

HCEO

hceconomics.org

**Human Capital and Economic Opportunity
Global Working Group**

Working Paper Series

Working Paper No. 2014-010

**Long-Term Impacts of Compensatory Preschool on Health and Behavior:
Evidence from Head Start**

Pedro Carneiro
Rita Ginja

April, 2014

Human Capital and Economic Opportunity Global Working Group
Economics Research Center
University of Chicago
1126 E. 59th Street
Chicago IL 60637
www.hceconomics.org

Long-Term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start

Pedro Carneiro * Rita Ginja^{†‡}

October 24, 2013

Abstract

This paper provides new estimates of the medium and long-term impacts of Head Start on health and behavioral problems. We identify these impacts using discontinuities in the probability of participation induced by program eligibility rules. Our strategy allows us to identify the effect of Head Start for the set of individuals in the neighborhoods of multiple discontinuities, which vary with family size, state and year. Participation in the program reduces the incidence of behavioral problems, health problems and obesity of male children at ages 12 and 13. It lowers depression and obesity among adolescents, and reduces engagement in criminal activities and idleness for young adults.

*University College London, CEMMAP, and IFS

[†]Uppsala University and UCLS.

[‡]Corresponding Author: Rita Ginja. Email: rita.ginja@nek.uu.se. Address: Department of Economics, Uppsala University, Box 513 SE-751 20 Uppsala, Sweden. We thank Joe Altonji, Sami Berlinski, Richard Blundell, Janet Currie, Julie Cullen, Michael Greenstone, Jeff Grogger, James Heckman, Isabel Horta Correia, Maryclare Griffin, Hilary Hoynes, Mikael Lindhal, Jens Ludwig, Costas Meghir, Robert Michael, Kevin Milligan, Lars Nesheim, Jesse Rothstein, Chris Taber, Frank Wjindmeier, and seminar participants at IFS, 2007 EEA Meetings, Universidade Católica Portuguesa, Banco de Portugal, 2008 RES Conference, 2008 SOLE meetings, 2008 ESPE Conference, 2008 Annual Meeting of the Portuguese Economic Journal, 2009 Winter Meetings of the Econometric Society, 2011 Nordic Labor Meetings, 2011 Workshop in Labor Markets, Families and Children (Stavanger) and the CIM Working Group (University of Chicago) for valuable comments. Pedro Carneiro gratefully acknowledges the financial support from the Leverhulme Trust and the Economic and Social Research Council (grant reference RES-589-28-0001) through the Centre for Microdata Methods and Practice, and the support of the European Research Council through ERC-2009-StG-240910-ROMETA and Orazio Attanasio's ERC-2009 Advanced Grant 249612 "Exiting Long Run Poverty: The Determinants of Asset Accumulation in Developing Countries". Rita Ginja acknowledges the support of Fundação para a Ciência e Tecnologia and the Royal Economic Society.

1 Introduction

The need to cut public spending, together with recent disappointing evaluations of Head Start and Sure Start, have put severe pressure on compensatory preschool programs both in the US and the UK. Opponents call for the outright termination of these programs, while supporters argue that they are needed now more than ever, as increasing numbers of families fall into poverty. Others propose maintaining the programs, as long as they are subject to comprehensive reform.

The Head Start Impact Study (HSIS) gained prominence in this debate. It evaluates Head Start, the main compensatory preschool program in the US, and it is the first experimental study of a large scale preschool program in the world. The study shows that Head Start has short term impacts on the cognitive and socio-emotional development of its participants, which disappear by first grade. While there are grounds on which this study can be criticized (e.g., Zigler, 2010), its main findings are notorious because of its transparent and rigorous design.¹ In parallel, an evaluation of Sure Start in the UK, although non-experimental and less influential than the Head Start Impact Study, finds that Sure Start also has limited impacts on the development of poor children.

We study impacts of Head Start on children using data from the Children of the National Longitudinal Survey of Youth (CNLSY) and a novel identification strategy in the context of this program, which relies on the eligibility criteria to the Head Start. Our paper shows that in spite of the lack of program impacts by first grade, there are important longer term impacts of Head Start on the health and behaviors, and on criminal behavior of male adolescents and young adults.² Our results are in line with the growing literature on the effectiveness of early childhood interventions, which shows that these programs have large long-term impacts on

¹Another experimental evaluation of Early Head Start (DHHS, 2006), a program for children ages 0-3, also shows small program impacts.

²Relative to comparable non-participants, male Head Start participants are 29% less likely to suffer from a chronic condition that requires the use of special equipment (such as a brace, crutches, a wheelchair, special shoes, a helmet, a special bed, a breathing mask, an air filter, or a catheter), 29% less likely to be obese at ages 12-13, less likely to show symptoms of depression at ages 16-17, and 22% less likely to have been sentenced to a correctional facility by ages 20-21. For the two youngest groups we find a significant improvement in an index of behavioral problems and among children 12-13 we are also able to detect improvements on health.

behavioral problems even when they have limited short term impacts on cognitive development. Short term evaluations of early childhood programs miss most of their potential impacts.

We identify the causal effects of Head Start using a (fuzzy) regression discontinuity design which explores the eligibility rules to the program. We determine the eligibility status for each child aged 3 to 5, by examining whether her family's income is above or below an income eligibility cutoff, which varies with year, state, family size, and family structure. In contrast with standard applications of regression discontinuity, there are multiple discontinuity points in our setup, which vary across families because they depend on year, state, family size and family structure. Therefore, our estimates are not limited to individuals located in the neighborhood of a single discontinuity, but they are applicable to a more general population. Finally, given that we exploit the income requirements to be eligible to Head Start, our estimate can be interpreted as the potential gains of relaxing marginally these requirements. Given that the first stage is only significant for males, the marginal entrant is a boy.

Beyond the HSIS (DHHS, 2010), described above as showing little or no effect of the program, there exist several non-experimental evaluations of Head Start which are also important, and it is worthwhile mentioning some of the most recent ones. Currie and Thomas (1995, 1999) compare siblings in families where at least one sibling attends Head Start and one does not. In contrast to HSIS, they find strong impacts of the program on a cognitive test (which fade-out for blacks, but not whites) and grade repetition. They use the CNLSY, which is the data set used in this paper. Currie, Garces and Thomas (2002) apply a similar strategy in the Panel Study of Income Dynamics (PSID), and show that the program has long lasting impacts on schooling achievement of adults, earnings, and crime. Also, relying on within family comparisons and using the CNLSY, Deming (2009) finds no effects on crime but positive effects on a summary measure of children's test scores and adult outcomes.

Ludwig and Miller (2007) explore a discontinuity in Head Start funding across US counties, at the time the program was launched (1965). They show that Head

Start has positive impacts on adolescents' and adults' health and schooling.³

Relative to all these studies, we evaluate a more recent variant of the program (and employ a novel empirical strategy): individuals in our sample enrol in the program from the 1980s to the late 1990s. This is relevant because Head Start has changed over the years and its costs have dramatically increased, closely approaching the costs of model interventions such as Perry Pre-School or Abecedarian. Furthermore, it means that, relative to the studies mentioned above, ours is likely to be more comparable to the recent Head Start Impact Study, which examines children who applied for Head Start in 2002. Ludwig and Miller (2007), and Garces, Currie and Thomas (2002), study the effects of attendance between the mid-1960s and the 1970s. Currie and Thomas (1995), and Deming (2009), analyze effects of Head Start for those who attended the program during the 1980s.⁴

This paper proceeds as follows. In the next section we describe Head Start in more detail. We discuss the identification strategy in Section 3. We present the data in Section 4. Results are presented in Section 5. Section 6 concludes.

2 Background: The Head Start Program

Head Start was launched in 1965 and currently it provides comprehensive education, health, nutrition, and parent involvement services to around 900,000 low-income children 0 to 5 years of age (of which 90% were 3-5 years old in 2009)⁵

³In addition, Currie and Neidell (2007) use the CNLSY to study the quality of Head Start centers and find a positive association between scores in cognitive tests and county spending in the program. They also find that children in programs that devote higher shares of the budget to education and health have fewer behavioral problems and are less likely to have repeated a grade. Frisvold and Lumeng (2007) explore an unexpected reduction in Head Start funding in Michigan to show strong effects of the program on obesity. Neidell and Waldfogel (2006) argue that ignoring spillover effects resulting from interactions between Head Start and non-Head Start children and/or parents underestimates the effects of the program in cognitive scores and grade repetition. Finally, Anderson, Foster and Frisvold (2010) find that Head Start is associated with a reduction in the probability that young adults smoke.

⁴There exist a few studies in the literature examining the long-term impact of universal preschool (Cascio, 2009, Magnuson et al, 2007, Berlinski et al, 2008, 2009, Havnes and Mogstad, 2011). They concern programs that affect a much larger fraction of the population, and generally show long-term impacts of preschool availability.

⁵According to the Head Start Office, in 2009, among those 3-5 years old, 36% of children were 3 years, 51% were 4 and 3% were 5 years old.

and until 1994 the program served children ages 3 to 5) and their families. Since the program targets all disadvantaged preschool age children, it differs from two other prominent early childhood experiments, the Perry Preschool Program and the Carolina Abecedarian Project, which in the 1960s and 1970s (respectively) served each just over 100 disadvantaged Black children, who have been followed up into adulthood.⁶

Head Start is primarily funded federally but grantees must provide at least 20 percent of the funding, which may include in-kind contributions, such as facilities to hold classes. Thus, the scale of the program implies that different grantees are heterogeneous in several dimensions, such as costs of personnel and space (depending on geographic location, for example) and type of sponsoring agency (school system or private nonprofit). However, each center must comply with publicly known standards which are described in the Head Start Act.

Centers may offer one or more out of three program options: center-based option, home-based option, or a combination of both, chosen based on the needs of the children and families served. By 1996, basic standards required that center-based programs employ two paid staff persons for each class and recommendations regarding the class size varied with the age of children: 17-20 children for 4-5 years old classes and 15-17 children for 3-years old classes. These figures imply a higher student-teacher ratio than that model early childhood interventions as the Perry Preschool Program or the Carolina Abecedarian Project: Perry Preschool Program had a teacher-student ratio of one teacher for 5.7 students whereas the Carolina Abecedarian Project ranged from 3 children per teacher for infants to 6:1 at age 5 (Cunha, Heckman, Lochner and Masterov, 2006). The Head Start ratio is closer to that of the Chicago Child Parent Center and Expansion Program, which was another 1960s intervention that served between 8-12 children per teacher (Fuerst and Fuerst, 1993).

Since it was launched, the needs of children and their families changed substantially and during the 1990s, Head Start shifted from a program where most children

⁶These two programs also differ in their intensity and age at which children start the intervention. The Abecedarian was a full-day program that started with children in the first months of life and the Perry operated half-day with 4 years old children.

were enrolled in part-day centers towards a full day program (by 2003 47% of those enrolled were served by full-day programs, that is, 6 or more hours/day; see GAO, 2003).

The criteria governing the selection of children into Head Start are advertised regularly by the Head Start Office (see the Head Start's Office web site). Children are eligible to participate if they are of preschool or kindergarten age and if they live in poverty. In addition, at least 10% of the children served per center must have some type of disability. Since these selection criteria are explored in our identification strategy, we defer the explanation of details to Section 3. Eligibility criteria have been mostly unchanged since the 1971, covering the entire period we analyze.⁷

There was a slow increase in enrolment in Head Start during the 1970s and 1980s, and a sharp increase in the early 1990s. Between 1991 and 1995 the enrolment increased by about 25% (from almost 600,000 to 750,000 children). Simultaneously, there was an increase in the funding per child: in the early 1990s the federal cost per child was about \$US5,300 per year (in \$US2009), whereas in the FY of 2009, Head Start operated 49,200 classrooms serving almost 1 million children at federal cost of \$US7,800/child. These numbers show that the effort to expand and improve the program means that today its costs per child reach about 85% of the cost of Perry Preschool.⁸

3 Empirical Strategy

Naive comparisons between the outcomes of those who have and have not participated in Head Start confound program impacts with differences in the underlying characteristics of participants and non-participants. This problem has been addressed in a variety of ways in recent papers, as mentioned above. In this paper we explore exogenous variation in participation in Head Start driven by program eligibility rules.

⁷See Table B.1 in the Appendix, which shows the main pieces of relevant legislation.

⁸According to Heckman et al. (2010) the estimates of initial costs of Perry Preschool (presented in Barnett, 1996), reached \$17,759/child over its two years (in 2006 \$US). This figure includes both operating costs (teacher salaries and administrative costs) and capital costs (classrooms and facilities).

Children ages 3 to 5 are eligible if their family income is below the federal poverty guidelines, or if their family is eligible for public assistance: AFDC (TANF, after 1996) and SSI (DHHS, 2011).⁹ Once a family becomes eligible in one program-year, it is also considered eligible for the subsequent program-year. Since program eligibility is a discontinuous function of family income, program participation is likely to be discontinuous in family income as well. Therefore, we can study the impact of participating in Head Start on a variety of outcomes using a regression discontinuity estimator.¹⁰

We start by estimating the following reduced form model:

$$Y_i = \phi + \gamma E_i + f(Z_i, X_i) + u_i \quad (1)$$

where E_i is an indicator of eligibility for Head Start, X_i is a set of all determinants of eligibility for each child except for family income (year, state, family size, family structure, measured at age 4), Z_i is family income (at age 4), and u_i is the unobservable.¹¹ We include state effects in our models not only because the criteria for eligibility are state-dependent but also to account for cross-state permanent heterogeneity that is associated with differences in generosity and services provided. The equation for E_i is:

$$E_i = 1 [Z_i \leq \bar{Z}(X_i)], \quad (2)$$

where $1[\cdot]$ denotes the indicator function.

$f(Z_i, X_i)$ is specified as a parametric but flexible function of its arguments, and $\bar{Z}(X_i)$ is a deterministic (and known) function that returns the income eligibility cutoff for a family with characteristics X_i (constructed from the eligibility rules). In modeling $f(Z_i, X_i)$ we rely on series estimation (widely used in other applications of this empirical strategy), restricting the sample to values of the forcing variable that are close to the cutoff points. In section 5 we study the sensitivity of our results

⁹AFDC denotes Aid to Families with Dependent Children, TANF denotes Temporary Assistance for Needy Families, and SSI denotes Supplemental Security Income.

¹⁰Eligibility criteria and the construction of eligibility status for each child are discussed in section 3.1.

¹¹ $f(Z_i, X_i)$ can be a different function in each side of the discontinuity. We empirically examine this case, but the estimates are too imprecise to be conclusive and therefore we did not include them in the paper.

to the choice of different functional forms for $f(Z_i, X_i)$. We use probit models whenever the outcome of interest is binary.

Three conditions need to hold for γ to be informative about the effects of Head Start on children outcomes. First, after controlling flexibly for all the determinants of eligibility, E_i must predict participation in the program, which we show to be true.

Second, families are not able to manipulate household income around the eligibility cutoff. This is the main assumption behind any regression discontinuity design. It is likely to hold in our case because the formulas for determining eligibility cutoffs are complex, and depend on family size, family structure, state and year, making it difficult for a family to position itself just above or just below the cutoff. In addition, there are standard ways to test for violations of this assumption (e.g., Imbens and Lemieux, 2008), and below we discuss them in detail.

Third, eligibility to Head Start should not be correlated with eligibility to other programs that also affect child outcomes. This assumption is potentially more likely to be violated than the first two, because there are other means tested programs which have eligibility criteria similar to those of Head Start (e.g., AFDC, SSI, or Food Stamps). The fact that the definition of income we use (see Appendix F) is specific to Head Start guarantees that those eligible through the Federal Poverty Guidelines have a cutoff not shared with other welfare programs, and this accounts for most of the children. Nevertheless, in order to assess the potential importance of this problem we implement the following procedure. While most welfare programs exist throughout the child's life, Head Start only exists when the child is between the ages of 3 and 5. If other programs affect outcomes of children, then eligibility to those programs in ages other than 3 to 5 should also affect children's outcomes. In contrast, if eligibility is correlated with children's outcomes only when measured between ages 3 and 5, then it probably reflects the effect of Head Start alone. Although we cannot definitely rule out the possibility that other programs confound the effects of Head Start (by operating exactly at the same ages), the results we present below are highly suggestive that this is not the case.

3.1 Eligibility Criteria

We construct each child's income eligibility status in the following way (a detailed description can be found in the Appendix F). First, the poverty status of each family is imputed by comparing family income with the relevant federal poverty line, which varies with family size and year. Second, eligibility for AFDC/TANF requires satisfying two income tests, and additional categorical requirements, all of which are state specific. In particular, the *gross income test* requires that total family income must be below a multiple of the state specific threshold, that is set annually and by family size at the state level. The second income test to be verified by applicants (but not by current recipients) is the *countable income test*, that requires total family income minus some disregards to be below the state threshold for eligibility (U.S. Congress, 1994). In addition, AFDC families must obey a particular structure: either they are female-headed families, or they are families where the main earner is unemployed. This means that children in two-parents households may be eligible for AFDC under the AFDC-Unemployed Parent program. In turn, eligibility for AFDC-UP is limited to those families in which the principal wage earner is unemployed but has a past work history, so we consider eligible those whose father (or step-father) worked on average less than 100 hours per month in the previous calendar year.¹²

We use total family income from the last calendar year available in the NLSY79, which relatively to the income measure used by the Head Start Office includes also Food Stamps. We also assess the sensitivity of our results to the measure of income used and we re-estimate the first stage and reduced form models using alternative definitions of income.¹³

A child can enrol in Head Start at ages 3, 4, or 5 and it is possible to construct eligibility status at each of these ages. As we show in Section 5, eligibility at age 4

¹²We do not impute SSI eligibility for two reasons. First, imputing SSI eligibility would require the imputation of categorical requirements which are complex to determine (e.g., Daly and Burkhauser, 2002), some of which we are unable to observe in the data (for example, in that NLSY there is no information on whether the health limitations of the parent that may be eligible fulfill the medical listings that determine eligibility). Second, SSI thresholds are below Poverty Guidelines and therefore these thresholds will not be binding (see U.S. Congress, 2004).

¹³See Appendix A.2 for the results for alternative income definitions.

is a better predictor of program participation than either eligibility at 3 or at 5, and in our data (and in the administrative records from the Head Start Office) 50-60% children enrol in Head Start when they are 4 (Head Start Office, 2011). Therefore, we focus on eligibility at age 4 in our main specification, but we also present results with eligibility at other ages.

When using regression discontinuity it is only possible to identify program impacts in the neighborhood of the cutoff. Since we explore multiple discontinuities, we can also learn about the range of neighborhoods of income over which we can identify program impacts. For this we can plot the distribution of cutoff values (for household income) to visualize the support of income values for which we are able to identify the effects of Head Start.¹⁴ About 98% of the children in our sample have eligibility determined by the federal poverty line criterion (with the remaining qualifying through AFDC/TANF eligibility).

3.2 Imperfect Compliance

It is important to note that eligibility rules for Head Start are not perfectly enforced (some ineligible children are able to enrol), and that take up rates among those eligible are far below 100%. There are several factors that influence the take up of social programs, such as shortage of funding to serve all eligible, barriers to enrollment, and social stigma associated with participation (e.g., Currie, 2006, Moffitt, 1983). Due to limited funding, Head Start enrolls less than 60% of all children in poverty who are between the ages of 3 and 4. Priority is given to the neediest among the poor.¹⁵

¹⁴Figure C.1 in Appendix displays the distribution of cutoff values (for household income) and this corresponds to the support of income values for which we are able to identify the effects of Head Start. Income cutoffs also vary across different family sizes, and in Figure G.1 in the Appendix G we plot the joint support of household income and family size over which we are able to estimate the relevant program effect.

¹⁵The problem of imperfect compliance is not unique to Head Start, but common across social programs. Only 2/3 of eligible single mothers used AFDC (Blank and Ruggles, 1996); 69% of eligible households for the Food Stamps program participated in 1994 (Currie, 2006); of the 31% of children eligible for Medicaid in 1996, only 22.6% were enrolled (Gruber, 2003); EITC has an exceptionally high take-up rate of over 80% among eligible taxpayers (Scholz, 1994); in 1998, participation in WIC (the Special Supplemental Nutrition Program for Women, Infants and Children) among those eligible was 73% for infants, 2/3 among pregnant women and 38% for children (Bitler,

The number of eligible individuals is also different from the number of actual participants because of lack of perfect enforcement of eligibility rules and of other factors affecting participation. In the addition, Head Start centers may enrol up to 10% of children from families whose income is above the threshold (without any cap on the income of these families). Thus, the discontinuity in the probability of take-up of Head Start around the income eligibility threshold is not sharp, but "fuzzy" (see Hahn, Todd and van der Klauww, 2001, Battistin and Rettore, 2008, and Imbens and Lemieux, 2008). As a result, γ in equation (1) does not correspond to the impact of Head Start on the outcome of interest. In order to determine the program impact, we estimate the following system, for the case where Y_i is continuous:

$$Y_i = \alpha + \beta HS_i + g(Z_i, X_i) + \varepsilon_i \quad (3)$$

$$HS_i = 1[\eta + \tau E_i + h(Z_i, X_i) + v_i > 0], \quad (4)$$

where equation (4) is estimated using a probit model (van der Klauww, 2002). $1[\cdot]$ denotes the indicator function. $P_i = \Pr(HS_i = 1|E_i, Z_i, X_i)$ is estimated in a first stage regression, and used to instrument for HS_i in a second stage instrumental variable regression (Hahn, Todd and van der Klauww, 2001). If Y_i is binary we use a bivariate probit. $g(\cdot)$ and $h(\cdot)$ are flexible functions of (Z_i, X_i) .¹⁶

4 Data

We use data from the Children of the National Longitudinal Survey of Youth of 1979 (CNLSY), which is a survey derived from the National Longitudinal Survey of Youth (NLSY79). The NLSY79 is a panel of individuals whose age was between 14 and 21 by December 31, 1978 (approximately half are women). The survey has been carried out since 1979 and we use data up to 2008 (interviews were annual up to 1994, and have been carried out every two years after that). The CNLSY is

Currie and Scholz, 2003).

