, as Table E1 below shows, our IW weights for lags of treatments are small and sum to $\leq .01$ for each county rollout-cohort. Overall, these IWs suggest that our estimate of the -2 indicator is not contaminated by rollout-cohort heterogeneity.

Figure E2: IW weights following Sun (2020)



Note: IW weights following Sun (2020). -1 is the excluded period.

Table E1: County Rollout-Cohort IW Weight Sums for Treatment Lags

County Rollout-Cohort	Sum of IW weights over Treatment Lags
1st Rollout	-.006
2nd Rollout	.010
3rd Rollout	005
4th Rollout	.005
5th Rollout	-.014

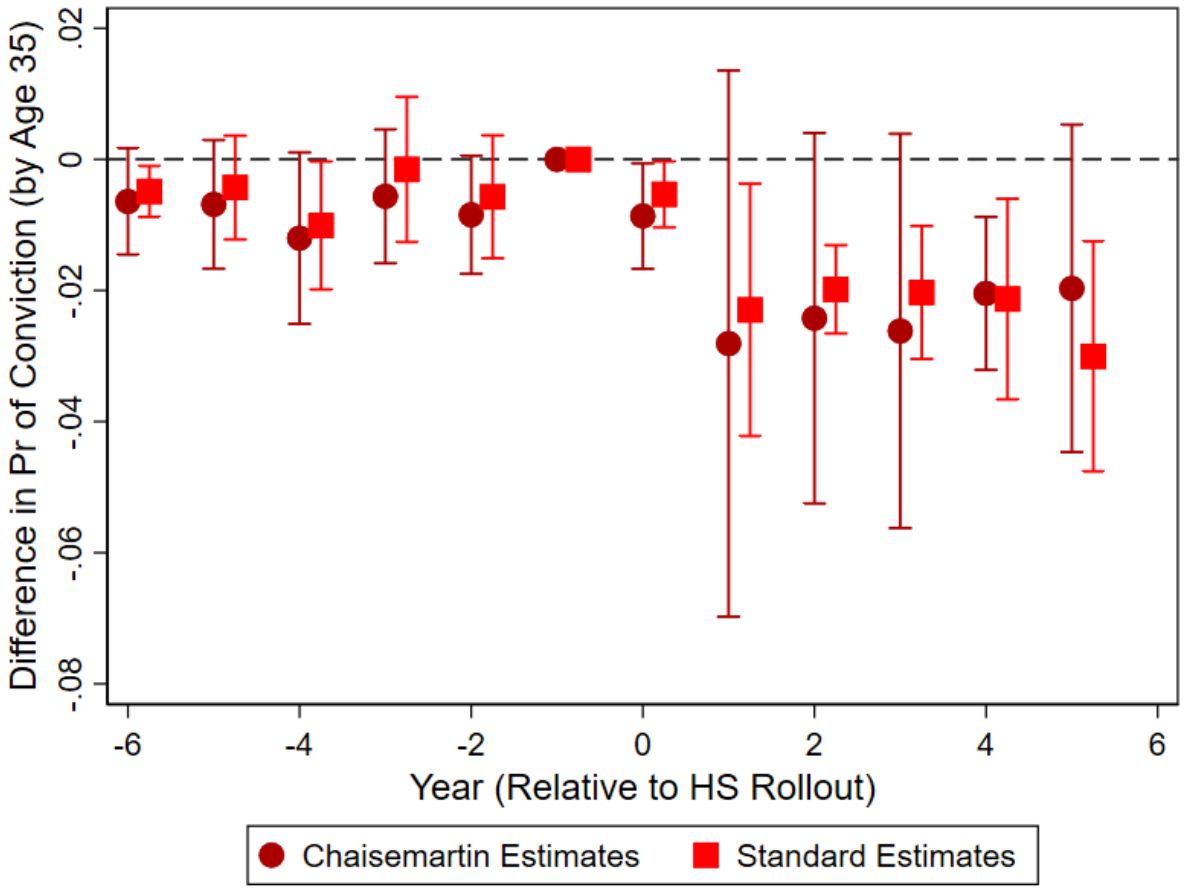
Note: Each row reports the sum of IW weights over treatment lags. Weights are calculated by implementing Sun (2020).

3. Accounting for Treatment Heterogeneity

While we have already provided evidence suggesting the lack of important treatment effect heterogeneity, another approach is to account for this treatment effect heterogeneity directly

in the estimation procedure. To achieve this, we implement the techniques of Chaisemartin (2018) to produce dynamic treatment effect estimates which directly estimate heterogeneous dynamic treatment effects and weights the corresponding effects together into the estimates (Figure E3).

Figure E3: Event Study of Head Start's Impact on Criminal Conviction – Treatment Heterogeneity



Note: Chaisemartin estimates analogous to our standard estimates in Figure 2 except that they account for the possibility of heterogeneous dynamic treatment effects.

Overall, Chaisemartin estimates for pre-treatment periods do not suggest differential movement of treatment and control groups prior to treatment. Period "0", the first period of treatment is when the first significant estimate appears. Furthermore, Chaisemartin post-treatment estimates give estimates very similar to our standard event study estimates. Chaisemartin estimates suggest an overall reduction of .0169, while our estimates suggest a reduction of .0131 (Row 1, Col 2 of Table 2). While the Chaisemartin estimate is not

significantly different from the standard estimate, this further suggests that, if anything, heterogeneity in the county rollout-cohorts is slightly attenuating our standard estimates towards zero.