¹⁶In Appendix G we also discuss how we can identify heterogeneous effects of Head Start. Unfortunately, even though our estimates of heterogeneous effects are interesting they are also imprecise.

a biennial survey which began in 1986 and contains information about cognitive, social and behavioral development of the children of the women surveyed in the NSLY79 (assembled through a battery of age specific instruments), from birth to early adulthood.

Children 3 to 5 years old are eligible to participate in the program if their family income is below an income threshold, which varies with household characteristics, state of residence, and year. Among the variables available in CNLSY there are those that determine income eligibility (total family income, family size, state of residence, Head Start cohort and an indicator of the presence of a father-figure in the child's household) along with outcomes at different ages. For reasons explained in Section 3, we will focus mainly in the outcomes of children potentially eligible for the program at age 4. In our data, the earliest year in which we can construct eligibility at age four is 1979 (for children born in 1975), since this is the first year in which income is measured in the survey (eligibility each year is determined by last year's income, which is precisely what is asked in the survey). Our final sample consists of children born between 1977 and 1996 (after imposing additional restrictions). Therefore, we study the effects of participating in Head Start throughout the 1980s and 1990s.¹⁷

Our treatment variable is an indicator for Head Start participation between ages 3 and 5. This is based on information collected by the CLNSY from 1988 onwards on whether the child currently attends nursery school or a preschool program, or whether she has ever been enrolled in preschool, day care, or Head Start. For participants we use the age at which the child first attended Head Start and the length of time attending the program to construct an indicator of Head Start attendance between ages 3 to 5. We use the variable "Ever enrolled in preschool?" to construct participation in preschool. Therefore, we distinguish three alternative child care arrangements between ages 3 to 5: Head Start, enrolment in other preschool or neither of the previous two (informal care at home or elsewhere). In the raw data, 90% of mothers who report that their child was enrolled in Head Start also report that their child was enrolled in preschool, possibly confounding the two child care

¹⁷All monetary values presented are here are through out the paper measured in 2000 values using the CPI-U from the Economic Report of the President (2012), unless mentioned otherwise.

arrangements. Therefore, as in Currie and Thomas (1995, 1999), we recode the preschool variable so that whenever a mother reports both Head Start and preschool participation, we assume that there was enrolment in Head Start alone. After recoding this variable, almost 20% of the children in the sample enrolled in Head Start, 40% attended other type of preschool, and the remaining attended neither. In our data, about 40% of participants enter Head Start at age 3, and 50% enter at age 4. In the CNLSY, 90% of Head Start participants attended the program for at most one year.¹⁸

Out of the 11,495 children surveyed by 2008, we drop 2285 children for whom we do not observe Head Start participation between ages 3 to 5.¹⁹ We further drop 1974 children for whom we are unable to assess income-eligibility status at age 4 because of lack of information on family income, family size, state of residence or mother's co-habitational status. We drop 855 children without information on income and family size before age 3 and birth weight. These variables are used as controls and we show in Section 5 that our results are not sensitive to the exclusion of these pre-determined control variables. We then exclude 948 children who are not observed at least once at ages 12-13, 16-17 and 20-21 and with missing information on the outcomes we analyze. Thus, the final sample consists of 5433 children. Although we discuss some results for females, for reasons that become clearer in Section 5, the bulk of the paper focuses on males.

¹⁸A back-of-envelope calculation, suggests that based on official numbers we would expect the Head Start participation rate to be around 5% in the 1980s and early 1990s, but 8% in 2000. This is because according to the US CENSUS about 20% of children ages 3 to 5 in the US are poor, which amounts to 1,663,440 (out of 9,207,040), 2,021,299 (out of 10,275,120) and 1,836,383 (out of 10,601,578) children for the years of 1980, 1990 and 2000, respectively (poverty in CENSUS is defined using poverty thresholds, whereas eligibility to Head Start is determined by the poverty guidelines), and for these years the number of children enrolled in the program is 376,300, 540,930 and 857,664. We have a larger estimate in our data, possibly because of two characteristics to the sampling of NLSY: (1) about 50% of our sample is an oversample of minorities and poor whites available in data and (2) the CNLSY contains an overestimate of children from young mothers. This explains why our number is comparable to the 19.4% Figure (in Currie and Thomas, 1995, who use the same data source). Currie, Garces and Thomas, 2002, estimate Head Start participation at 10% in the PSID, and Ludwig and Miller, 2007, have participation rates of 20 to 40% in the counties close to their relevant discontinuity (based on data from the National Educational Longitudinal Study). As a further note, the NLSY79 also includes a subsample of members of the military, which we exclude from our work.

¹⁹Table D.1 in Appendix includes the details of construction of our sample.

Since we rely on a discontinuity in the probability of participation around a threshold, we restrict the sample to children whose family income at eligible age was near the income eligibility cutoff for the program since points away from the discontinuity should have no weight in the estimation of program impacts (see e.g., Black, Galdo, and Smith, 2005, Lee and Lemieux, 2010). Therefore, we focus on the sample of children whose income was between 15% and 185% of the relevant income cutoff (we also present estimates using alternative intervals for income). Within this window of data we observe 2833 children (2550 at ages 12-13, 2416 at ages 16-17, 1977 at 20-21 years old; of these 1595 children are present at all age groups).²⁰

MAYBE THIS TABLE SHOULD COME TO THE MAIN TEXT Table B.2 in Appendix summarizes the data and it includes covariates that relate to family and child characteristics. It shows means, standard deviations and the number of available observations for each variable. It is clear that the children in our sample come from fairly disadvantaged backgrounds: 38% of their mothers are high school dropouts, and only 10% ever enrolled in college (although not presented in table, these figures are 26% and 22%, respectively, when we use all children in the CNLSY). Their average annual family income is only slightly above \$18000 (deflated to 2000; as opposed to \$42443 for the whole sample), 10% of children are reported to have been of low birth weight, 31% of these children were enrolled in Head Start, 52% were in other types of preschool, and 17% were in neither. In the whole sample of children 8% of children are reported to have been of low birth

²⁰One potential problem of our approach is the large fraction of individuals which are dropped from the sample due to missing information for the assessment of the eligibility status at age four. Missing information could be a problem for our identification strategy if there are different response rates on either side of the cutoff. In practice, if income is missing we cannot check if a given individual is on one side or the other side of the discontinuity at the age in which we measure eligibility. However, we can do something very close to this, by looking at variables measured at other ages. In particular, in table D.2 in Appendix we re-estimate the reduced form model of equation 1 using as dependent variable a dummy for whether the child has missing information on any of these pre-age 3 controls that we add to the specification. In other words, we check if there is any difference in non-response on pre-age 3 controls on either side of the discontinuity. We cannot reject the null that the coefficient on eligibility in this regression is equal to zero, i.e., we cannot reject the null that non-response on pre-age 3 controls is the same in both sides of the discontinuity. Although this is just suggestive that selective non-response is not a major problem, it is reassuring. We thank this point to an anonymous referee.

weight, 20% of these children were enrolled in Head Start, 62% were in other types of preschool, and the remainder did not attend any of these. A detailed description of all outcome variables used in our analysis is included table B.3 in Appendix and their mean and standard deviation are presented in table B.4 also in Appendix B.

5 Results

5.1 First Stage Estimates

We start by checking whether the discontinuity in eligibility status also induces a discontinuity in the probability of Head Start participation by estimating equation (4). We present estimates for the three main samples we analyze (children ages 12-13, adolescents 16-17 and young adults 20-21) and by gender. Table 1 presents estimates of τ in equation (4). Eligibility is measured at age four, the age at which most children first enrol in Head Start. The marginal effect included is the average marginal change in participation as a result of a change in the eligibility status.²¹ Function $h(Z_i, X_i)$ consists of a cubic in log family income and family size at age 4, an interaction between these two variables, a dummy indicating the presence of a father figure (father or step-father) in the household at age 4, indicators for gender, race and age, and indicators for year and state of residence at age 4. All standard errors in the paper are clustered at the level of the state-year, since eligibility rules are determined at this level (in the Appendix we also present estimates where clustering is done only at the state level).

It is clear from Table 1, that across age groups, eligibility at age four is a strong predictor of program participation for males, although the estimated effect is well

²¹This is defined by:

$$\frac{1}{N} \sum_{i=1}^N \{\Pr(HS_i = 1 | E_i = 1, Z_i, X_i) - \Pr(HS_i = 1 | E_i = 0, Z_i, X_i)\} =$$

$$\frac{1}{N} \sum_{i=1}^N [\Phi(\eta + \tau + h(Z_i, X_i)) - \Phi(\eta + h(Z_i, X_i))]$$

where N is the number of children in the regression sample, and Φ is the standard normal distribution function (we obtain similar results if we take the average of marginal effects using observations only in a small neighborhood of each cutoff).

below 100%. This is an indication of weak take-up of the program at the margin of eligibility (common to many social programs).²² Our paper is novel in obtaining estimates of how the take-up of Head Start changes for individuals near the eligibility threshold, as their eligibility status change. This can be interpreted as the increase in participation generated by a small change in eligibility thresholds.²³

We choose to focus on eligibility at age four as the main determinant of participation in Head Start because eligibility at age four is a better predictor of participation than either eligibility at age 3 or eligibility at age 5 (see table B.5 in Appendix). Therefore, the population of children for whom we are able to estimate the impact of Head Start are those for whom small changes in eligibility criteria induce them to enrol in Head Start. We are not able to estimate the impact of Head Start on those who are permanently and substantially below the poverty line, since they are unlikely to be located close to the eligibility cutoffs.

When using a RD setup it is standard practice to present a graphical analysis of the problem. Relative to the standard setting which has a single discontinuity, our setup makes use of a range of discontinuities. One graphical representation of the problem which does not correspond exactly to the specification of our model takes a measure of family income relative to each family's income eligibility cutoff, and defines this variable as "distance to the eligibility cutoff". Figure 1 plots Head Start participation at age 4 for males and females entering our analysis of outcomes at ages 12-13, 16-17 and 20-21, against the relative distance of family income to the income eligibility cutoff (at age 4). We divide the sample into bins of this variable (of size 0.05) and compute cell means for participation. We then run local linear regressions of each variable on the distance to cutoff on either side of the

²²This was discussed briefly above, but low take-up could be partially driven by the fact some children start the program at either ages three or five when they are also eligible, but it is more likely that it results from several other factors, such as the lack of available funds to cover all eligible children (Head Start was never fully funded), stigma associated with program participation (Moffitt, 1983), or the fact that throughout most of the period we study (1980s and 1990s) most of the centers offered only part-day programs, which do not satisfy the needs of working families for full-day programs (Currie, 2006).

²³Most of the evidence of how newly eligible to social programs respond in terms of participation comes from Medicaid expansions throughout the 1980s and early 1990s, namely Cutler and Gruber (1996), Currie and Gruber (1996), Card and Shore-Sheppard (2002) and Lo Sasso and Buchmueller (2002).

discontinuity (we use a bandwidth = 0.3, but the results are unchanged if we use bandwidths of 0.2 or 0.4). These figures suggest discontinuities of about 15% in program participation at the eligibility cutoff for the sample of boys, but no jump in the probability of participation for females. This is exactly what our regressions show in Table 1.

5.1.1 Gender Differences in first stage estimates

It is interesting and surprising that changes in eligibility status are not associated with changes in participation in Head Start for females. This result holds across races, as reported in table B.6 in Appendix B. It is difficult to understand why there is such a gender discrepancy. The fact that the change in eligibility status is only associated with changes in participation for boys and not for girls suggests that the marginal entrant into Head Start is a boy. It also implies that, using this strategy, we cannot estimate the impacts of Head Start for girls.²⁴

This gender difference in program take-up is also present in the Head Start Impact Study, which randomizes eligibility across children wait-listed for a few oversubscribed Head Start centers. Using the data from this study, when we regress an HS enrollment dummy on a dummy indicating whether the child had won the lottery to access HS (with no controls), we estimate that winning the lottery leads to a 72% increase in the probability of enrolment for (4 year old) boys, but only a 63% increase for girls. Although this variation in eligibility is different than the one used in our paper (see table B.7 in Appendix B), these results still show that differential gender responses to eligibility status are not exclusive to our paper.

In an attempt to understand the reasons behind the gender difference in the response to changes in eligibility, we start by dividing the sample by race, mother's cohabitation status when child was 4 and area of residence at age 4, and redoing our analysis for each group.²⁵ Our results show that gender differences in responses to

²⁴Gender is not the only demographic on which we find differences in the impact of program eligibility on program enrolment. In the appendix we also report that the discontinuity in the probability of participation is larger for Black boys than for non-Blacks, so the marginal entrant is more likely to be Black (see table B.6).

²⁵The results by race are included in table B.6. For brevity we do not include the first stage estimates by mother's cohabitation status and area of residence when child was 4, but these are

eligibility are not driven by any particular demographic group. Furthermore, these differences are also not driven by an earlier or later enrolment of girls (relative to boys) in Head Start (at either ages 3 and 5, the jump in the probability of participation is only statistically significant for boys - see Table B.5 in the Appendix).

Second, we studied whether there exist gender differences in other parental investments, and labor market outcomes of parents. Simple plots of the density for a measure of quality of the home environment (the HOME score) show that, between the ages of 0 and 10, HOME investments tend to be higher for girls than for boys.²⁶

We also regressed measures of maternal labor supply and the HOME score (and two of its subscores) on child's gender, age and year indicators. We do not find any gender gap in the labor market outcomes of mothers (in particular, on the number of weeks worked per year). However, maternal investments measured by the HOME score are on average 0.1SD lower for boys than girls (the same holds for its subscores). Even when we focus on families with multiple children of different gender the HOME score is 0.09SD lower on average among boys relatively to girls. Note that the HOME score is constructed from mother's answers to variety of questions including mostly information about maternal attitudes towards children. Our findings that mothers engage more with girls is consistent with others in literature (see Lundberg, 2005, for a review).²⁷

In sum, gender differences in how the take-up of Head Start responds to changes in eligibility are not exclusive to our dataset. They are also present in the data from the Head Start Impact Study. Although the magnitudes of the gender differences are not the same in the two datasets (nor is the nature of the changes in eligibil-

available from the authors.

²⁶See figure C.3 in Appendix. For each age we perform the Kolmogorov-Smirnov test for the equality of the distributions for the score of the two genders. The p-values are the following: 0.339 (age 0), 0.067 (age 2), 0.000 (age 4), 0.157 (age 6), 0.017 (age 8) and 0.002 (age 10). The differences in these densities for most of the ages presented here, and the same is true for those ages not shown. The HOME score is available from ages 0 to 14 and it aims to measure the quality of the cognitive stimulation and emotional support provided by a child's family. More than half of the HOME items are obtained from maternal reports.

²⁷The results for the estimated gender gap are in Panels A and B of table B.8 in the Appendix.

Following Dahl and Moretti, 2008, we also tried to look for evidence of whether parents in the NLSY continue childbearing until they have a son. We do not find robust evidence of this in the CNLSY sample, which could be due to the fact that this is a much smaller sample than the CENSUS, and thus we may lack power to test such hypothesis.

ity), the qualitative patterns are similar: enrolment rates of males respond more to changes in eligibility than enrolment rates of girls. In addition, gender differences in parental investments have already been documented in the literature for developing and developed countries (e.g., Dahl and Moretti, 2008, Lundberg, 2005, Baker and Milligan, 2013). The gender differences in program take-up we observe may be one more manifestation of the same phenomenon. These differences could be driven by differences in technology or differences in preferences (or even in expectations), but as Lundberg (2005) points out, it is very difficult to distinguish different explanations.

5.1.2 Understanding the comparison group

In order to be able to interpret our results it is central to understand in which type of child care would children enrol in the absence of the program. As we explained in Section 4, we consider three possible child care arrangements between ages 3 and 5: "Head Start", "Other Preschool", "Informal care". Table 2 shows how participation in these three alternative child care arrangements responds to eligibility. We regress the dummy variables indicating participation in each type of child care on eligibility and the remaining control variables. There are two panels in the table, corresponding to males and females. Each panel is divided by age group: those for whom we have outcomes at ages 12-13 (columns 1-3), those with outcomes at 16-17, and those with outcomes at ages 20-21 (columns 7-9).

We start by discussing Panel A for boys. Columns 2 and 3 show that, for the youngest cohort, when an individual becomes Head Start eligible there is a statistically significant movement out of "Other Preschool". In contrast, columns 4-6 show instead that children in slightly older cohorts are more likely to leave "Informal Care" when they become eligible for Head Start. Finally, for the oldest cohort of children (columns 7-9), there is movement out of both "Other Preschool" and "Informal Care" in response to a change in eligibility status, but movement out of the "Informal Care" seems to be relatively more important. Changes in a child's eligibility status are not associated with changes in participation in any of the three types of care among girls.²⁸

²⁸The estimates for the marginal change in the take-up of the three child care alternatives do not

It is useful to contrast our control group with those used in previous studies, since differences in the population of interest across studies may lead to differences in results. Currie and Thomas (1995), Currie, Garces and Thomas (2002), and Deming (2009) compare siblings that attended Head Start vs. either "Other Preschool" or "Other type of care". In contrast, the HSIS, 2010, compares Head Start children with children in the waiting lists of about 380 centers, who attended a mixture of alternative care settings (around 60% of children in the control group participated in some type of child care or early education programs during the first year of the study, with 13.8% and 17.8% of the 4 and 3-year-old in the control group, respectively, participating in Head Start itself). Since we use the same data set as Currie and Thomas, 1995 (and Deming, 2009), we can further compare the characteristics of the individuals induced to enrol in Head Start by eligibility at age four (the relevant population in our study), and the characteristics of children in families where one sibling enrolls in Head Start and the other does not, which are the relevant populations in earlier papers on this topic. For most of the measures we analyze, the group of children for whom we identify impacts of Head Start is less disadvantaged than the population of children studied in sibling studies. Both groups of children are more likely to live in poor families before age 3 than the average, but when we compare the relevant population in siblings studies with the relevant population in our study, the former is more likely to be poor, to have less educated mothers, to have mothers with lower levels of AFQT, and to not have been breastfed.²⁹

5.2 Balancing Checks

In this section we perform standard balancing checks, examining whether individuals just above and just below the eligibility thresholds look similar in terms of their observable characteristics. We take a set of pre-program variables that should not be affected by participation in the program, and we use them as dependent variables in equation (1). If our procedure is valid then the estimate of γ in these regressions

change if a multinomial logit model is estimated instead of separate probit models for each choice.

²⁹These results are included in Table B.9 in Appendix B, which presents the relative likelihood of compliers having a given characteristic compared to the population at large; see Angrist and Pischke, 2009.

should be equal to zero.

The relevant variables are: the child's average MOTOR score before she turned three (a measure of the physical and social development for very young children), birthweight, mother's education, maternal grandmother's education, marital status of the mother before the child turned 3, mother's AFQT score, average log family income and family size between the ages of 0 and 2, and several variables related to the mother's family environment when she was 14 years old (whether the mother lived in a Southern state, whether she lived with her parents, how many siblings she had, and whether she lived in a rural area). Eligibility is measured at age 4, as explained above.

The results are presented in Table 3, and the sample includes all children for whom we observe outcomes at ages 12-13. Results for other older age groups are similar, and are available from the authors. Most estimates of γ are small (compared with the mean and standard deviation of each variable also included in table) and, when taken individually, almost all of them are statistically insignificant.³⁰

Furthermore, because we are testing multiple hypotheses simultaneously we should adjust the relevant p-values. Once we do that, following the procedure suggested in Algorithms 4.1 and 4.2 of Romano and Wolf (2005), we cannot reject the hypothesis that there is no significant relationship between any of these variables and eligibility (even in the case of the two variables for which coefficients are individually statistically different from zero: birth weight, mother married before child turned 3 and whether mother lived in a rural area by age 14).³¹ Plots of local linear

³⁰In order to better understand the magnitude of these estimates we conducted the following exercise. Take a few of our main outcomes of interest, such as BPI at ages 12-13, and CESD by ages 16-17. Then regress each outcome on each of the variables in table (3), and compute predicted values for each regression. We can now rerun the regressions on table (3) using these predicted values instead of the variables that generated them, allowing us to translate the coefficients in table (3) into magnitudes of the outcomes of interest. We do not report this in a table, but describe the results briefly in the text (for all boys): in terms of BPI, all the coefficients in table (3) are between -0.0081 and 0.016 (expressed as a fraction of a standard deviation), and for CESD up to ages 16 to 17 they are between -0.0068 and 0.007 (expressed as a fraction of a standard deviation). All these figures are very small.

³¹Since we are examining the impact of a program on multiple variables (as opposed to a single variable) we need to account for that when doing hypothesis testing. Several multiple hypotheses testing procedures exist, but the most recent one is developed in Romano and Wolf (2005), which accounts for non-independence across outcomes, and has more power than most of its predecessors

regression estimates of these variables measured before the child turned three years old on the distance to the cutoff (by eligibility status) are continuous around the threshold.³²

Throughout the rest of the paper we augment our basic specification of $f(Z_i, X_i)$ with some of these variables as additional covariates to reduce sampling error and small sample bias (e.g., Lee and Lemieux, 2010). In particular, we add a cubic on log of average family income and average family size between ages 0 and 2, an interaction between the two, and a cubic on the child's birth weight.

5.3 Estimates from the Reduced Form Equations

5.3.1 Indices of Outcomes

Table 4 presents estimates for γ in equation (1), where the dependent variables are summary indices of the set of outcomes we analyze at each age group studied (12-13, 16-17 and 20-21). In order to construct these indices we first standardize each individual outcome variable, and then we average them, using weights that ensure that outcomes which are highly correlated with each other receive less weight whereas outcomes that are uncorrelated and thus represent new information receive more weight. In particular, the weight is the inverse of the variance covariance matrix (see Anderson, 2008). Each index is then re-standardized to have mean zero and standard deviation one for a clearer interpretation of results.

We use 15 variables to construct the summary index at ages 12-13, 8 variables for the index at ages 16-17, and 6 variables for the index at ages 20-21. For children ages 12-13, the overall index can be divided into three subindexes: cognition, which includes mainly test scores; behaviors, which includes measures of behavioral problems; and health, which includes a variety of health indicators. Given the small number of variables used at ages 16-17 and 20-21, which mostly refer

(namely Westfall and Young, 1993).

³²These results are presented in figure C.4 in the Appendix.

An alternative and more direct test of manipulation, developed by McCrary (2007), checks whether there is bunching of individuals just before the discontinuity. This test is not practical with multiple discontinuities unless we have a large sample size. However, when we implement it using a single discontinuity (using percentage distance to the eligibility cutoff as the running variable) we find no evidence of income manipulation, as shown in figure C.5 in Appendix.

to behavioral measures, we opted to construct one single index for these samples. The variables composing these indices have their sign switched when needed, so that positive direction always indicates a "better" outcome. Therefore, a positive coefficient on eligibility is interpreted as a positive effect of the program. Table B.3 in Appendix lists the variables used in the indices.

Table 4 is divided into three panels, one for each age group. Since there are no impacts of eligibility on Head Start participation for girls all the relevant coefficients should be zero for this sample. We report results for this sample as a check to our procedure. Estimates in Panel A show that, among boys 12-13 years old, eligibility to Head Start leads to an overall improvement in the summary index. Panel B shows that for boys 16-17 being eligible for Head Start also leads to better outcomes for boys (column 4), which are reflected into the overall sample (columns 6). We do not detect a statistically significant relationship between the index at ages 20-21 and eligibility to Head Start (see Panel C), although the estimated coefficient for boys is positive and large (roughly of the same magnitude as the estimate in panel B, but the sample at ages 20-21 is about 3/4 of that for adolescents). Furthermore, for this age group the analysis of individual outcomes below shows statistically significant program impacts on a few outcomes, even after accounting for multiple hypothesis testing.³³

5.3.2 Individual Outcomes

In tables 5-7 we present the effects for the individual components of the index, only for boys (estimates for girls and for the whole sample are displayed in Appendix tables B.11-B.13).

For each outcome we report (i) the mean ("control mean") of the outcome for

³³In table 4 the coefficient on eligibility for girls ages 12-13 is weakly significant, but (1) none of its subcomponents (cognitive, health or behaviors) is statistically associated to eligibility in table B.10 in Appendix and (2) none of the estimates for its individual components in table B.11 in Appendix is significant when we adjust the p-values to account for multiple hypotheses testing. Additionally, in table B.10 for males there is no association between the index of cognition and eligibility to Head Start (columns 1), but there is a positive relation between eligibility to Head Start and the behavioral and health indexes for boys (column 4 and 7, respectively) and for the whole sample (column 6 and 9). Among boys 12-13 years old Head Start is associated with an improvement of 23.3% and 23.6% of a SD in behavioral and health problems, respectively.

those individuals just above the cutoff, (ii) the number of observations in each regression, (iii) the coefficient on eligibility (column labeled "ITT", or intention to treat) and its standard error³⁴, and (iv) whether the hypothesis that the coefficient is equal to zero is rejected at the 10% level of significance using the algorithm of Romano and Wolf (column labeled "RW $p_v < 0.1$ "), which accounts for multiple hypothesis testing. Throughout our discussion we consider that the program has a statistically significant effect on a specific outcome only in the cases where we can reject the null that the effect is zero using this procedure.

Table 5 looks at outcomes at ages 12-13. We find that for boys Head Start eligibility leads to a reduction in the probability of being overweight, on the probability of having a health condition that requires the use of special equipment (such as a brace, crutches, a wheelchair, special shoes, a helmet, a special bed, a breathing mask, an air filter, or a catheter) and a reduction in behavior problems as measured by the BPI.³⁵

Our results for cognitive tests are imprecise, but overall we do not find evidence of impacts of Head Start participation on any of the tests we study. This is not consistent with the findings of Currie and Thomas (1995) and Deming (2009), but it is consistent with the findings of HSIS. Note that the children in our analysis are a mixture of the older cohorts studied in Currie and Thomas (1995) and Deming (2009) and younger cohorts closer to those in the HSIS, so our results could be close to either of these sets of studies. In addition, as shown above we study a less disadvantaged group than Currie and Thomas (1995) and Deming (2009) and our focus is not on the widely analyzed PPVT because this test is administered fairly infrequently, when compared to the other tests we study (nevertheless, our results are essentially the same when we analyze the PPVT).³⁶

³⁴For the discrete outcomes we also present the average marginal effect of eligibility on the outcome being analyzed in each row in *italic*.

³⁵We also analyzed the frequency of dental check-ups. This is an important outcome as one of the services provided to Head Start children and it is one outcome where the Head Start Impact Study, 2010, found effect sustained until the end of kindergarten. We did not find any effects on whether the child has had any dental check either the last 12 or 24 months at ages 6-7, 9-10 or 12-13. We do not report these outcomes in our main tables as information on dental check-ups is only available since 1992, and the sample size in estimations is about 75% of that used for the other outcomes for these age groups.

³⁶For comparison, in the appendix we also present estimates of the impacts of Head Start partic-

Table 6 shows estimates of the impact of eligibility to Head Start on outcomes for adolescents ages 16-17. We find that eligibility to Head Start leads to a decrease in the probability of being overweight and a reduction in symptoms of depression, measured by the CESD.

Finally, table 7 includes estimates for young adults ages 20-21 years. We show that HS eligibility leads to a decrease in the probability of ever being sentenced for a crime and idleness by ages 20-21 among males. These impacts are statistically significant even after accounting for multiple hypothesis testing, even though we could not detect an impact of eligibility to HS on the summary index used in table 4.³⁷

The different panels in figure 2 display the graphical representation of our replication versus pre-school and other arrangements using a siblings comparison strategy, as in Currie and Thomas (1995) and Deming (2009). These results are included in table B.14 in Appendix. Because we focus on a different cohort of participants than these papers, we present two columns for each outcome: (1) for children that could have mainly attended the program in the 1980s, born up to 1986 (as Currie and Thomas, 1995, and Deming, 2009) and (2) for children that could have enrolled in the 1990s (born after 1986). We present estimates for four outcomes: (1) an index created following Deming (2009) which is the average of PIAT-Math, PIAT-Reading Recognition and PPVT, (2) for PPVT, (3) for PIAT-RR and (4) for BPI. The first column for each outcome replicates the findings in Currie and Thomas, 1995, and Deming, 2009, but for the later cohort the effect on test scores is not the same as in those papers, which is mainly because we have greatly extended the sample to include younger cohorts of children.

³⁷Kling, Ludwig and Katz, 2005, find that youth tend to underreport antisocial behavior, namely arrests. Our measure of crime and other social behaviors rely on self-reported information, however, this underreport will only bias our estimates if it occurs differentially on either side of the cutoff. We do not suspect that this is the case, since our balancing checks (table 3) show that we cannot reject the null that those just eligible and just ineligible are similar in terms of pre-HS characteristics. Thus, there is no association between HS eligibility and some characteristics which could be associated with different reporting of behaviors. Other concern with adult outcomes, namely crime, is the fact that they could be driven by attrition. To understand if those that attrite from the sample at ages 20-21 (but observed in data at ages 12-13 or 16-17) are systematically different in terms of likelihood to commit crime than those that do not attrite we perform the following exercise. We use the sample of children around the cutoff and estimate regressions versions of equation 1: we estimate regressions of several child and family outcomes before 20-21 years old on an indicator for whether the individual will be missed from the data at ages 20-21, this indicator interacted with eligibility at age 4 and eligibility to HS at age 4, and we include the controls in table 4. We cannot detect any significant pattern in terms of how prone to crime those with outcome missing at ages 20-21 are. To be more precise, those present in sample at ages 12-13 or 16-17, but missing at age 20-21, seem have the same pre-age 20 characteristics on either side of the cutoff. This suggests that our results of the effects on crime are not driven by some differential pattern of attrition among just eligible and just ineligible. We thank this point to an anonymous referee.

sults for a selected set of outcomes (for the sample of males). As in figure 1, we use a bandwidth equal to 0.3 (results are similar if we use bandwidths 0.2 or 0.4). The figures suggest that there are discontinuities in the level of the outcomes we study at the eligibility cutoffs, and they have the same sign as those reported in the tables above. However, for all the outcomes we consider there is a fair amount of oscillation in both sides of the discontinuity.³⁸

5.3.3 Sensitivity to functional form, sample size and effects of other programs

Table 8 shows that our results are robust to a battery of sensitivity checks, namely, functional form of running variable, sample size and effects of other programs. We use one outcome for each age group studied, the summary index of table 4. In tables B.15-B.17 in the Appendix we include additional estimates for each of these three exercises, where selected individual components of the index are used as the dependent variable (those components for which we find the strongest impacts of the program). We focus on males, which is the sample driving our results.

We start by examining Panel A, which shows changes to the specification and to the set of controls. The first row presents our basic specification, giving us the main results presented so far. In the second row we exclude several control variables from the model, namely those corresponding to pre-age 4 characteristics, while in the third row we expand the set of pre-age 4 variables we include in the model (see the note to the table). In fourth and fifth rows we change the order of the polynomial in income and family size, from cubic, to either quadratic or quartic. Results are fairly similar across rows.

Panel B of Table 8 shows the sensitivity of our results to the size of the window of data used around the discontinuity. We construct these intervals based on values of family income, taken as a proportion of the household specific cutoff. The third row in the panel is the benchmark displaying our main results. The other rows

³⁸When we redo the regressions in tables 4-7 using distance to the eligibility cutoff as the running variable instead of family income and family size our estimates which are slightly smaller than the ones we report as our basic specification, but always of the same sign and similar magnitude. Although they do not remain statistically significant for the outcomes where we saw the strongest effects at ages 12-13, they remain statistically significant at ages 16-17 and ages 20-21. These results are shown in table B.22 in Appendix.

present different window sizes going from very small (first row) to an income three times the cutoff (last row). If the window is very small so is the sample size, and the estimates become more noisy. If the window is too large we are using large amounts of data that are not very relevant for the parameter of interest, which can make the problem of misspecifying the polynomial worse. Our results are robust to reasonable changes in window size, only changing substantially when the window is very large.

Panel C examines whether our estimates are potentially capturing the impacts of other programs. As mentioned in section 3, eligibility to Head Start is correlated with eligibility to other programs, such as AFDC, Medicaid, or SSI³⁹. It is therefore possible that the estimates in tables 4-7 confound the effects of Head Start with those of other programs. However, while most of these programs exist during several years of the child's life, Head Start is only available when the child is between ages 3 and 5. This fact allows us to assess whether confounding effects from other programs are likely to be important.

Our reasoning is as follows. Suppose that we estimate equation (1) using eligibility (as well as the covariates) measured at different ages of the child. If participation in other programs is driving our results, E_i should have a strong coefficient even when measured at ages other than 3 to 5. Otherwise, we can be confident that our estimates reflect the impact of Head Start, since it is (possible but) unlikely that other programs affect child development only if the child enrolls at ages 3 to 5, but have no effect if she enrolls either at ages 0, 1, 2, 6 or 7.⁴⁰

In this last panel we present estimates of the impact of eligibility to Head Start at different ages on the summary index. Each row represents a different regression, where the age of eligibility (and the corresponding controls) varies from 2 to 6. Across the rows, the largest and strongest estimates occur consistently at age 4, and

³⁹In results available from the authors, it is possible to confirm that our eligibility variable is also a good predictor of participation in these other programs.

⁴⁰This reasoning will work if the set of individuals who are at the margin of eligibility at ages 3 to 5, are different from those who are at the margin of eligibility at ages 0, 1, 2, 6 and 7. If they were all the same individuals it would be impossible to distinguish eligibility to Head Start (only at ages 3 to 5) from eligibility to other programs (at all ages). Furthermore, it is not possible to rule out that other programs have most of its influence at ages 3-5.

sometimes 5.⁴¹ We take this as suggestive evidence that, by using our procedure, we are capturing the impact of Head Start and not of other programs.

5.3.4 Additional results

We also perform additional robustness checks, which we summarize here. The details are included in Appendix A. First, we include an analysis of the main results separated by race groups, although in this case we focus on males alone (table B.18). We find that for children ages 12-13 the overall effects are driven by the non-Black with an improvement on the summary index (although the effects on obesity come from the Black sample). One additional effect (robust to multiple hypotheses testing) found for non-Black children is a decrease in the probability of enrolment in special education, which is consistent with improvement in White children's school performance also find by Currie and Thomas (1995). Among adolescents 16-17, the effects are mainly driven by the Black sample, with an improvement in the summary index. The effects on being overweight are driven by Black adolescents (which is consistent with the findings of Frisvold, 2011), whereas the impacts on mental health come from the non-Black sample. Finally, the effects on crime related activities among young adults are due to reduced engagement in criminal activities by the non-Black.

Second, we study whether there are differences in program impacts across cohorts of children, because they may tell us something about changes in the program over time. We only do this for the youngest age group (12-13), for whom we have a larger sample (see table B.19; again only the sample of boys is used). We separated children 12-13 years of age into two groups: those who could have been eligible to attend the program in the 1980s (born between 1977 and 1984), and those who

⁴¹We present further evidence in table B.17 in Appendix, where we use include estimates for individual outcomes, but also for eligibility measured between ages 0 and 7. We find that across columns the strongest effects appear when eligibility is measured at age 4, sometimes at age 5 (BPI and overweight in panels A.1 and B.1). The only exception is in panel A.3 - for the significant association between use of special equipment and eligibility at age 7; there is also a significant association between the need to use special equipment and eligibility at ages 0 and 1, but the coefficient goes in the opposite sign. In Panel C.1 there is a positive mild association between "ever sentenced" at ages 20-21 and eligibility at ages 1 and 6; in panel C.2 there is also a positive mild association between "idleness" at ages 20-21 and eligibility at ages 2 and 7.

could have attended it in the 1990s (born in 1985-2000). The reduced form estimates show that most of the effects we find in the overall sample are driven by the set of children who attended the program in the 1980s.

Third, we investigated the mechanisms behind the effects found, studying whether Head Start is associated with a response in parental labor supply around Head Start age, and if it exists a reinforcing or compensatory response of parents with respect to child investments as a response to the program (see Gelber and Isen, 2013). We start by analyzing labor supply of mothers and her spouses in the years prior and during Head Start age (table B.20) to learn whether parents (especially, mothers) use the fact that children are in child care for job search (parents can use the Head Start years to improve their current and future employment prospects, through the services offered by the program). We find that during the period in which children can attend HS there is a drop in the weekly hours worked by mothers of boys at the cutoff, suggesting that there is not an immediate recover of labor market prospects for mothers of children that just become eligible. Regarding parental investments (table B.21) we cannot rule out a zero relation between of eligibility at age 4 and a measure of quality of home environment in the period subsequent to program.

Finally, we also show that our results are robust to three additional sensitivity checks. First, we test for discontinuities in outcomes at non-discontinuity points (table B.23). Second, we allow for serial correlation in state specific shocks by clustering the standard errors at state level (table B.24). Third, we re-estimate equation (1) including only the individuals we observe for all three age groups (12-13, 16-17 and 20-21). Our main conclusions hold in this smaller sample, although the estimates become more imprecise either when we study individual outcomes or use the summary indexes (table B.25).

5.4 Estimates from the Structural Equations

The reduced form analysis of a summary index that aggregates several variables presented in table 4 tells us that HS has overall positive effects for 12-13 and 16-17 males. These positive effects represent strong effects of eligibility to Head Start on behavioral problems, on being overweight, and on the need to use special health

equipment at ages 12 and 13. Among adolescent boys there are strong effects on depression and obesity, and table 7 shows large effects on criminal activity and idleness among young adults. The effects in tables (4)-(7) are our main results, but the estimates in these tables do not correspond to the quantitative impact of the program on individuals because compliance with the program is imperfect, and eligibility does not equal participation. These estimates need to be scaled up by the estimated effect of eligibility on participation, and the best way of doing this is to estimate equation (3) jointly with (4) (Lee and Lemieux, 2010). In doing so, the estimated effects became quite imprecise, reflecting some instability in the procedure. In spite of this, in all cases but one the essential patterns of tables (4)-(7) remain unchanged.⁴²

Table 9 shows estimates of β coming from the system consisting of (3) and (4), for the sample of males. The table reports estimates of β , as well as average marginal effects of Head Start for the discrete outcomes (labeled *Marginal Effect*). Since in the first stage there is only a significant association between eligibility at age 4 and program participation for males, table 9 focus only in this sample. In panel A we present the effect at ages 12-13, and we estimate that participation in Head Start leads to a 26% reduction in the probability of being overweight, a 29% reduction in the probability of needing special health equipment, a 0.6 standard deviation decrease in the behavior problems index for the whole sample and a 129% standard deviation improvement in the summary index (see columns 1-4). We should point that the structural estimates on the summary index are implausibly large, which result from the very fuzzy discontinuity in the first stage (see table 1).⁴³

Panel B presents estimates for ages 16-17. Surprisingly for this age group we

⁴²Behind the instability problem may be the fact that either one or both equations in this system are non-linear and our specifications include a large number of location and time indicator variables. This is particularly true when we estimate bivariate probit models, which involve maximizing non-concave likelihood functions with more than one local maximum. For each outcome we started the optimization routine using the estimates where Head Start participation is considered exogenous, and the results we report correspond to the maximum values of the likelihood that we found. The optimization algorithm used for each outcome is presented in the note of table 9.

⁴³The 2SLS estimates for the summary index by area at ages 12-13 are as follows: cognitive -0.622 (0.468); behaviors 0.687 (0.490) and health 1.272 (0.572) (standard errors in parenthesis).

cannot find any impact of the program on being overweight (perhaps because of numerical difficulties in our procedure), or on the summary index, but we estimate that the program leads to a 0.55 standard deviations decrease in the depression score. Finally, at ages 20 or 21 (Panel C) we find a 22% reduction in the probability of ever being sentenced of a crime (the effect on idleness is not significant), but no effects on the overall index.

In summary, tables 4-7 and table 9 (and the subsequent sensitivity analysis) present a picture of strong effects of Head Start on behavioral and health outcomes of children, which are sustained at least until early adulthood. It is interesting that in the case of behavioral outcomes we were able to find a consistent set of large and statistically significant results, while that is not true for cognitive outcomes (as in the HSIS). As stressed by Cameron, Heckman, Knudsen and Schonkoff (2007), this may be due to the fact that non-cognitive skills are more plastic than cognitive skills, and early childhood interventions are more likely to have sustained effects on the former than on the latter.⁴⁴

6 Summary and Conclusions

In this paper we study the impact of Head Start (a preschool program for poor children) on the risky behaviors and health of children, adolescents and young adults. A recent experimental evaluation of this program, the HSIS, reports little or no effects on children outcomes. However, its focus is on short terms impacts, while our paper focuses on mid to long-term impacts of the program.

Identification of the effects of the program is based on the fact that the probability of program participation is a discontinuous function of household income (and family size) because of the program's eligibility rules, enabling us to use a "fuzzy" regression discontinuity design. There is a large range of discontinuity cut-offs, which vary with family size, family structure, year and state. Therefore, we are able to identify the effect of the program for a wide range of individuals.

We find that Head Start decreases behavioral problems, prevalence of chronic

⁴⁴In table B.26 in the Appendix we also include estimates using a linear probability model for discrete outcomes. These results show that our findings hold under a linear regression model.

conditions and obesity at ages 12 to 13, depression and obesity at ages 16 and 17 and crime at ages 20-21.

The parameter we identify can be interpreted as the effect of expanding marginally the eligibility requirements to the program, and the effects we find are large, sustained and remarkably robust to a battery of tests. A simple cost-benefit analysis (see Appendix E) shows that the program has an internal rate of return of at least 4% (this is higher than the interest rate of other investments, for example, during 2013 the yield curve rates for 30 years US T-Bonds has been fluctuating between 2.8% and 3.8%). These impacts show the potential for preschool programs to improve outcomes of poor children, even when they are universal programs such as HS.

References

- [1] Administration of Children and Families, Department of Health and Human Services, 2011, <http://eclkc.ohs.acf.hhs.gov/>.
- [2] Anderson, Kathryn, James Foster and David Frisvold, 2010. "Investing In Health: The Long-Term Impact Of Head Start On Smoking," *Economic Inquiry*, Western Economic Association International, vol. 48(3), pages 587-602, 07.
- [3] Anderson, Michael L., 2008. "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, American Statistical Association, vol. 103(484), pages 1481-1495.
- [4] Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricists Companion*. Princeton: Princeton University Press.
- [5] Baker, Michael and Kevin Milligan, 2013. "Boy-Girl Differences in Parental Time Investments: Evidence from Three Countries", NBER Working Paper No. 18893, March 2013.
- [6] Barnett, Steven, 1996, "Cost-Benefit Analysis of Preschool Education", PowerPoint presentation, <http://nieer.org/resources/files/BarnettBenefits.ppt>.

- [7] Battistin, E. and Enrico Rettore, 2008, "Ineligibles and Eligible Non-Participants as a Double Comparison Group in Regression Discontinuity Designs", *Journal of Econometrics*, 2008, Volume 142, Issue 2, pp. 715-730.
- [8] Berlinski, Samuel, Sebastian Galiani, and Paul Gertler. 2009. The Effect of Pre-Primary Education on Primary School Performance. *Journal of Public Economics*, 93(1-2): 219-34.
- [9] Berlinski, Samuel, Sebastian Galiani, and Marco Manacorda. 2008. Giving Children a Better Start: Preschool Attendance and School-Age Profiles. *Journal of Public Economics*, 92(5-6): 1416-40.
- [10] Bitler, Marianne, Janet Currie and John Karl Scholz, 2003, "WIC Participation and Eligibility", *Journal of Human Resources*, v38, 2003, 1139-1179.
- [11] Black, Dan A., Jose Galdo, and Jeffrey A. Smith, 2005, "Evaluating the Regression Discontinuity Design Using Experimental Data." Mimeo.
- [12] Blank, Rebecca and Patricia Ruggles, 1996, "When Do Women Use AFDC & Food Stamps? The Dynamics of Eligibility vs. Participation", *The Journal of Human Resources*, 31(1), 1996, 57-89.
- [13] Cameron, Heckman, Knudsen, Shonkoff, 2006, "Economic, Neurobiological and Behavioral Perspectives on Building America's Future Workforce", NBER Working Paper, 12298, June 2006, National Bureau of Economic Research.
- [14] Card, David, and Lara D. Shore-Sheppard, 2002, "Using Discontinuous Eligibility Rules to Identify the Effects of the Federal Medicaid Expansions on Low Income Children", NBER Working Paper, 9058, July 2002, National Bureau of Economic Research.
- [15] Cascio, Elizabeth U. 2009. Do Investments in Universal Early Education Pay off? Long-term Effects of Introducing Kindergartens into Public Schools. National Bureau of Economic Research Working Paper 14951.
- [16] Cunha, Flavio, James J. Heckman, Lance Lochner and Dimitriy Masterov, 2006, "Interpreting the Evidence on Life Cycle Skill Formation," *Handbook of the Economics of Education*, Elsevier.
- [17] Currie, Janet, 2006, "The Take-up of Social Benefits," in Alan Auerbach, David Card, and John Quigley (eds.) *Poverty, the Distribution of Income, and Public Policy*, (New York: Russell Sage) 2006.

- [18] Currie, Janet, Eliana Garces and Duncan Thomas, 2002, "Longer Term Effects of Head Start", *The American Economic Review*, 92:4.
- [19] Currie, Janet and Jonathan Gruber, 1996, "Health insurance eligibility and Child Health: lessons from recent expansions of the Medicaid program," *The Quarterly Journal of Economics*, May, 1996, 431-466.
- [20] Cutler, David and Jonathan Gruber, 1996, "Does Public Insurance Crowdout Private Insurance?", *The Quarterly Journal of Economics*, CXI, May 1996, 391-430.
- [21] Currie, Janet and Matthew Neidell, 2007, "Getting Inside the 'Black Box' of Head Start Quality: What Matters and What Doesn't," *Economics of Education Review*, Vol. 26, Issue 1, February 2007, pg. 83-99.
- [22] Currie, Janet and Duncan Thomas, 1995, "Does Head Start make a difference?", *The American Economic Review*, 85(3), 341 – 364.
- [23] Currie, Janet and Duncan Thomas, 1999, "Does Head Start help Hispanic children?", *The Journal of Public Economics*, 74 (1999), 235-262,
- [24] Daly, Mary and Richard V. Burkhauser, 2002, The Supplemental Security Income Program, Federal Reserve Bank of San Francisco, WP 2002-20.
- [25] Deming, David, 2009, "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start", *American Economic Journal: Applied Economics*, American Economic Association, vol. 1(3), pages 111-34, July.
- [26] Department of Health and Human Services, 2006, Early Head Start Benefits Children and Families, http://www.acf.hhs.gov/programs/opre/ehs/ehs_research.
- [27] Department of Health and Human Services, 2010, "Head Start Impact Study: Final Report", Administration for Children and Families, Washington, DC.
- [28] Department of Health and Human Services, 2011, Income Eligibility for Enrollment in Head Start and Early Head Start Programs <http://www.acf.hhs.gov/programs/ohs/>.
- [29] Department of Health and Human Services, 2010, Head Start Impact Study: Final Report, Executive Summary, January 2010.

- [30] Economic Report of the President, 2012, United States Government Printing Office Washington, February 2012.
- [31] Frisvold, D., 2011, Head Start Participation and Childhood Obesity, September 2011, mimeo.
- [32] Frisvold, D. and J. Lumeng, 2007, Expanding Exposure: Can Increasing the Daily Duration of Head Start Reduce Childhood Obesity?, November 2007.
- [33] Fuerst, J.S. and D. Fuerst(1993). Chicago Experience with an Early Childhood Program: The Special Case of the Child Parent Center Program. *Urban Education*, 28(1):69-96.
- [34] Gelber, Alexander and Adam Isen, 2013, Children's Schooling and Parents' Behavior: Evidence from the Head Start Impact Study, *Journal of Public Economics*.
- [35] General Accounting Office, 1998, Head Start Programs: Participant Characteristics, Services, and Funding <http://www.gao.gov/archive/1998/he98065.pdf>
- [36] General Accounting Office, 2003, Better Data and Processes Needed to Monitor Underenrollment, <http://www.gao.gov/new.items/d0417.pdf>
- [37] Gruber, Jonathan, 2003, "Medicaid" in Means Tested Transfer Programs in the United States Robert Moffitt (ed) (Chicago: University of Chicago Press for NBER).
- [38] Havnes, Tarjei, and Magne Mogstad. 2011. "No Child Left Behind: Subsidized Child Care and Children's Long-Run Outcomes." *American Economic Journal: Economic Policy*, 3(2): 97-129.
- [39] Hahn, J., P. Todd and W. Van der Klaauw, 2001, Identification and Estimation of Treatment Effects with a Regression Discontinuity Design, *Econometrica* 69 (2001), 201-209.
- [40] Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev and Adam Yavitz, 2010. "The rate of return to the HighScope Perry Preschool Program", *Journal of Public Economics*, Elsevier, vol. 94(1-2), pages 114-128, February.
- [41] Imbens, Guido and Thomas Lemieux, 2007, Regression Discontinuity Designs: a guide to practice, NBER Technical Working Paper 337, April 2007.

- [42] Imbens, Guido and J. D. Angrist, 1994, Identification and estimation of local average treatment effects. *Econometrica* 62(2), 467-475, March 1994.
- [43] Kling, J. R., J. Ludwig, and L. F. Katz, 2005, "Neighborhood Effects on Crime for Female and Male Youth: Evidence from a Randomized Housing Voucher Experiment", *Quarterly Journal of Economics*, 120, 87–130.
- [44] Lee, David S., and Thomas Lemieux. 2010. Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, 48(2): 281–355.
- [45] Lo Sasso, Anthony, and Thomas C. Buchmueller, 2002. "The Effect of the State Children's Health Insurance Program on Health Insurance Coverage", NBER Working Paper 9405, National Bureau of Economic Research, December 2002.
- [46] Ludwig, J., and D. Miller, 2007, Does Head Start Improve Children's Life Chances? Evidence from a Regression Discontinuity Design, *The Quarterly Journal of Economics* 122(1), 159-208.
- [47] Ludwig, J. and D. Phillips, 2007, The Benefits and Costs of Head Start, NBER WP 12973.
- [48] Lundberg, S., 2005. "Sons, Daughters, and Parental Behaviour," *Oxford Review of Economic Policy*, Oxford University Press, vol. 21(3), pages 340-356, Autumn.
- [49] Magnuson, Katherine A., Christopher Ruhm, and Jane Waldfogel. 2007. Does Prekindergarten Improve School Preparation and Performance?, *Economics of Education Review*, 26(1): 33–51.
- [50] McCrary, J., 2007, Testing for Manipulation of the Running Variable in the Regression Discontinuity Design, *Journal of Econometrics*, forthcoming.
- [51] Moffitt, Robert A., 1983, "An Economic Model of Welfare Stigma", *The American Economic Review*, 73(5), 1023-1035.
- [52] Neidell, Matthew and Jane Waldfogel, 2006, "Spillover effects of early education: evidence from Head Start", January 2006, manuscript.
- [53] National Collaborative on Childhood Obesity Research, 2009, http://www.nccor.org/downloads/ChildhoodObesity_020509.pdf. Accessed on Jan. 2012.

- [54] Romano, Joseph P. and Michael Wolf, 2005, "Stepwise Multiple Testing as Formalized Data Snooping", *Econometrica*, Vol. 73, No. 4, Jul., 2005, pp. 1237-1282.
- [55] Scholz, John Karl, 1994, "The Earned Income Tax Credit: Participation, Effectiveness", *National Tax Journal*, March, 1994, 59-81.
- [56] U.S. Committee on Ways and Means. Green Book 1994, Washington DC, U.S. Government Printing Office, July 1994.
- [57] Westfall, P. H. and S. S. Young, 1993, *Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment*, New York, Wiley.
- [58] Zigler, E., 2010, *Putting the National Head Start Impact Study into a Proper Perspective*, National Head Start Association (U.S.).

Figures

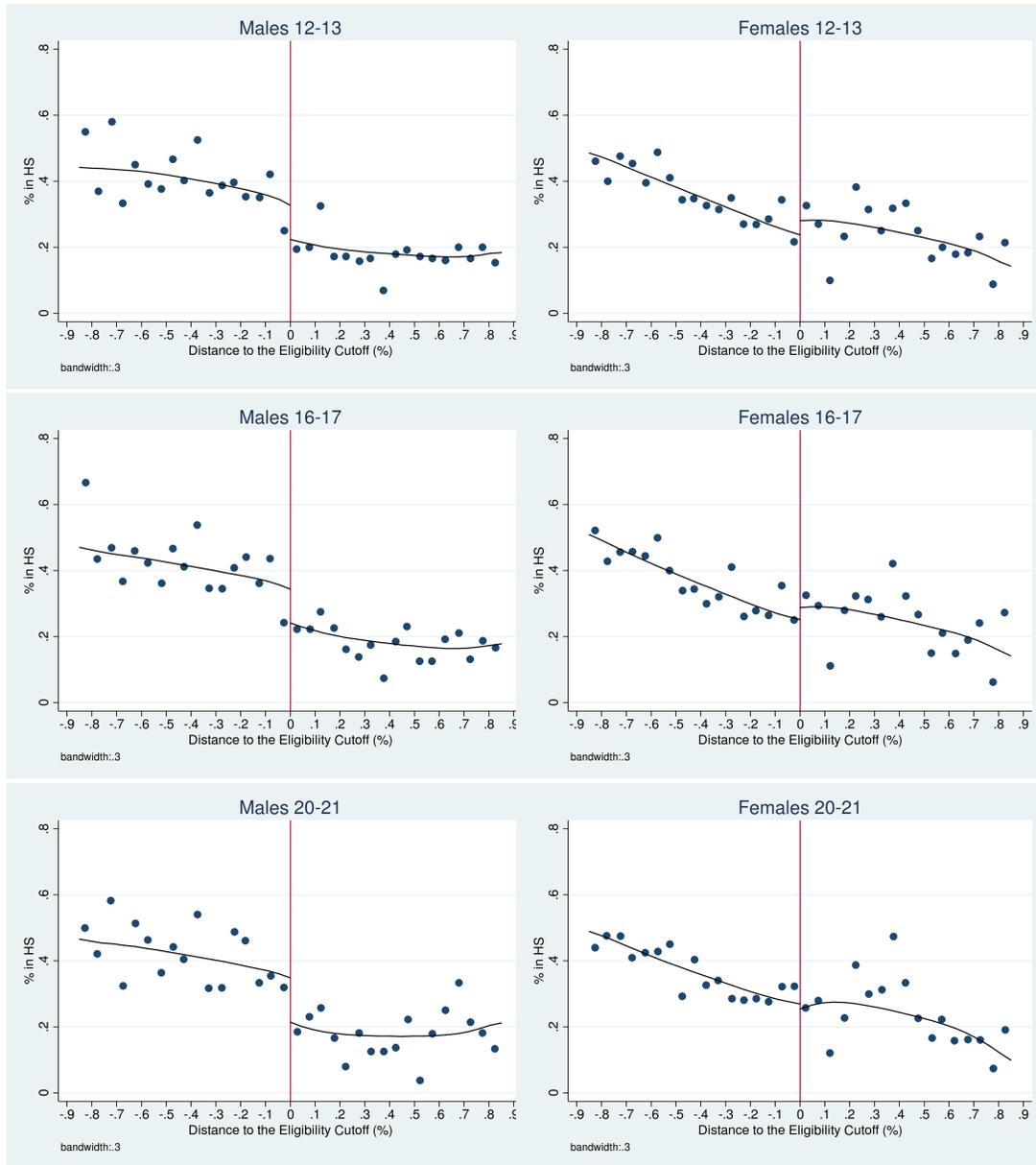


Figure 1: Proportion of children in Head Start, by eligibility status.

Note: The continuous lines are local linear regression estimates of Head Start participation on percentage distance to cutoff; regressions were run separately on both sides of the cutoff and the bandwidth was set to 0.3. Circles in figures represent mean Head Start participation by cell within intervals of 0.05 of distance to cutoff. The kernel used was Epanechnikov.

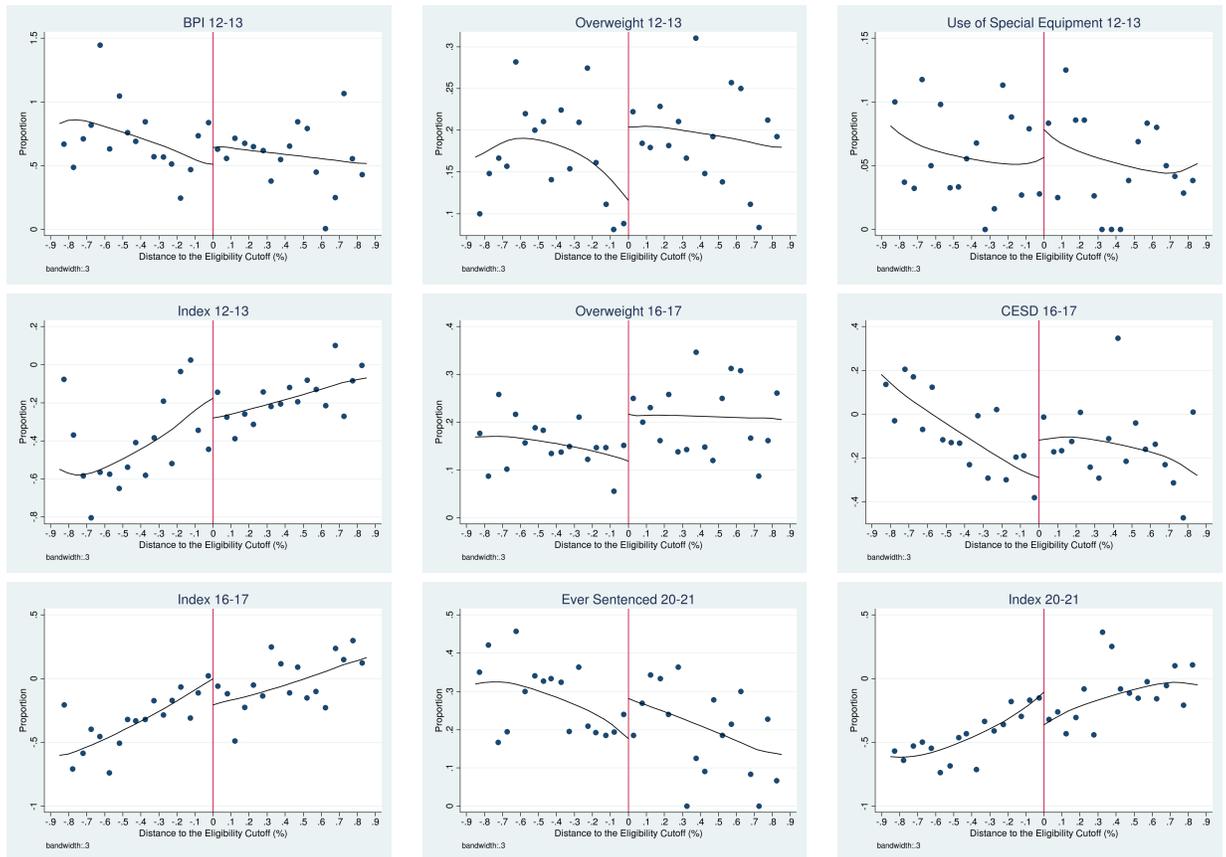


Figure 2: Average outcomes by eligibility status, Bandwidth = 0.3.

Note: The continuous lines are local linear regression estimates of several outcomes on percentage distance to cutoff. The bandwidth was set to 0.3. Circles in figures represent the mean outcome by cell within intervals of 0.05 of distance to cutoff. The kernel used was Epanechnikov. The sample only includes boys.

Table 1: First Stage Estimates.

	(1)	(2)	(3)
Sample	All	Males	Females
Panel A: ages 12-13			
1[HS Eligible at 4]	0.278** [0.118]	0.684*** [0.169]	-0.048 [0.170]
<i>Marginal Effect</i>	0.248	0.209	-0.015
Observations	2,550	1,294	1,256
Control Mean	0.432	0.215	0.272
SD	0.089	0.412	0.446
Panel B: ages 16-17			
1[HS Eligible at 4]	0.313** [0.123]	0.640*** [0.176]	0.014 [0.176]
<i>Marginal Effect</i>	0.252	0.198	0.004
Observations	2,416	1,228	1,188
Control Mean	0.435	0.224	0.275
SD	0.101	0.418	0.448
Panel C: ages 20-21			
1[HS Eligible at 4]	0.311** [0.137]	0.744*** [0.197]	-0.003 [0.186]
<i>Marginal Effect</i>	0.229	0.225	-0.001
Observations	1,977	953	1,024
Control Mean	0.421	0.190	0.261
SD	0.100	0.394	0.441

Note: The table reports results of probit estimates of an indicator for Head Start participation on income eligibility. The marginal effect is the average marginal change in the probability of Head Start participation across individuals as the eligibility status changes and all other controls are kept constant. Controls excluded from the table include cubic in log family income and family size at age 4, an interaction between these two variables, a dummy indicating the presence of a father figure in the household at age 4, race and age dummies, and dummies for year and state of residence at age 4.

The F-test for the exclusion of variable "1[HS Eligible at 4]" from the linear probability model equivalent to the probit model estimated in table are: 5.8 for the whole sample, 17.3 for males and 0.03 for females. Thus, for the sample of males the F-test is above the value of 10, usually used to assess about the weakness of instruments.

Robust standard errors are reported in brackets clustered at state-year at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 2: Control Group - Alternative Child Care.

Sample Program	(1)	(2)		(3)	(4)		(5)		(6)	(7)	(8)	(9)
		HS	Preschool	Informal	HS	Preschool	Informal	HS	Preschool	Informal	HS	Preschool
		Ages 12-13										
1[HS Eligible at 4]	0.684*** [0.169]	-0.392*** [0.146]	-0.325 [0.210]	0.640*** [0.176]	-0.217 [0.148]	-0.624*** [0.227]	0.744*** [0.197]	-0.322* [0.166]	-0.678*** [0.260]			
Marginal Effect	0.209	-0.136	-0.068	0.198	-0.076	-0.132	0.225	-0.113	-0.128			
		Ages 16-17										
Observations	1,294	1,302	1,267	1,228	1,236	1,210	953	951	869			
Control Mean	0.216	0.589	0.195	0.224	0.575	0.201	0.190	0.642	0.168			
SD	0.413	0.493	0.397	0.418	0.496	0.402	0.394	0.481	0.375			
		Ages 20-21										
1[HS Eligible at 4]	-0.0484 [0.170]	0.0945 [0.160]	-0.135 [0.183]	0.0142 [0.176]	0.0587 [0.164]	-0.169 [0.186]	-0.00272 [0.186]	0.121 [0.181]	-0.262 [0.239]			
Marginal Effect	-0.015	0.034	-0.031	0.004	0.021	-0.040	-0.001	0.045	-0.058			
Observations	1,256	1,271	1,239	1,188	1,195	1,174	1,024	1,031	952			
Control Mean	0.272	0.572	0.156	0.275	0.569	0.156	0.261	0.580	0.159			
SD	0.446	0.496	0.363	0.448	0.497	0.364	0.441	0.495	0.367			

Note: The table reports results of probit regressions of different child care arrangements at ages 3-5 on income eligibility at age four (sample of boys). The marginal effect is the average marginal change in the probability of participation in an arrangement across individuals as the eligibility status changes and all other controls are kept constant. Controls excluded from table include: cubic in log family income and family size at age 4, an interaction between these two variables, a dummy indicating the presence of a father figure in the household at age 4, race and age dummies, and dummies for year and state of residence at age 4. Robust standard errors are reported in brackets clustered at state-year at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 3: Balancing results: Pre-Head Start age outcomes.

	(1)	(2)	(3)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Birth weight	Mother's Educ. 0-2	Grandmother's Educ.	Mom married before age 3	Mother's AFQT	Family Income 0-2	Family Size 0-2	Mom lived in south at 14	Lived with parents at 14	Mom's siblings at 14	Mom lived in rural area at 14
Panel A: All											
1 HS Eligible at 4	-0.209*	0.178	-0.217	-0.030	0.632	-0.031	-0.110	-0.023	-0.013	-0.077	-0.037
	[0.109]	[1.884]	[0.251]	[0.025]	[1.724]	[0.055]	[0.122]	[0.025]	[0.040]	[0.253]	[0.032]
RW algorithm	No	No	No	No	No	No	No	No	No	No	No
H0 rejected at 10%	1,116	2,550	2,371	2,550	2,470	2,550	2,550	2,445	2,540	2,542	2,536
Observations	0.164	116.660	9.955	0.897	29.615	9.934	4.273	0.417	0.634	4.602	0.182
Panel B: Boys											
1 HS Eligible at 4	-0.107	0.110	-0.335	-0.057*	-2.618	-0.129	-0.213	-0.049	-0.049	0.146	-0.082*
	[0.170]	[2.840]	[0.370]	[0.035]	[2.506]	[0.079]	[0.162]	[0.036]	[0.057]	[0.333]	[0.048]
RW algorithm	No	No	No	No	No	No	No	No	No	No	No
H0 rejected at 10%	576	1,294	1,193	1,294	1,260	1,294	1,294	1,243	1,290	1,289	1,287
Observations	0.026	119.443	10.249	0.907	33.742	9.971	4.315	0.411	0.645	4.381	0.182
Panel C: Girls											
1 HS Eligible at 4	-0.318*	0.213	-0.145	-0.001	3.485	0.040	-0.012	0.001	0.002	-0.181	0.019
	[0.186]	[2.328]	[0.362]	[0.038]	[2.379]	[0.078]	[0.165]	[0.035]	[0.057]	[0.388]	[0.044]
RW algorithm	No	No	No	No	No	No	No	No	No	No	No
H0 rejected at 10%	540	1,256	1,178	1,256	1,210	1,256	1,256	1,202	1,250	1,253	1,249
Observations	0.310	113.767	9.661	0.886	25.222	9.896	4.228	0.423	0.623	4.830	0.182

Note: The table reports OLS estimates of family and child's outcomes measured before age three on income eligibility. Controls excluded from table include cubic in log family income and family size at age 4, an interaction between these two variables, a dummy indicating the presence of a father figure in the household at age 4, race and age dummies, and dummies for year and state of residence at age 4. The sample used includes children ages 12-13. Robust standard errors are reported in brackets clustered at state-year at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 4: Reduced Form Estimates: Effect of Head Start. Dependent Variable: Global Summary Index.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Sample	Male	Female	All	Male	Female	All	Male	Female	All
	Ages 12-13			Ages 16-17			Ages 20-21		
1[HS Eligible at 4]	0.313** [0.128]	0.184* [0.097]	0.210*** [0.079]	0.266** [0.121]	0.063 [0.124]	0.163* [0.092]	0.194 [0.139]	0.009 [0.139]	0.073 [0.096]
Observations	1,294	1,256	2,550	1,228	1,188	2,416	953	1,024	1,977

Note: This table presents estimates for γ in equation 1. Controls excluded from table: cubic in log family income and family size at age 4, an interaction between these two variables, cubic in log of average family income and family size for ages 0-2, interaction between these two variables, cubic on child's birth weight, dummy for the presence of a father figure in the household at age 4, race and age dummies, and year and state of residence at age 4 effects. "Control Mean" is the mean outcome among observations just above the cutoff (at most 25% above the cutoff). Marginal effects for discrete outcomes in *italic*. Robust standard errors are reported in brackets clustered at state-year at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5: Reduced Form Estimates: Ages 12-13 (sample: males).

	(1)	(2)	(3)	(4)	(5)
	Control Mean	N	ITT	Marginal Effect	RW p-v <0.1
Behaviors					
Drug Use	0.156	1,268	0.118 [0.189]	0.027	No
Overweight	0.198	1,242	-0.379** [0.165]	-0.095	Yes
Grade Retention	0.286	1,285	-0.243 [0.161]	-0.079	No
Alcohol Use	0.467	1,289	-0.249 [0.160]	-0.085	No
School Damage	0.143	1,209	0.0334 [0.161]	0.008	No
Ever smoke	0.359	1,281	-0.146 [0.160]	-0.048	No
marginal effect					
Special Education	0.239	1,254	-0.250 [0.169]	-0.075	No
BPI	0.654	1,211	-0.274** [0.125]		Yes
Health					
Health requires use sp. equip.	0.0938	1,111	-0.777*** [0.287]	-0.101	Yes
Health requires freq. visits to doctor	0.190	1,273	-0.323* [0.184]	-0.083	No
Health requires use of medicines	0.207	1,251	-0.214 [0.179]	-0.056	No
Health limitations	0.0683	1,115	-0.168 [0.216]	-0.028	No
Cognitive					
PIAT-M	0.047	1,197	0.027 [0.100]		No
PIAT-R	0.156	1,196	-0.238* [0.133]		No
PIAT-RC	-0.161	1,181	-0.144 [0.113]		No

Note: Probit (OLS for BPI, PIAT) estimates. Controls excluded from table: cubic in log family income and family size at age 4, an interaction between these two variables, cubic in log of average family income and family size for ages 0-2, interaction between these two variables, cubic on child's birth weight, dummy for the presence of a father figure in the household at age 4, race and age dummies, and year and state of residence at age 4 effects. "Control Mean" is the mean outcome among observations just above the cutoff (at most 25% above the cutoff). Marginal effects for discrete outcomes in column (4). Robust standard errors are reported in brackets clustered at state-year at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 6: Reduced Form Estimates: Ages 16-17 (sample: males).

	(1)	(2)	(3)	(4)	(5)
	Control Mean	N	ITT	Marginal Effect	RW p-v <0.1
In High School	0.914	1,080	0.070 [0.229]	0.009	No
Overweight	0.230	1,167	-0.471** [0.191]	-0.118	Yes
Birth Control	0.612	904	0.093 [0.172]	0.033	No
Health Status	0.669	1,213	0.206 [0.169]	0.070	No
Ever Drunk	0.108	1,167	-0.234 [0.192]	-0.046	No
Ever Sex	0.688	1,214	-0.076 [0.164]	-0.023	No
Ever Sentenced	0.136	1,165	-0.092 [0.193]	-0.018	No
CESD	-0.099	1,053	-0.333*** [0.108]		Yes

Note: Probit (OLS for CESD) estimates. Controls excluded from table: cubic in log family income and family size at age 4, an interaction between these two variables, cubic in log of average family income and family size for ages 0-2, interaction between these two variables, cubic on child's birth weight, dummy for the presence of a father figure in the household at age 4, race and age dummies, and year and state of residence at age 4 effects. "Control Mean" is the mean outcome among observations just above the cutoff (at most 25% above the cutoff). Marginal effects for discrete outcomes in column (4). Robust standard errors are reported in brackets clustered at state-year at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 7: Reduced Form Estimates: Ages 20-21 (sample: males).

	(1)	(2)	(3)	(4)	(5)
	Control Mean	N	ITT	Marginal Effect	RW p-v <0.1
High School Diploma	0.669	943	0.282 [0.189]	0.090	No
Birth Control	0.567	765	0.125 [0.197]	0.045	No
Ever Sentenced	0.284	943	-0.402** [0.200]	-0.114	Yes
Idle	0.112	874	-0.525** [0.250]	-0.090	Yes
Ever in College	0.462	948	-0.052 [0.175]	-0.018	No
Ever worked	0.817	922	-0.162 [0.219]	-0.042	No

Note: Probit estimates. Controls excluded from table: cubic in log family income and family size at age 4, an interaction between these two variables, cubic in log of average family income and family size for ages 0-2, interaction between these two variables, cubic on child's birth weight, dummy for the presence of a father figure in the household at age 4, race and age dummies, and year and state of residence at age 4 effects. "Control Mean" is the mean outcome among observations just above the cutoff (at most 25% above the cutoff). Marginal effects for discrete outcomes in column (4). Robust standard errors are reported in brackets clustered at state-year at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 8: Sensitivity Analysis. Dependent Variable: Global Summary Index (sample: boys).

Ages	(1) 12-13	(2) 16-17	(3) 20-21
Panel A: Functional Form			
Basic	0.313** [0.128]	0.266** [0.121]	0.194 [0.139]
No pre-HS age controls	0.301** [0.126]	0.257** [0.123]	0.154 [0.137]
All controls	0.412*** [0.139]	0.296** [0.134]	0.189 [0.160]
Quadratic	0.304** [0.127]	0.254** [0.122]	0.188 [0.138]
Quartic	0.328** [0.128]	0.268** [0.127]	0.199 [0.143]
Panel B: Trimming around cutoff			
[50% – 150%]	0.291* [0.148] 826	0.174 [0.136] 778	0.208 [0.160] 586
[25% – 175%]	0.356*** [0.132] 1,188	0.227* [0.129] 1,133	0.213 [0.145] 876
[15% – 185%]	0.313** [0.128] 1,294	0.266** [0.121] 1,228	0.194 [0.139] 953
[0% – 300%]	0.026 [0.105] 1,900	0.095 [0.090] 1,791	0.0686 [0.112] 1,382
Panel C: Eligibility at other ages			
1[HS Eligible at age 2]	-0.160 [0.143]	0.0311 [0.138]	-0.307* [0.165]
1[HS Eligible at age 3]	-0.0621 [0.162]	-0.190 [0.129]	-0.106 [0.159]
1[HS Eligible at age 4]	0.313** [0.128]	0.266** [0.121]	0.194 [0.139]
1[HS Eligible at age 5]	0.315** [0.146]	0.150 [0.129]	-0.0378 [0.129]
1[HS Eligible at age 6]	0.007 [0.127]	-0.151 [0.134]	-0.166 [0.146]

Note: In Panel A, "Basic" is the specification used throughout the paper (cubic in log family income and family size at age 4, an interaction between these two variables, a dummy for the presence of a father figure in the child's household at age 4, cubic in average log family income and average family size between ages 0 and 2, an interaction between the two, and cubic in birth weight, race and age dummies and dummies for year and state of residence at age 4). "No pre-Head Start age controls" includes the same controls than in column (1), except those measured before age 3. "All Controls" includes the same controls than in column (1) and dummies for highest grade completed by mother before child turned 3, maternal AFQT score, maternal grandmother's highest grade completed and indicators of maternal situation at 14 years old (whether the mother lived in a Southern, whether she lived with parents and whether she lived in a rural area). "Quadratic" and "Quartic" are the same specification as "Basic" but using polynomials up to the second and fourth order, respectively, in (log) income and family size variables. Panel B includes estimates using the same specification as table (4), but different trimming of data around the cutoff. Panel C includes reduced form results of table (4) with income eligibility measured at different ages between 2 and 6.

Robust standard errors in brackets clustered at state-year at age of eligibility. *, **, *** significant at 10%, 5%, and 1%, respectively.

Table 9: Estimates of structural equations.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Panel A: Ages 12-13			Panel B: Ages 16-17			Panel C: Ages 20-21			
	Overweight	Special Equipment	BPI	Index	Overweight	CESD	Index	Ever Sentenced	Idle	Index
Head Start	-1.255*** [0.328]	-1.743*** [0.101]	-0.647 [0.582]	1.294*** [0.509]	0.0144 [1.897]	-0.552 [0.489]	0.459 [0.497]	-0.824*** [0.288]	-0.561 [1.836]	-0.206 [0.500]
Marginal Effect	-0.290	-0.293			0.004			-0.218	-0.085	
Observations	1,242	1,111	1,211	1,294	1,167	1,053	1,228	943	874	953

Note: Participation in program is instrumented with eligibility status at age four. Controls excluded from table include cubic in log family income and family size at age 4, an interaction between these two variables, cubic in log of average family income and family size for ages 0-2, the interaction between these two variables, cubic on child's birth weight, a dummy indicating the presence of a father figure in the household at age 4, race and age dummies, and dummies for year and state of residence at age 4.

For the discrete outcomes (overweight, ever sentenced and idle) estimates obtained by bivariate probit allow for a tolerance of 0.0001 in the likelihood using the Newton-Raphson algorithm. Standard errors are obtained using the observed information matrix. The marginal effect is average marginal change in outcome across individuals as the participation in Head Start between ages 3 and 5 changes and all other controls are kept constant.

For the continuous outcomes (BPI and CESD) the standard errors for the 2SLS estimates they are obtained by block-bootstrap (500 replications; the block unit is state-year of residence when child was 4). * significant at 10%; ** significant at 5%; *** significant at 1%.

Appendix for: Long-Term Impacts of Compensatory Preschool on Health and Behavior: Evidence from Head Start

A Sensitivity Analysis and Additional Results

A.1 Additional Results

In this Appendix we include additional results that support our analysis. We start by presenting an analysis of the main results separated by race groups and cohorts, although in this case we focus on males alone. We do not include these results in the main text given the substantial decrease in precision when the sample is divided. These results are presented in Table (B.18) in the Appendix.

We find that for children ages 12-13 the overall effects are driven by the non-black with an improvement on the summary index. Regarding individual outcomes, the effects on reduction of chronic health conditions are present on both races, but HS impacts on BPI are driven by the non-Black, whereas the effects on obesity come from the Black sample. One additional effect (robust to multiple hypotheses testing) found for non-Black children is a decrease in the probability of enrolment in special education, which is consistent with improvement in white children's school performance also found by Currie and Thomas (1995). Among adolescents 16-17, the effects are mainly driven by the Black sample, with an improvement in the summary index. The effects on being overweight are driven by Black adolescents (which is consistent with the findings of Frisvold, 2011), whereas the impacts on mental health come from the non-Black sample. Finally, the effects on crime related activities among young adults are due to reduced engagement in criminal activities by the non-Black.

It would also be interesting to check whether there are differences in program impacts across cohorts of children, because they may tell us something about changes in the program over time. We only do this for the youngest age group (12-13), for whom we have a larger sample. When looking at cohort differences, one needs to be careful about changes occurring in the control group across cohorts. In particular, we mention in the main text (table 2) that, relative to the younger cohorts, there is a stronger substitution of Head Start for informal care in the older cohorts. This is consistent with the recent expansion in state provided care.

Our estimates are shown in Table (B.19). We separated children 12-13 years of age into two groups: those who could have been eligible to attend the program in the 1980s (born between 1977 and 1984), and those who could have attended it in the 1990s (born in 1985-2000). The reduced form estimates for the four individual outcomes presented in the table (indicators for a health condition that requires the use of special equipment, being overweight, grade retention, enrolment in special education classes, and an index of behavioral problems-BPI) and for the summary index show that most of the effects we find in the overall sample are driven by the set of children who attended the program in the 1980s. This is interesting and

perhaps surprising, given the large increase in expenditure in Head Start in recent years, but which can perhaps be explained by the change in the control group over time.⁴⁵

We now turn to understand more about the mechanisms behind the effects found. We start by studying whether Head Start is associated with a response in parental labor supply around Head Start age, and if it exists a reinforcing or compensatory response of parents with respect to child investments as a response to the program (see Gelber and Isen, 2013). We start by analyzing labor supply of mothers and her spouses in the years prior and during Head Start age in table B.20 with the aim of learning whether parents (especially, mothers) use the fact that children are in child care for job search. Studying the parental labor supply is also useful since in the years subsequent to Head Start children will be enrolled in school, and thus parents can use the Head Start years to improve their current and future employment prospects, through the services offered by the program. Table B.20 includes three sets of vertical panels, which include estimates for the average number of hours worked per week by the mother and her spouse in the years prior to the Head Start age, at the moment eligibility is assessed and during the HS period. The table further includes three horizontal panels with the estimates for the whole sample of children used in the estimates presented in table 4 (and by child's gender). This table shows that there no relation between the weekly hours worked by mother prior to age 3 and eligibility at age 4. When children become eligible there is a drop in the number of weekly hours worked by mothers at the cutoff, especially among mothers of boys, which is consistent with the fact that the effect identified in this paper is driven by participation in the program of children whose parents suffered a labor market shock that cause their eligibility to the program.⁴⁶ Interestingly, during the period in which children can attend HS there is still a drop in the weekly hours worked by mothers of boys at the cutoff, suggesting that there is not an immediate recover of labor market prospects for mothers of children that just become eligible.

Finally, table B.21 presents estimates using as outcome a measure of parental investments, the HOME score. This table includes three vertical panels, with the estimates for the whole sample of children and for each gender separately. We use the panel dimension of the CNLSY and include multiple estimates per child

⁴⁵Since eligibility cutoffs vary with income and family size, in principle we could examine how the impacts of Head Start varied with these two variables, within the support of the cutoffs. This would be an extra dimension of heterogeneity. Our attempts to do so resulted in imprecise estimates, as we show in the Appendix G. Estimates for equation (6) are presented in Figures (G.2)-(G.3) and they show larger effects for children living in larger families.

⁴⁶The mother suffers a labor market shock, since there is no association between spouses labor supply and eligibility at age 4.

between ages 4 (HS age) and 14. We estimate a version of equation (1), but since we have multiple observations per child we include an interaction between eligibility at age 4 and the child's age group (6-9 or 10-14; 4-5 being the omitted category). Column 1 shows that eligibility at age 4 is associated to a mild improvement in parental investments for the whole sample after age 10 relatively to HS age, which is driven by males (column 2). However, the p-values in the bottom of the table show that we cannot rule out a zero relation between of eligibility at age 4 and HOME score between ages 10-14.

Other Robustness Analysis We present here three additional robustness tests. First, we test for discontinuities in outcomes at non-discontinuity points and we are unable to detect non-zero impacts where impacts do not exist (see Table B.23). Columns (1) and (2) of Table B.23 include estimates of size of the discontinuity when we set the cutoff to be 50% above and below, respectively, of the original cutoff. We estimate that only 2 out of the 18 coefficients that should be zero is statistically significant, and they have the opposite sign to the original estimate. In columns (3) and (4) the fake and the real discontinuities and only two of the 18 coefficients that should be zero are found to be statistically different from zero.)

Second, we allow for serial correlation in state specific shocks by clustering the standard errors at state level. A version of the results presented in Tables (B.11)-(B.13) with these standard errors is presented in Table (B.24) and the estimates of the standard errors of the coefficients are very similar to those in our main results.

Finally, we re-estimate equation (1) including only the individuals we observe for all three age groups (12-13, 16-17 and 20-21), i.e., forcing the sample to be the same for regressions at different ages. It is useful to see how our results would change if we were to use exactly the same sample of children to estimate the impacts of Head Start at different ages, since we would then be able to isolate age effects (assuming time effects are not important), but only for a limited set of cohorts. Our estimates for this exercise are presented in Table B.25, and show that our main conclusions hold in this smaller sample, although they become more imprecise either when we study individual outcomes or use the summary indexes.

A.2 Measurement Error in Income

Our estimates of the size of discontinuity in participation and effects of Head Start at the cutoff might be underestimated due to measurement error in the running variable. This is especially problematic here because we are measuring income using survey data. Therefore, we assess the robustness of our results to alternative measures of family income constructed in the CNLSY.

The family income measure we use in our main specification is the Total Net Family Income for the years of 1981-2000. This variable includes the following components if all are not missing: respondent's and spouse's military income, their salaries and farm/business income and spouse's unemployment insurance; alimony; child support (own and spouse's); AFDC and other public assistance; respondent's and spouse's educational benefits and other kinds of scholarships, fellowships, or grants; other veteran benefits; worker compensation or disability payments received by respondent; total amount of money received by the respondent (or wife/husband) from any source excluding the previous categories, such as savings, payments from Social Security, net rental income or any other regular or periodic sources of income; total income received by the respondent from adults that also live in the household and are related to her (excluding spouse and children); total amount of money received by the respondent (or wife/husband) from persons living outside household (if living in the same dwelling), outside her home in city of permanent residence (if living in Dorm, fraternity or sorority) or not living with respondent (if in military); and, finally, food stamps.⁴⁷

We consider two additional income measures in our sensitivity analysis. The first one is the same as the one above subtracting Food Stamps, which is perhaps a better approximation to the measure of income officially used to determine eligibility than the one we use in the baseline specification. The second one is the sum of the non-missing income components described in the previous paragraph. This is very much like the baseline measure of income used in the paper, but with the inclusion of non-missing components as opposed to requiring that each of them is not missing to allow for a larger sample size. The three income measures analyzed in this paper are highly correlated but they are not exactly the same.

Estimates of the equation describing participation in Head Start when we use these two alternative measures of income are included in Tables A.1. These are very similar to those reported in Table 1. The impact of Head Start eligibility on outcomes estimated for alternative income measures is presented in Table A.2 for the samples ages 12-13, 16-17 and 20-21. There is some sensitivity across different measures of income, but overall the main results remain the same.

⁴⁷Cole and Currie (1994) document inconsistencies in the income measures reported in NLSY79. Therefore, we follow their recommendations to minimize these inconsistencies. First, we check if income categories reported by siblings living with parents are consistent among them. If not, we choose the mode among the values reported by the siblings. If there is no unique mode, we use the value reported by the oldest child that lives with parents (this consistency check is only performed between 1979-1986, which are the years with information for whether the member of NLSY79 is living with the parents). Second, whenever we use individual components of income, we identify those cases in which both spouses are present in the NLSY79, and if the reported earnings for one spouse are not consistent with the own report, we use the own report.

References

- [1] Nancy Cole and Janet Currie, 1994. "Reported Income in the NLSY: Consistency Checks and Methods for Cleaning the Data," NBER Technical Working Papers 0160, National Bureau of Economic Research, Inc.

Table A.1: First Stage: Alternative measures of income.

	(1)	(2)	(3)	(4)	(5)	(6)
Age group	12-13		16-17		20-21	
Sample	Males	Females	Males	Females	Males	Females
Panel A: Alternative 1						
1[HS Eligible at 4]	0.496***	0.0311	0.518***	0.102	0.664***	-0.0935
	[0.186]	[0.208]	[0.190]	[0.212]	[0.210]	[0.238]
Marginal Effect	0.149	0.009	0.157	0.032	0.201	-0.0285
Observations	1,057	1,035	1,003	978	764	867
Control Mean	0.227	0.255	0.221	0.263	0.200	0.252
Panel B: Alternative 2						
1[HS Eligible at 4]	0.536***	0.119	0.424**	0.181	0.599***	0.141
	[0.195]	[0.182]	[0.201]	[0.188]	[0.227]	[0.208]
Marginal Effect	0.164	0.037	0.130	0.058	0.186	0.0439
Observations	1,052	1,065	989	1,017	769	901
Control Mean	0.264	0.284	0.267	0.307	0.214	0.287

Note: Probit estimates of Head Start participation on income eligibility. Controls excluded: cubic in log family income and family size at age 4, an interaction between these two variables, a dummy indicating the presence of a father figure in the household at age 4, dummies for race, child's age and for year and state of residence at age 4. *Alternative measure 1* is equal to the measure of income used in paper, excluding Food Stamps. *Alternative 2* is equal to the sum of different income components. Robust standard errors are reported in brackets clustered at state-year at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table A.2: Reduced Form Estimates with alternative income measures (sample: boys).

Age groups	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables	Overweight	12-13 Use of Special Equipment	BPI	Overweight	16-17 CESD	Ever Sentenced	20-21 Idle
	Panel A: Alternative 1						
1[HS Eligible at 4]	-0.422** [0.166]	-0.831** [0.336]	-0.333*** [0.105]	-0.293 [0.200]	-0.213* [0.123]	-0.244 [0.293]	-0.233 [0.328]
Marginal Effect	-0.100	-0.111		-0.070		-0.069	-0.039
Observations	957	853	947	916	822	748	693
	Panel B: Alternative 2						
1[HS Eligible at 4]	-0.108 [0.136]	-0.543* [0.288]	-0.368*** [0.114]	-0.233 [0.203]	-0.250** [0.103]	-0.349 [0.218]	0.0375 [0.263]
Marginal Effect	-0.028	-0.071		-0.058		-0.100	0.006
Observations	1,210	1,068	1,180	1,123	1,025	921	863

Note: Probit estimates for several outcomes (OLS estimates for BPI and CESD). Controls excluded from table include cubic in log family income and family size at age 4, an interaction between these two variables, cubic in log of average family income and family size for ages 0-2, the interaction between these two variables, cubic on child's birth weight, a dummy indicating the presence of a father figure in the household at age 4, race and age dummies, and dummies for year and state of residence at age 4. Marginal effect is average marginal change in outcome across individuals as the eligibility status changes and all other controls are kept constant. "Control Mean" is the mean outcome among observations just above the cutoff, that is, at most 25% above the cutoff. Robust standard errors are reported in brackets clustered at state-year at age four level.

* significant at 10%; ** significant at 5%; *** significant at 1%.

B Tables

Table B.1: Summary of legislation

Date	Law Number	Title	Description
1964	88-452	Economic Opportunity Act	Anti-poverty bill to "strengthen, supplement, and coordinate efforts in furtherance" of a policy of the U.S. "to eliminate the paradox of poverty in the midst of plenty". Head Start was not mention in the original act, but it was considered part of the Community Action Program.
1966	P.L. 89-794	Economic Opportunity Act Amendments of 1966	A section was added to Title II making Head Start a part of the Economic Opportunity Act.
1967	P.L. 90-222	Economic Opportunity Act Amendments of 1967	"Follow Through" was added in Title II, to continue services for Head Start children when they enter kindergarten and elementary school. This program was administered by the Office of Education.
1969	P.L. 91-177	Economic Opportunity Act Amendment of 1969	A provision was added allowing children from families above the poverty level to receive Head Start services for a fee.
1972	P.L. 92-424	Economic Opportunity Act Amendment of 1972	A fee schedule for non-poor participants in Head Start was required; fees were prohibited for families below the poverty line. Added a requirement that at least 10 percent of Head Start's enrollment include children with disabilities.
1973	93-202	Postponement of a Head Start Fee Schedule	Prior approval by Congress was required before any Head Start fee schedule could be established.
1974	P.L. 93-644	Head Start, Economic Opportunity, and Community Partnership Act of 1974	Reauthorized Head Start through the fiscal year of 1978. Head Start funds should be allocated to states proportionately based upon each state's relative number of children living in families with income below the poverty line and the relative number of public assistance recipients in each state.
1978	P.L. 95-568	Economic Opportunity Act Amendment of 1978	Reauthorized Head Start for three more years. Minor changes to the law.
1981	P.L. 97-35 (42 USC 9831 et. Seq.)	Economic Opportunity Amendment of 1981	The Head Start Act was attached to the OBRA of 1981. To "promote school readiness by enhancing the social and cognitive development of low-income children."
1984	P.L. 98-558	Human Services Reauthorization Act of 1984	Head Start Reauthorization for 2 years. In 1984, the Indian and Migrant branches of Head Start became separate regions; prohibited changes in methods for determining eligibility for low income if they would reduced participation of persons in the program. HS may provide services to children age 3 to the age of compulsory school attendance.
1989	P.L. 101-120	Head Start Supplemental Authorization Act of 1989	Reauthorized Head Start for FY of 1990.
1990	P.L. 101-597	National Health Service Corps Revitalization Amendments of 1990	Minor amend to Head Start Act (library of congress 101th congress pub laws)
1990	P.L. 101-501	Head Start Reauthorization Act of 1990.	Reinforced importance of parental involvement, improved information on Head Start programs.
1992	P.L. 102-763	Head Start Improvement Act	Facilities purchase; Extended waivers for non-federal regulations; Establishment of transportation regulations; Health services to younger siblings; Protection of the quality set-aside; Literacy and child development training to parents; Elimination of priority status to a grantee once funded.
1994	P.L. 103-218	Head Start Act Amendments of 1994	Reauthorized Head Start for the years of 1995 through 1998. Required the development of quality standards (including revising the Program Performance Standards), the development of performance measures, and improved monitoring of local programs. It authorized family-centered programs for infants and toddlers. It established new standards for classroom teachers and family service workers.
1998	P.L. 105-285	Coats Human Services Reauthorization Act of 1998	Reauthorized Head Start for 5 years.
2007	110-134	Improving Head Start for School Readiness Act of 2007	Allows grantees to serve additional children from families with income up to 130% of poverty to be served; formula allocation remains at 100% of poverty; expansion of both Head Start and Early Head Start programs with additional funds going to states serving fewer than 60 percent of eligible children; establishes standards for the curriculum of teachers.

Note: Source of regulations relevant to Head Start: 45 CFR (Code of Federal Regulations), Parts 1301 to 1311. Additional Program Instructions and Information Memorandums can be found at the Early Childhood Learning and Knowledge Center web site: <http://eclkc.ohs.acf.hhs.gov/hslc>.

Table B.2: Descriptive Statistics: Covariates.

	(1)	(2)	(3)
Variable	Obs	Mean	SD
Treatment between ages 3-5			
Head Start	2833	0.31	0.46
Other preschool	2833	0.52	0.50
Other child care	2833	0.17	0.38
Mother's Characteristics			
AFQT	2742	-7.56	19.66
HS Dropout	2833	0.38	0.49
HS graduate	2833	0.52	0.50
College	2833	0.10	0.30
Characteristics at entry (age 4)			
Total Family Income	2833	18338.34	10022.31
Father Figure present	2833	0.52	0.50
Family Size	2833	4.46	1.81
Poor	2833	0.57	0.49
Eligible	2833	0.60	0.49
Child's characteristics			
Birth weight (ounces)	2833	114.36	22.08
Low birth weight	2833	0.10	0.30
Breastfed	2806	0.35	0.48

Note: This table reports means and standard deviations for control variables in our sample. Statistics are reported for males and females whose controls are all not missing and whose family income at age four is between 15% and 185% of the maximum level of income that would allow participation in Head Start. We report means and standard deviation using only one observation per individual. All monetary values are deflated to 2000 values.

Table B.3: Description of variables.

Variables		Ages
PIAT-Math	Test that measures a child's attainment in mathematics as taught in mainstream education. Standardized score with population mean 0 and standard deviation 1 (normed within age).	12-13
PIAT-RR	Test that measures word recognition and pronunciation ability. Standard score with population mean 0 and standard deviation 1 (normed within age).	12-13
PIAT-RC	Measures child's ability to derive meaning from sentences. Standard score with population mean 0 and standard deviation 1 (normed within age).	12-13
BPI	The Behavior Problems Index measures the frequency, range, and type of childhood behavior problems for children age four and over Standard score with population mean 0 and standard deviation 1 (normed within age).	12-13
Grade Repetition	Ever repeated a grade up to a given age	12-13
Special Education	Attending classes for remedial work	12-13
Overweight	Indicator that takes value 1 if the individual's Body Mass Index (BMI) is above the 95th percentile for her/his age and gender.	12-13 16-17
School damage	Ever damaged school property	12-13
Alcohol Use	Ever tried alcohol	12-13
School damage	Ever damaged school property	12-13
Drug Use	Ever tried cocaine or marijuana	12-13
Ever smoked	Ever smoked	12-13
Health: equipment	Child has had between ages 6 and 12-13 health condition required use of special equipment, such as a brace, crutches, a wheelchair, special shoes, a helmet, a special bed, a breathing mask, an air filter, or a catheter and so on	12-13
Health: doctor	Child has had between ages 6 and 12-13 health condition that required medical attention	12-13
Health: medicines	Child has had between ages 6 and 12-13 health condition that required regular use of medicines	12-13
Any limitation	Child has health problem that limits school attention, work or play activity	12-13
CESD	Center for Epidemiological Studies Depression Scale: percentile score that measures symptoms of depression (higher scores are negative)	16-17
High School	Attending high school	16-17
Ever Sex	Ever had sexual relations	16-17
Health Status	Self-reported health status (1 if has good/excellent health status)	16-17
Ever Drunk	Ever got drunk	16-17
Birth control	Use of birth control when have sexual intercourse (only if ever had sexual relations)	16-17 20-21
Ever sentenced	Indicator for whether the individual has ever been convicted of any charge other than minor traffic violations or sentenced to a corrections institution/jail/reform school by a give age (based on self-reported information; the sample for analysis is unchanged if information about the place of residence, which includes prison, is used).	16-17 20-21
High School Diploma	1 if has high school diploma as opposed to be GED or not having high school diploma	20-21
Ever in college	ever in college up to a given age	20-21
Ever worked	Ever worked for pay	20-21
Idle	If the respondent is not enrolled in school/college and reports zero wages	20-21
Indexes		
Behaviors	BPI, drugs use, dummy for overweight child, grade retention, alcohol use, ever tried a cigarette, cause of damage in school facilities, special education attendance.	12-13
Health	Any health condition that limits school work or attendance and play activity, frequent use of medicines, frequent visits to doctor, health condition that requires use of need of special equipment.	12-13
Cognitive	PIAT-Math, PIAT-Reading Recognition, PIAT-Reading Comprehension.	12-13
Global	All above	12-13
Global	Overweight indicator, indicator for ever getting drunk at least once by ages 16-17, high school enrolment, self-report health status good/excellent, ever had sexual relations, use of birth control methods, indicator for whether the individual has ever been convicted of any charge other than minor traffic violations or sentenced to a corrections institution/jail/reform school, CESD.	16-17
Global	Idle, indicator for whether the individual has ever been convicted of any charge other than minor traffic violations or sentenced to a corrections institution/jail/reform school, college enrolment, indicator for high school diploma, ever worked for pay, use of birth control methods when engaging in sexual relationships.	20-21

Table B.4: Descriptive Statistics: Outcomes.

	(1)	(2)	(3)
Variable	Obs	Mean	SD
Age 12-13			
BPI	2392	0.58	1.02
Overweight	2438	0.19	0.39
Special Education	2464	0.23	0.42
Grade Repetition	2537	0.29	0.45
Health: Use special equipment	2049	0.06	0.23
Health: needs doctor	2472	0.16	0.37
Health: needs medicine	2428	0.16	0.37
Any health limitation	2212	0.08	0.27
PIAT-Math	2376	-0.20	0.89
PIAT-RR	2376	-0.07	1.05
PIAT-RC	2349	-0.39	0.89
School damage	2357	0.15	0.36
Alcohol Use	2516	0.39	0.49
Drug Use	2424	0.15	0.36
Ever smoke	2530	0.35	0.48
Index: Global	2550	-0.14	1.02
Index: Cognitive	2384	-0.29	0.98
Index: Behaviors	2550	-0.20	1.04
Index: Health	2550	0.00	1.01
Age 16-17			
Overweight	2267	0.17	0.37
Ever sentenced	2173	0.11	0.32
CESD	2063	0.04	0.98
In High School	2151	0.92	0.27
Birth control	1836	0.60	0.49
Ever Drunk	2236	0.13	0.33
Ever Sex	2377	0.65	0.48
Health Status	2390	0.63	0.48
Index: Global	2416	-0.19	1.04
Age 20-21			
Ever sentenced	1904	0.19	0.39
High School Diploma	1965	0.70	0.46
Idle	1854	0.11	0.32
Birth control	1639	0.59	0.49
Ever worked	1932	0.74	0.44
Ever in college	1969	0.49	0.50
Index: Global	1977	-0.23	1.07

Note: This table reports means and standard deviations for outcome variables in our sample. Statistics are reported for males and females whose controls are all not missing and whose family income at age four is between 15% and 185% of the maximum level of income that would allow participation in Head Start. We report means and standard deviation using only one observation per individual.

Table B.5: Participation in Head Start and eligibility.

Sample	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Eligibility age at 3			Eligibility age at 4			Eligibility age at 5		
	All	Males	Females	All	Males	Females	All	Males	Females
1[HS Eligible]	0.189 [0.119]	0.128 [0.177]	0.124 [0.171]	0.278** [0.118]	0.684*** [0.169]	-0.048 [0.170]	0.206* [0.121]	0.457*** [0.171]	-0.031 [0.179]
Marginal Effect	0.060	0.0389	0.039	0.248	0.209	-0.015	0.068	0.148	-0.01
Observations	2,302	1,190	1,112	2,550	1,294	1,256	2,123	1,082	1,041
Control Mean	0.264	0.282	0.239	0.432	0.216	0.272	0.270	0.233	0.302
SD	0.442	0.451	0.428	0.089	0.413	0.446	0.445	0.424	0.461

Note: Table of probit estimates of Head Start participation between ages 3-5 on income eligibility between ages 3-5. The sample includes children ages 12-13 included in the analysis. The controls excluded from the table include cubic in log family income and family size at age 4, an interaction between these two variables, a dummy indicating the presence of a father figure in the household at age 4, race and age dummies, and dummies for year and state of residence at age 4. Robust standard errors are reported in brackets clustered at state-year at eligibility. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.6: First Stage Estimates.

	(1)	(2)	(3)	(4)
Sample	Black		Non-Black	
	Males	Females	Males	Females
Panel A: ages 12-13				
1[HS Eligible at 4]	0.873***	0.172	0.474**	-0.293
	[0.259]	[0.260]	[0.234]	[0.238]
<i>Marginal Effect</i>	<i>0.273</i>	<i>0.058</i>	<i>0.126</i>	<i>-0.077</i>
Observations	563	566	708	655
Control Mean	0.359	0.348	0.140	0.225
SD	0.484	0.480	0.349	0.420
Panel B: ages 16-17				
1[HS Eligible at 4]	0.811***	0.115	0.450*	-0.127
	[0.274]	[0.268]	[0.243]	[0.254]
<i>Marginal Effect</i>	<i>0.255</i>	<i>0.040</i>	<i>0.124</i>	<i>-0.035</i>
Observations	535	550	664	599
Control Mean	0.361	0.348	0.150	0.223
SD	0.484	0.480	0.359	0.419
Panel C: ages 20-21				
1[HS Eligible at 4]	1.062***	0.229	0.496*	-0.183
	[0.327]	[0.296]	[0.261]	[0.281]
<i>Marginal Effect</i>	<i>0.314</i>	<i>0.076</i>	<i>0.134</i>	<i>-0.049</i>
Observations	413	484	518	509
Control Mean	0.300	0.346	0.144	0.209
SD	0.464	0.480	0.353	0.409

Note: The table reports results of probit regressions of Head Start participation on income eligibility. The marginal effect is the average marginal change in the probability of Head Start participation across individuals as the eligibility status changes and all other controls are kept constant. The controls excluded from the table are cubic in log family income and family size at age 4, an interaction between these two variables, a dummy indicating the presence of a father figure in the household at age 4, race and age dummies, and dummies for year and state of residence at age 4. Robust standard errors are reported in brackets clustered at state-year at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.7: First Stage: Head Start Impact Study.

	(1)	(2)
Panel A: Full Sample		
HS Assignment	0.689*** [0.01]	0.662*** [0.02]
HS Assignment x Child is Male		0.054* [0.02]
N	4442	4442
Panel B: Age 4 Cohort		
HS Assignment	0.671*** [0.02]	0.627*** [0.02]
HS Assignment x Child is Male		0.090** [0.03]
N	1993	1993
Panel C: Age 3 Cohort		
HS Assignment	0.703*** [0.01]	0.692*** [0.02]
HS Assignment x Child is Male		0.022 [0.03]
N	2449	2449

Note: Data from the Head Start Impact Study. Table reports the coefficients of regressions of HS participation indicator on an indicator of HS assignment (no controls). In column (2) we also include an indicator for a male child (and its interaction with HS assignment). Standard errors shown below coefficient estimates in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.8: Gender gap in parental investments and maternal labor supply.

	(1) Annual weeks worked by mother	(2) HOME score	(3) Cognitive Stimulation	(4) Emotional Support
	Panel A: All			
1[Male]	-0.379 [0.323]	-0.105*** [0.018]	-0.121*** [0.018]	-0.041** [0.016]
Observations	111,304	48,353	45,485	42,898
	Panel B: Multi-children families			
1[Male]	-0.048 [0.100]	-0.090*** [0.008]	-0.100*** [0.008]	-0.038*** [0.009]
Observations	73,888	32,373	30,542	28,707

Note: This table presents coefficients for the male dummy from a regression of the dependent variable on gender, age and year indicators. The sample is restricted to children ages 0-14, when HOME score is constructed. Panel B of the table controls also for mother fixed effects. Robust standard errors clustered by child are presented in brackets. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.9: Compilers characteristics for eligibility and siblings instruments.

Instrument Variable	(1)	(2)	(3)
	$P[X_i = 1]$	Eligibility $P[X_i = 1 HS_{it} > HS_{0t}] / P[X_i = 1]$	Siblings $P[X_i = 1 HS_{it} > HS_{0t}] / P[X_i = 1]$
Family Characteristics			
Poor family (age 0-2)	0.48	1.35	1.58
Mother's Education before age 3			
Mom is High School dropout	0.27	0.95	1.50
Mom is High School graduate	0.51	0.98	0.99
Father-figure at home (age 4)	0.71	0.62	0.73
Mothers AFQT above sample median	0.50	0.84	0.55
Child's Characteristics			
Child was breastfed	0.44	1.08	0.60
Motor score before age 3 above sample median	0.49	0.60	0.86

Note: Column (1) reports the distribution of the population by characteristics, $P[X_i = 1]$, for the sample of boys (2765 children that we observe at ages 12-13, 16-17 and 20-21). We use this data to perform separate estimations of the first stage equation for each group. Columns (2) and (3) reports the relative likelihood of an individual belonging to a particular group (in the compliers group) compared to the population at large. These figures are obtained from the first stage coefficient for each group divided by the overall first stage coefficient. For column (2) we use the same specification than in table 1 and the sample is also restricted around cutoff. In column (3) we present the relative probability for those children with siblings (one in Head Start and other in other preschool or home care).

Table B.10: Reduced Form Estimates: Effect of Head Start. Subindexes measured at ages 12-13.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	
	Male	Female	All	Male	Female	All	Male	Female	All	
Index				Cognitive			Only behaviors			Health
1[HS Eligible at 4]	-0.136 [0.117]	-0.031 [0.099]	-0.099 [0.074]	0.233* [0.122]	0.146 [0.111]	0.149* [0.083]	0.236* [0.133]	0.135 [0.099]	0.169** [0.081]	
Observations	1,201	1,183	2,384	1,294	1,256	2,550	1,294	1,256	2,550	

Note: This table presents estimates for γ in equation 1. Controls excluded from table: cubic in log family income and family size at age 4, an interaction between these two variables, cubic in log of average family income and family size for ages 0-2, interaction between these two variables, cubic on child's birth weight, dummy for the presence of a father figure in the household at age 4, race and age dummies, and year and state of residence at age 4 effects. Robust standard errors are reported in brackets clustered at state-year at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.11: Reduced Form Estimates: Ages 12-13.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Control Mean	N	All ITT	RW pv<0.1	Control Mean	N	Males ITT	RW pv<0.1	Control Mean	N	Girls ITT	RW pv<0.1
Behaviors												
Drug Use	0.140	2,424	0.0170 [0.141] 0.004	No	0.156	1,268	0.118 [0.189] 0.027	No	0.123	1,156	-0.262 [0.198] -0.045	No
<i>marginal effect</i>												
Overweight	0.190	2,438	-0.308** [0.120] -0.080	Yes	0.198	1,242	-0.379** [0.165] -0.095	Yes	0.182	1,196	-0.218 [0.172] -0.056	No
<i>marginal effect</i>												
Grade Retention	0.251	2,537	-0.130 [0.119] -0.041	No	0.286	1,285	-0.243 [0.161] -0.079	No	0.216	1,252	-0.0605 [0.181] -0.017	No
<i>marginal effect</i>												
Alcohol Use	0.415	2,516	-0.163 [0.107] -0.056	No	0.467	1,289	-0.249 [0.160] -0.085	No	0.359	1,227	-0.118 [0.142] -0.039	No
<i>marginal effect</i>												
School Damage	0.134	2,357	-0.121 [0.135] -0.027	No	0.143	1,209	0.0334 [0.161] 0.008	No	0.124	1,148	-0.401* [0.208] -0.072	No
<i>marginal effect</i>												
Ever smoke	0.342	2,530	-0.0876 [0.115] -0.029	No	0.359	1,281	-0.146 [0.160] -0.048	No	0.324	1,249	-0.151 [0.169] -0.049	No
<i>marginal effect</i>												
Special Education	0.218	2,464	-0.147 [0.118] -0.041	No	0.239	1,254	-0.250 [0.169] -0.075	No	0.197	1,210	-0.134 [0.180] -0.032	No
<i>marginal effect</i>												
BPI	0.573	2,392	-0.0809 [0.0832]	No	0.654	1,211	-0.274** [0.125]	Yes	0.486	1,181	0.118 [0.120]	No

Reduced Form Estimates: Ages 12-13 (cont.).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Control Mean	N	All ITT	RW pv<0.1	Control Mean	N	Males ITT	RW pv<0.1	Control Mean	N	Girls ITT	RW pv<0.1
Health												
Health requires use sp. equip. <i>marginal effect</i>	0.0751	2,049	-0.535*** [0.191] -0.064	Yes	0.0938	1,111	-0.777*** [0.287] -0.101	Yes	0.0526	938	-0.595** [0.273] -0.058	No
Health requires freq. visits to doctor <i>marginal effect</i>	0.153	2,472	-0.253* [0.138] -0.059	No	0.190	1,273	-0.323* [0.184] -0.083	No	0.114	1,199	-0.242 [0.207] -0.046	No
Health requires use of medicines <i>marginal effect</i>	0.165	2,428	-0.0966 [0.131] -0.023	No	0.207	1,251	-0.214 [0.179] -0.056	No	0.120	1,177	-0.0436 [0.201] -0.008	No
Health limitations <i>marginal effect</i>	0.0639	2,212	-0.230 [0.162] -0.033	No	0.0683	1,115	-0.168 [0.216] -0.028	No	0.0592	1,097	-0.441** [0.221] -0.047	No
Cognitive												
PIAT-M	-0.044	2376	-0.080 [0.065]	No	0.047	1,197	0.027 [0.100]	No	-0.138	1,179	-0.181** [0.0899]	No
PIAT-R	0.088	2376	-0.055 [0.087]	No	0.156	1,196	-0.238* [0.133]	No	0.017	1,180	0.169 [0.113]	No
PIAT-RC	-0.236	2349	-0.114 [0.0730]	No	-0.161	1,181	-0.144 [0.113]	No	-0.312	1,168	-0.049 [0.097]	No

Note: Probit (OLS for BPI, PIAT) estimates. Controls excluded from table: cubic in log family income and family size at age 4, an interaction between these two variables, cubic in log of average family income and family size for ages 0-2, interaction between these two variables, cubic on child's birth weight, dummy for the presence of a father figure in the household at age 4, race and age dummies, and year and state of residence at age 4 effects. "Control Mean" is the mean outcome among observations just above the cutoff (at most 25% above the cutoff). Marginal effects for discrete outcomes in italic. Robust standard errors are reported in brackets clustered at state-year at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.12: Reduced Form Estimates: Ages 16-17.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Control	N	All	RW	Control	N	Males	RW	Control	N	Girls	RW
	Mean		ITT	p<0.1	Mean		ITT	p<0.1	Mean		ITT	p<0.1
In High School	0.924	2,151	0.075	No	0.914	1,080	0.070	No	0.935	1,071	0.076	No
<i>marginal effect</i>			[0.161]				[0.229]				[0.245]	
			0.010				0.009				0.009	
Overweight	0.190	2,267	-0.275**	Yes	0.230	1,167	-0.471**	Yes	0.145	1,100	0.004	No
<i>marginal effect</i>			[0.134]				[0.191]				[0.186]	
			-0.067				-0.118				0.001	
Birth Control	0.580	1,836	0.099	No	0.612	904	0.093	No	0.548	924	0.117	No
<i>marginal effect</i>			[0.121]				[0.172]				[0.194]	
			0.037				0.033				0.042	
Health Status	0.593	2,390	0.239**	Yes	0.669	1,213	0.206	No	0.510	1,177	0.262	No
<i>marginal effect</i>			[0.114]				[0.169]				[0.164]	
			0.087				0.070				0.100	
Ever Drunk	0.104	2,236	-0.180	No	0.108	1,167	-0.234	No	0.100	1,069	-0.276	No
<i>marginal effect</i>			[0.158]				[0.192]				[0.237]	
			-0.035				-0.046				-0.047	
Ever Sex	0.658	2,377	-0.002	No	0.688	1,214	-0.076	No	0.624	1,163	0.049	No
<i>marginal effect</i>			[0.123]				[0.164]				[0.173]	
			-0.0005				-0.023				0.017	
Ever Sentenced	0.112	2,173	0.026	No	0.136	1,165	-0.092	No	0.0827	1,008	0.064	No
<i>marginal effect</i>			[0.155]				[0.193]				[0.227]	
			0.004				-0.018				0.008	
CESD	0.0134	2061	-0.113	No	-0.099	1,053	-0.333***	Yes	0.146	1,008	0.098	No
			[0.089]				[0.108]				[0.155]	

Note: Probit (OLS for CESD) estimates. Controls excluded from table: cubic in log family income and family size at age 4, an interaction between these two variables, cubic in log of average family income and family size for ages 0-2, interaction between these two variables, cubic on child's birth weight, dummy for the presence of a father figure in the household at age 4, race and age dummies, and year and state of residence at age 4 effects. "Control Mean" is the mean outcome among observations just above the cutoff (at most 25% above the cutoff). Marginal effects for discrete outcomes in italic. Robust standard errors are reported in brackets clustered at state-year at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.13: Reduced Form Estimates: Young Adults.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Control Mean	N	All ITT	RW pv<0.1	Control Mean	N	Males ITT	RW pv<0.1	Control Mean	N	Girls ITT	RW pv<0.1
In High School Diploma	0.716	1,965	0.088 [0.123] 0.028	No	0.669	943	0.282 [0.189] 0.090	No	0.763	1,022	-0.082 [0.189] -0.023	No
<i>marginal effect</i>												
Birth Control	0.550	1,639	0.087 [0.125] 0.033	No	0.567	765	0.125 [0.197] 0.045	No	0.536	874	0.065 [0.186] 0.024	No
<i>marginal effect</i>												
Ever Sentenced	0.212	1,904	-0.239 [0.154] -0.060	No	0.284	943	-0.402** [0.200] -0.114	Yes	0.138	961	-0.286 [0.228] -0.054	No
<i>marginal effect</i>												
Idle	0.102	1,854	-0.146 [0.178] -0.026	No	0.112	874	-0.525** [0.250] -0.090	Yes	0.092	980	0.044 [0.255] 0.007	No
<i>marginal effect</i>												
Ever in College	0.536	1,969	-0.074 [0.125] -0.027	No	0.462	948	-0.052 [0.175] -0.018	No	0.607	1,021	-0.076 [0.187] -0.027	No
<i>marginal effect</i>												
Ever worked	0.309	1,931	0.007 [0.145] 0.002	No	0.229	934	0.192 [0.206] 0.060	No	0.389	997	-0.169 [0.184] -0.057	No
<i>marginal effect</i>												

Note: Probit estimates. Controls excluded from table: cubic in log family income and family size at age 4, an interaction between these two variables, cubic in log of average family income and family size for ages 0-2, interaction between these two variables, cubic on child's birth weight, dummy for the presence of a father figure in the household at age 4, race and age dummies, and year and state of residence at age 4 effects. "Control Mean" is the mean outcome among observations just above the cutoff (at most 25% above the cutoff). Marginal effects for discrete outcomes in *italic*. Robust standard errors are reported in brackets clustered at state-year at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.14: Mother Fixed Effects

Year of birth	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1972-1986	1986-2000	1972-1986	1986-2000	1972-1986	1986-2000	1972-1986	1986-2000
	Test Score				PIAT-RR			
	PPVT				BPI			
Head Start	0.134 [0.087]	0.132 [0.122]	0.215* [0.126]	-0.058 [0.204]	0.200** [0.095]	0.193 [0.121]	0.033 [0.083]	0.036 [0.091]
Head StartXAge 7-10	-0.096 [0.081]	-0.193** [0.092]	-0.169 [0.128]	-0.188 [0.204]	-0.117 [0.090]	-0.209* [0.108]	0.007 [0.077]	-0.035 [0.089]
Head StartXAge 11-14	-0.085 [0.085]	-0.183* [0.100]	-0.085 [0.121]	-0.052 [0.201]	-0.197** [0.095]	-0.164 [0.119]	-0.060 [0.087]	-0.088 [0.096]
Preschool	0.155 [0.142]	-0.203 [0.142]	0.095 [0.195]	0.176 [0.260]	0.013 [0.146]	-0.114 [0.146]	0.022 [0.133]	0.055 [0.119]
PreschoolXAge 7-10	-0.233* [0.139]	0.174 [0.114]	-0.097 [0.202]	-0.378 [0.293]	-0.230* [0.134]	0.061 [0.113]	-0.063 [0.130]	0.017 [0.094]
PreschoolXAge 11-14	-0.276* [0.146]	0.195 [0.122]	-0.046 [0.185]	-0.184 [0.259]	-0.213 [0.144]	0.016 [0.129]	-0.070 [0.128]	0.031 [0.122]
Observations	5,276	3,228	2,804	1,311	5,008	3,085	4,944	3,051
control mean	-0.440	-0.285	-1.147	-1.077	-0.125	0.0109	0.679	0.388
SD	0.881	0.942	1.179	1.361	0.955	0.967	0.997	1.074
P-Value Test: HS 7-10	0.487	0.526	0.590	0.266	0.193	0.885	0.454	0.986
P-Value Test: HS 11-14	0.373	0.605	0.102	0.559	0.956	0.792	0.646	0.555

Note: Controls excluded from table include child's gender, first born status, and age at test and year fixed effects and mother fixed effects.

We restrict the sample to children who were over four years old by 1990 (in the first column for each outcome) and to children born after 1986 (on the second column for each outcome). We further restrict the sample to children to whom it is possible to recover information for pre-Head Start age (in particular, on birth weight and on whether a child was breastfed). Finally, we restrict the sample to families with at least two age-eligible children. The sample used in includes only children aged 5-14 years old. The unit of observation is child-by-age.

Robust standard errors are reported in brackets clustered at family level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.15: Reduced Form Estimates: Other specifications (only boys).

	(1)	(2)	(3)	(4)	(5)
	Basic	No-Pre HS age controls	All controls	Quadratic	Quartic
Panel A: Ages 12-13					
A.1: BPI					
1[HS Eligible at 4]	-0.274** [0.125]	-0.250** [0.125]	-0.252* [0.141]	-0.261** [0.124]	-0.312** [0.127]
A.2: Overweight					
1[HS Eligible at 4]	-0.379** [0.165]	-0.380** [0.159]	-0.488** [0.204]	-0.358** [0.160]	-0.398** [0.170]
Marginal Effect	-0.0950	-0.0966	-0.115	-0.0904	-0.0997
A.3: Health condition that requires use of special equipment					
1[HS Eligible at 4]	-0.777*** [0.287]	-0.632** [0.275]	-1.001*** [0.336]	-0.706*** [0.260]	-0.779*** [0.281]
Marginal Effect	-0.101	-0.0839	-0.139	-0.0927	-0.103
Panel B: Ages 16-17					
B.1: Overweight					
1[HS Eligible at 4]	-0.471** [0.191]	-0.493*** [0.187]	-0.412* [0.216]	-0.483** [0.190]	-0.470** [0.194]
Marginal Effect	-0.118	-0.126	-0.102	-0.122	-0.118
B.2: CESD					
1[HS Eligible at 4]	-0.333*** [0.108]	-0.329*** [0.108]	-0.348*** [0.120]	-0.327*** [0.106]	-0.325*** [0.112]
Panel C: Ages 20-21					
Ever Sentenced					
1[HS Eligible at 4]	-0.402** [0.200]	-0.312 [0.199]	-0.477** [0.229]	-0.403** [0.198]	-0.443** [0.208]
Marginal Effect	-0.114	-0.0928	-0.133	-0.115	-0.126
C.2: Idle					
1[HS Eligible at 4]	-0.525** [0.250]	-0.486** [0.247]	-0.576** [0.276]	-0.513** [0.242]	-0.454* [0.261]
Marginal Effect	-0.090	-0.084	-0.097	-0.088	-0.076

Note: "Basic" is specification used throughout the paper (cubic in log family income and family size at age 4, an interaction between these two variables, a dummy for the presence of a father figure in the child's household at age 4, cubic in average log family income and average family size between ages 0 and 2, an interaction between the two, and cubic in birth weight, race and age dummies and dummies for year and state of residence at age 4). "No pre-Head Start age controls" includes the same controls than in column (1), except those measured before age 3. "All Controls" includes the same controls than in column (1) and dummies for highest grade completed by mother before child turned 3, maternal AFQT score, maternal grandmother's highest grade completed and indicators of maternal situation at 14 years old (whether the mother lived in a Southern, whether she lived with parents and whether she lived in a rural area). "Quadratic" and "Quartic" are the same specification as "Basic" but using polynomials up to the second and fourth order, respectively, in (log) income, family size and birth weight variables. Robust standard errors in brackets clustered at state-year at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.16: Reduced Form Estimates: Trimming around cutoff (only boys).

	(1)	(2)	(3)	(4)	(5)	(6)
	[75%-125%]	[50%-150%]	[25%-175%]	[15%-185%]	300%	Full Sample
Panel A: Ages 12-13						
A.1: BPI						
1[HS Eligible at 4]	0.123	-0.168	-0.323**	-0.274**	-0.169*	-0.0693
	[0.205]	[0.153]	[0.134]	[0.125]	[0.0936]	[0.0818]
Observations	359	774	1,110	1,211	1,793	2,327
A.2: Overweight						
1[HS Eligible at 4]	-1.008***	-0.461**	-0.399**	-0.379**	-0.0147	-0.0924
	[0.368]	[0.204]	[0.167]	[0.165]	[0.135]	[0.112]
Observations	300	772	1,128	1,242	1,852	2,392
A.3: Health condition that requires use of special equipment						
1[HS Eligible at 4]	-0.506	-0.698**	-0.779***	-0.777***	-0.246	-0.244
	[0.499]	[0.352]	[0.293]	[0.287]	[0.195]	[0.166]
Observations	244	588	985	1,111	1,811	2,360
Panel B: Ages 16-17						
B.2: Overweight						
1[HS Eligible at 4]	-0.760**	-0.577**	-0.445**	-0.471**	-0.377***	-0.360***
	[0.337]	[0.239]	[0.195]	[0.191]	[0.145]	[0.131]
Observations	303	724	1,076	1,167	1,755	2,248
B.2: CESD						
1[HS Eligible at 4]	-0.453**	-0.249**	-0.303***	-0.333***	-0.178**	-0.0919
	[0.208]	[0.124]	[0.111]	[0.109]	[0.0825]	[0.0683]
Observations	315	671	973	1,054	1,534	1,960
Panel C: Ages 20-21						
Ever Sentenced						
1[HS Eligible at 4]	-0.0346	-0.574**	-0.457**	-0.402**	-0.0568	-0.0574
	[0.396]	[0.244]	[0.213]	[0.200]	[0.146]	[0.127]
Observations	256	590	867	943	1,378	1,694
C.2: Idle						
1[HS Eligible at 4]	-3.337***	-0.535*	-0.711**	-0.525**	-0.264	-0.262
	[1.173]	[0.284]	[0.277]	[0.250]	[0.212]	[0.187]
Observations	173	503	806	874	1,246	1,529

Note: Estimates using the same specification as table (4), but different trimming of data around the cutoff. The sample includes only males. Robust standard errors are reported in brackets clustered at state-year at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.17: Income eligibility to Head Start at ages 0-7 (only boys).

Eligible at age	(1) 0	(2) 1	(3) 2	(4) 3	(5) 4	(6) 5	(7) 6	(8) 7
Panel A: Ages 12-13								
A.1: BPI								
1[HS Eligible at 4]	-0.000322 [0.142]	-0.0768 [0.146]	0.128 [0.139]	0.167 [0.136]	-0.274** [0.125]	-0.236* [0.136]	0.0533 [0.123]	-0.0978 [0.141]
Observations	984	1,087	1,172	1,110	1,211	1,012	1,093	983
A.2: Overweight								
1[HS Eligible at 4]	0.0365 [0.214]	-0.129 [0.195]	0.0649 [0.183]	-0.0200 [0.198]	-0.379** [0.165]	-0.294 [0.201]	-0.112 [0.191]	0.143 [0.218]
Observations	946	1,092	1,211	1,126	1,242	1,029	1,130	970
A.3: Health condition that requires use of special equipment								
1[HS Eligible at 4]	0.539** [0.271]	0.845*** [0.280]	-0.349 [0.255]	0.0767 [0.298]	-0.777*** [0.287]	-0.426 [0.304]	-0.000698 [0.231]	-0.744*** [0.263]
Observations	848	910	1,123	1,041	1,111	863	1,039	866
Panel B: Ages 16-17								
B.1: Overweight								
1[HS Eligible at 4]	-0.303 [0.224]	-0.159 [0.203]	-0.0379 [0.198]	-0.130 [0.212]	-0.471** [0.191]	-0.471** [0.194]	0.0993 [0.212]	-0.0337 [0.211]
Observations	912	1,045	1,089	1,090	1,167	1,019	1,045	942
B.2: CESD								
1[HS Eligible at 4]	0.0463 [0.121]	0.104 [0.111]	-0.00376 [0.133]	-0.0413 [0.120]	-0.333*** [0.108]	0.0333 [0.108]	-0.103 [0.124]	-0.0535 [0.130]
Observations	831	974	1,009	990	1,053	923	939	840
Panel C: Ages 20-21								
Ever Sentenced								
1[HS Eligible at 4]	0.115 [0.212]	0.373* [0.206]	0.286 [0.197]	-0.0419 [0.204]	-0.402** [0.200]	-0.236 [0.188]	0.317* [0.187]	0.0761 [0.225]
Observations	767	871	913	914	943	926	931	833
C.2: Idle								
1[HS Eligible at X]	-0.0939 [0.284]	-0.327 [0.286]	0.817** [0.325]	-0.0369 [0.317]	-0.525** [0.250]	-0.0229 [0.232]	0.179 [0.259]	0.503** [0.248]
Observations	691	793	834	829	874	864	864	775

Note: Table includes reduced form results of table (4) with income eligibility measured at different ages between 0 and 7. The sample includes only males. Robust standard errors are reported in brackets clustered at state-year at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.19: Reduced Form Estimates: Ages 12-13 (effects by year of birth; only boys).

Year of birth	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	1977-1984	1985-2000	1977-1984	1985-2000	1977-1984	1985-2000	1977-1984	1985-2000
	Panel A: Indexes							
	Cognitive		Only behaviors		Health		Global	
I[HS Eligible at 4]	-0.145 [0.0959]	-0.0531 [0.128]	0.258** [0.108]	0.0142 [0.143]	0.184* [0.103]	0.154 [0.142]	0.293*** [0.0983]	0.109 [0.139]
Observations	1,394	989	1,501	1,052	1,501	1,052	1,501	1,052
	Panel B: Individual Variables							
	Health condition requires special equipment		Overweight		Special Education		BPI	
I[HS Eligible at 4]	-0.823*** [0.265]	-0.322 [0.292]	-0.317** [0.160]	-0.294 [0.195]	-0.290* [0.160]	-0.0161 [0.185]	-0.0395 [0.106]	-0.126 [0.143]
Observations	1,289	826	1,434	1,007	1,421	1,029	1,419	1,007
Control Mean	0.0640	0.0787	0.169	0.217	0.177	0.263	0.676	0.456
Marginal Effect	-0.090	-0.040	-0.076	-0.082	-0.076	-0.005		

Note: Probit (and OLS for BPI, index) estimates. Sample: Males. Controls excluded from table include cubic in log family income and family size at age 4, an interaction between these two variables, cubic in log of average family income and family size for ages 0-2, the interaction between these two variables, cubic on child's birthweight, a dummy indicating the presence of a father figure in the household at age 4, race and age dummies, and dummies for year and state of residence at age 4. Marginal effect is average marginal change in the probability of Head Start participation across individuals as the eligibility status changes and all other controls are kept constant. Control Mean is the mean outcome among observations just above the cutoff, that is, at 25% above the cutoff. Robust standard errors are reported in brackets clustered at state-year at age four level. *, **, *** significant at 10%, 5% and 1%, respectively.

Table B.20: Labor market outcomes of mother and her spouse.

	(1)	(2)	(3)	(4)	(5)	(6)
Age Variable	Prior to eligibility Hours per week mother spouse		At eligibility Hours per week mother spouse		At HS age Hours per week mother spouse	
	Panel A: All sample					
Eligible 4	-1.746 [1.335]	-0.479 [1.251]	-4.440*** [1.227]	-1.285 [1.368]	-4.000*** [1.317]	-1.176 [1.321]
Observations	2,736	1,483	2,864	1,507	2,855	1,496
Control Mean	27.25	43.36	27.57	43.30	27.44	42.97
	Panel B: Males					
Eligible 4	-1.474 [1.734]	-1.747 [1.717]	-5.265*** [1.655]	-3.371 [2.060]	-5.082*** [1.755]	-1.282 [2.006]
Observations	1,388	757	1,448	772	1,446	762
Control Mean	26.43	42.24	26.95	43.49	26.73	43.52
	Panel C: Females					
Eligible 4	-2.069 [2.061]	-0.078 [1.776]	-3.644** [1.843]	0.483 [1.832]	-3.178 [1.951]	-1.083 [1.910]
Observations	1,348	726	1,416	735	1,409	734
Control Mean	28.07	44.57	28.18	43.09	28.14	42.41

Note: The age groups per column are as follow: "Prior to eligibility" includes labor supply of parents when the child is 0-1 years old; "At eligibility" includes labor supply of parents when child is 3-4 years old; and "At HS age" includes labor supply of parents when the child is 4-5 years old. Notice that eligibility at age four is determined using income from the calendar year prior to that in which the child turn 4 (and, therefore, incorporates parental labor market outcomes also in the previous calendar year).

Controls excluded from table include cubic in log family income and family size at age 4, an interaction between these two variables, dummy for the presence of a father figure in the household at age 4, race, gender and age dummies, and year and state of residence at age 4 effects.

"Control Mean" is the mean outcome among observations just above the cutoff (at most 25% above the cutoff). * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.21: Parental investments in children: HOME score.

	(1)	(2)	(3)
Sample	All	Males	Females
Eligible 4	-0.062 [0.061]	-0.114 [0.085]	0.014 [0.089]
Eligible 4XAge6-9	0.001 [0.040]	0.016 [0.060]	-0.020 [0.057]
Eligible 4XAge10-14	0.096** [0.045]	0.135** [0.064]	0.051 [0.063]
Observations	14,642	7,392	7,250
Control Mean	-0.487	-0.548	-0.427
P-Values			
Joint test on "Eligible 4"	0.021	0.047	0.398
Effect ages 6-9	0.294	0.212	0.944
Effect ages 10-14	0.560	0.786	0.442

Note: The dependent variable is the HOME score. Controls excluded from table include cubic in log family income and family size at age 4, an interaction between these two variables, dummy for the presence of a father figure in the household at age 4, race, gender and age dummies, and year and state of residence at age 4 effects and indicators for age group (6-9 and 10-14; 4-5 is the omitted category).

"Control Mean" is the mean outcome among observations just above the cutoff (at most 25% above the cutoff). The unit of observation is child-by-age. Robust standard errors are reported in brackets clustered at state (at age 4) level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.22: Reduced Form Estimates: Functions of distance to cutoff as running variable.

	(1)	(2)
	Quadratic	Distance to cutoff Cubic
Panel A: Ages 12-13		
A.1: Index		
1[HS Eligible at 4]	0.104 [0.109]	0.143 [0.126]
A.2: BPI		
1[HS Eligible at 4]	-0.117 [0.106]	-0.150 [0.126]
A.3: Overweight		
1[HS Eligible at 4]	-0.149 [0.145]	-0.155 [0.173]
Marginal Effect	-0.0379	-0.0395
A.4: Health condition that requires use of special equipment		
1[HS Eligible at 4]	-0.319 [0.208]	-0.315 [0.244]
Marginal Effect	-0.0400	-0.0395
Panel B: Ages 16-17		
B.1: Index		
1[HS Eligible at 4]	0.136 [0.112]	0.280** [0.133]
B.2: Overweight		
1[HS Eligible at 4]	-0.329** [0.166]	-0.390** [0.188]
Marginal Effect	-0.0838	-0.0997
B.3: CESD		
1[HS Eligible at 4]	-0.270*** [0.0921]	-0.197* [0.114]
Panel C: Ages 20-21		
C.1: Index		
1[HS Eligible at 4]	0.096 [0.128]	0.158 [0.138]
C.2: Ever Sentenced		
1[HS Eligible at 4]	-0.267 [0.182]	-0.465** [0.200]
Marginal Effect	-0.0807	-0.139

Note: Table includes reduced form estimates for selected outcomes using several specifications. The sample includes only males.

Controls excluded from the table include polynomials of the difference between family income and cutoff instead of polynomials in log income and family size at age 4, a dummy for the presence of a father figure in the child's household at age 4, race and age dummies and dummies for year and state of residence at age 4. Robust standard errors in brackets clustered at state-year at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.23: Reduced Form estimates using fake income cutoff.

	(1)	(2)	(3)	(4)
Cutoff at	0.5	-0.5	0,0.5	-0.5,0
Panel A: Ages 12-13				
A.1: BPI				
Fake cutoff	-0.009 [0.131]	0.332*** [0.122]	-0.074 [0.139]	0.284** [0.126]
1[HS Eligible at 4]			-0.294** [0.133]	-0.212* [0.128]
A.2: Overweight				
Fake cutoff	0.020 [0.177]	0.047 [0.176]	-0.082 [0.180]	-0.044 [0.184]
1[HS Eligible at 4]			-0.426** [0.169]	-0.415** [0.172]
A.3: Special Equipment				
Fake cutoff	-0.275 [0.262]	0.260 [0.235]	-0.448 [0.278]	0.090 [0.240]
1[HS Eligible at 4]			-0.880*** [0.325]	-0.753*** [0.288]
A.4: Global Index				
Fake cutoff	0.041 [0.118]	-0.197 [0.136]	0.116 [0.120]	-0.131 [0.138]
1[HS Eligible at 4]			0.346*** [0.131]	0.289** [0.129]
Panel B: Ages 16-17				
B.1: Overweight				
Fake cutoff	-0.251 [0.187]	0.153 [0.179]	-0.366* [0.196]	0.042 [0.191]
1[HS Eligible at 4]			-0.588*** [0.205]	-0.495** [0.201]
B.2: CESD				
Fake cutoff	0.134 [0.096]	0.117 [0.117]	0.073 [0.094]	0.037 [0.120]
1[HS Eligible at 4]			-0.316*** [0.108]	-0.322*** [0.111]
B.3: Global Index				
Fake cutoff	-0.044 [0.107]	-0.273** [0.139]	0.005 [0.109]	-0.225 [0.142]
1[HS Eligible at 4]			0.261** [0.123]	0.207* [0.123]
Panel C: Ages 20-21				
Ever Sentenced				
Fake cutoff	-0.060 [0.218]	0.112 [0.191]	-0.151 [0.222]	0.038 [0.196]
1[HS Eligible at 4]			-0.412** [0.204]	-0.369* [0.203]
C.2: Global Index				
Fake cutoff	0.161 [0.138]	-0.200 [0.139]	0.208 [0.139]	-0.172 [0.141]
1[HS Eligible at 4]			0.217 [0.139]	0.132 [0.142]

Note: Estimates using the same specification as table (4), but different cutoffs are set at the relative distance indicated in each column. The sample includes only males. Robust standard errors are reported in brackets clustered at state-year at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.24: Reduced Form clustering SE by state of residence at age 4.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Age groups			12-13			16-17			20-21	
Variables	Index	Overweight	Use of Special Equipment	BPI	Index	Overweight	CESD	Index	Ever Sentenced	Idle
I[HS Eligible at 4]	0.313*** [0.103]	-0.379*** [0.127]	-0.777*** [0.306]	-0.274** [0.120]	0.266** [0.118]	-0.471** [0.213]	-0.333*** [0.104]	0.194 [0.123]	-0.402* [0.209]	-0.525* [0.291]
Marginal Effect		-0.095	-0.101			-0.118			-0.114	-0.090

Note: Estimates using the same specification as table (4). The sample includes only males. Robust standard errors are reported in brackets clustered by state at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.25: Reduced Form accounting for attrition: Children surveyed at ages 12-13, 16-17 and 20-21.

Age groups Variables	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	Index	Overweight	Use of Special Equipment	BPI	Index	Overweight	CESD	Index	Ever Sentenced	Idle
1[HS Eligible at 4]	0.389** [0.167]	-0.309 [0.228]	-0.487 [0.397]	-0.235 [0.165]	0.370** [0.151]	-0.304 [0.244]	-0.409*** [0.149]	0.202 [0.146]	-0.552*** [0.214]	-0.399 [0.293]
Observations	776	724	588	713	776	710	622	776	768	713
Control Mean	-0.360	0.206	0.0899	0.721	-0.275	0.182	-0.127	-0.278	0.303	0.097
Marginal Effect		-0.0747	-0.0592			-0.071			-0.156	-0.066

Note: This table presents estimates for γ in equation 1. Controls excluded from table: cubic in log family income and family size at age 4, an interaction between these two variables, cubic in log of average family income and family size for ages 0-2, interaction between these two variables, cubic on child's birth weight, dummy for the presence of a father figure in the household at age 4, race and age dummies, and year and state of residence at age 4 effects. "Control Mean" is the mean outcome among observations just above the cutoff (at most 25% above the cutoff). Marginal effects for discrete outcomes in italic.

Sample restricted to children used in to estimate effects of Head Start at ages 12-13, 16-17 and 20-21. The sample includes only males.

Robust standard errors are reported in brackets clustered at state-year at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table B.26: Linear Probability Model: Two Stage Least Squares Estimates.

	(1)	(2)	(3)	(4)	(5)
	Ages 12-13		Ages 16-17	Ages 20-21	
	Overweight	Special Equipment	Overweight	Ever Sentenced	Idle
Head Start	-0.357** [0.169]	-0.322** [0.138]	-0.432** [0.205]	-0.149 [0.191]	-0.099 [0.142]
Observations	1,242	1,111	1,167	943	874

Note: Participation in program is instrumented with eligibility status at age four. Controls excluded from table include cubic in log family income and family size at age 4, an interaction between these two variables, cubic in log of average family income and family size for ages 0-2, the interaction between these two variables, cubic on child's birth weight, a dummy indicating the presence of a father figure in the household at age 4, race and age dummies, and dummies for year and state of residence at age 4. The standard errors are obtained by block-bootstrap (500 replications; block is the cell year-state of residence at age 4). * significant at 10%; ** significant at 5%; *** significant at 1%.

C Figures

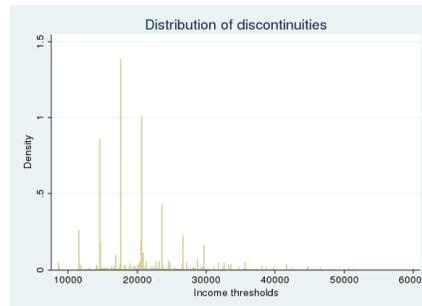


Figure C.1: Distribution of Income thresholds at age 4.

Note: Figure includes all children used in the regressions whose family income at age 4 was 15-185% of the discontinuity level of income.

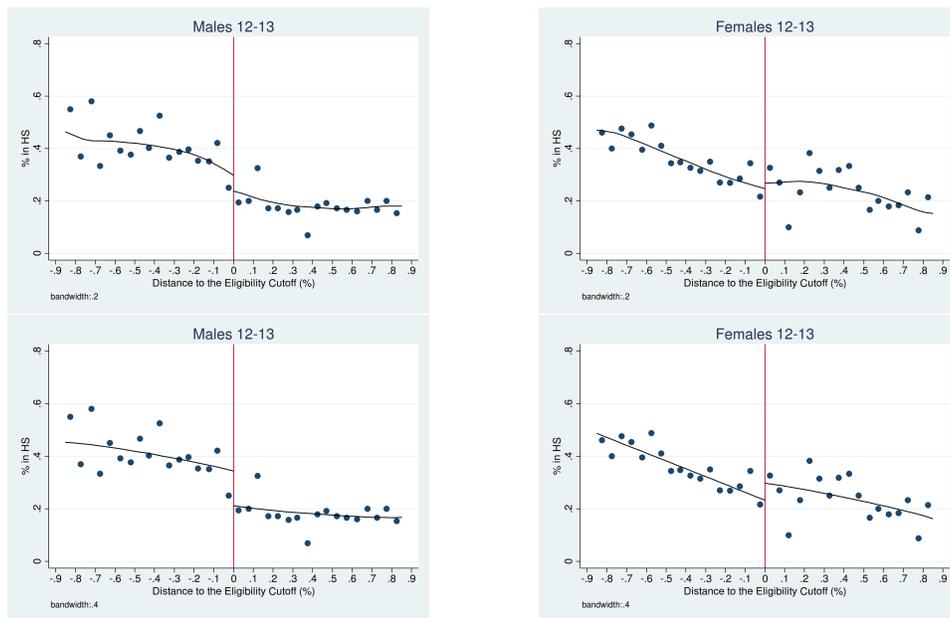


Figure C.2: Proportion of children in Head Start, by eligibility status.

Note: The continuous lines in figure are local linear regression estimates of HS participation on percentage distance to cutoff. The bandwidth is 0.2 in the first two graphs and 0.4 in the bottom two. Circles in figures represent mean participation by cell within intervals of 0.05 of distance to cutoff. The kernel used was Epanechnikov.

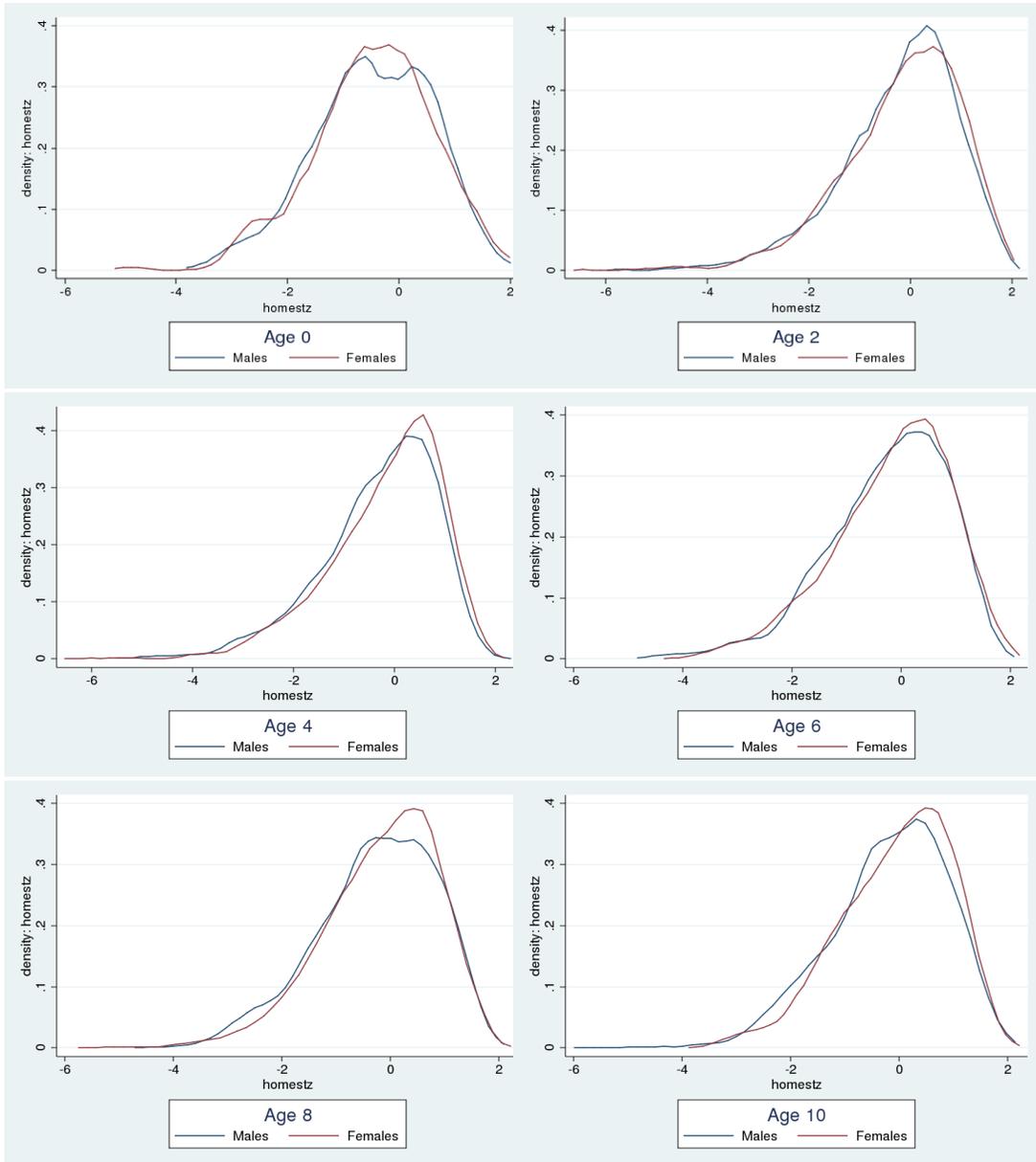


Figure C.3: Density estimates of maternal investments by gender.

Note: The graphs density estimates for HOME score by gender for several ages. For each age we perform the KolmogorovSmirnov test for the equality of the distributions for the score of the two genders. The p-values are the following: 0.339 (age 0), 0.067 (age 2), 0.000 (age 4), 0.157 (age 6), 0.017 (age 8) and 0.002 (age 10).



Figure C.4: Average outcomes by eligibility status: Falsification, Bandwidth = 0.3. Note: The continuous lines in figure present local linear regression estimates of several outcomes on percentage distance to cutoff. The graphs in the first two rows present estimates for males and the graphs in the bottom two rows present estimates for females. The kernel used was Epanechnikov.

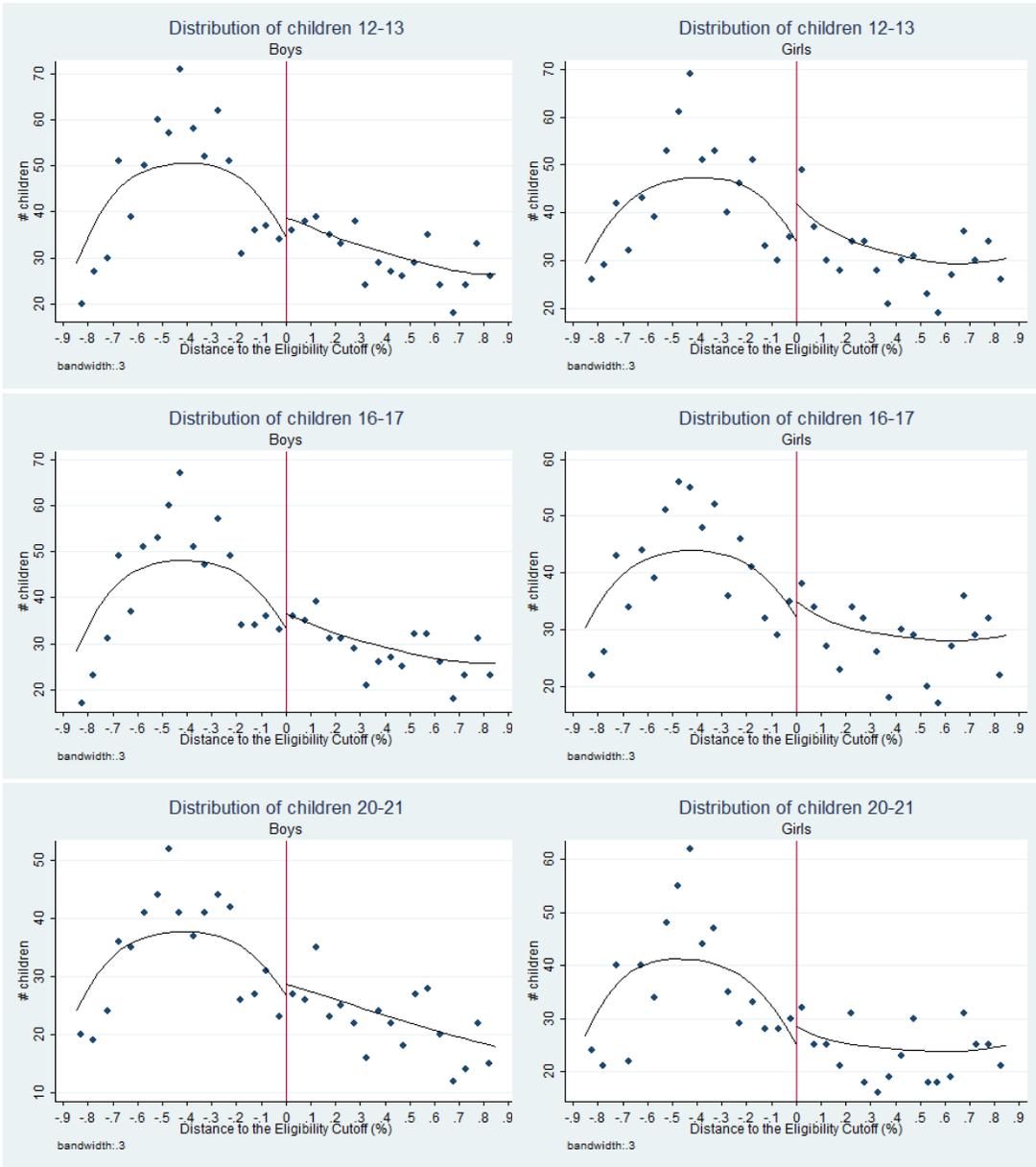


Figure C.5: Distribution of children around the cutoff.

Note: The continuous lines in figure are local linear regression estimates of the number of children per interval of width 0.05 of relative income-distance to cutoff; regressions were run separately on both sides of the cutoff and the bandwidth was set to 0.3. Circles in figures represent mean number of children by cell within intervals of 0.05 of distance to cutoff. The kernel used was Epanechnikov.

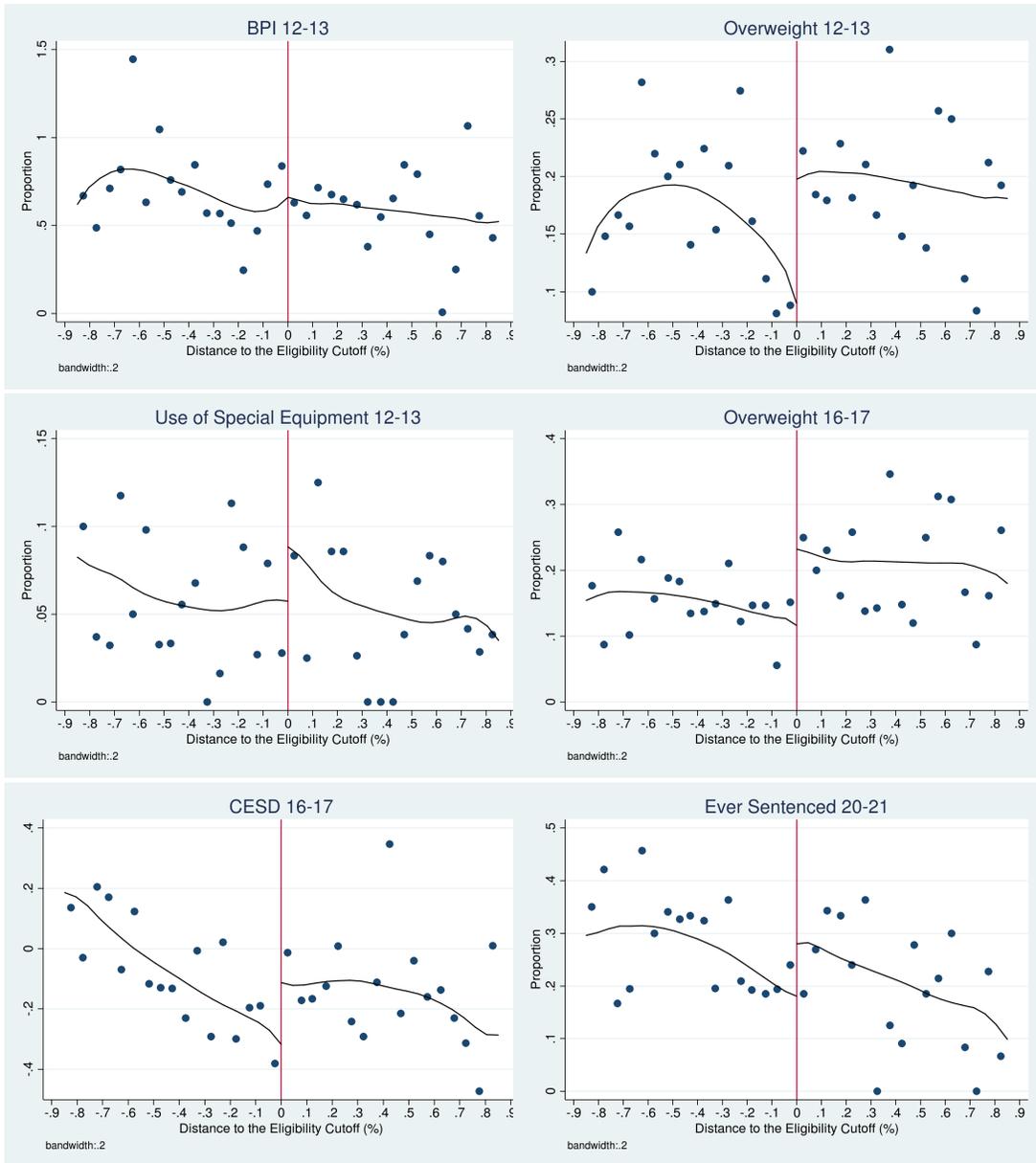


Figure C.6: Average outcomes by eligibility status, Bandwidth = 0.2.

Note: The continuous lines in figure present local linear regression estimates of several outcomes on percentage distance to cutoff. This figure replicates estimates of figure 2, using a bandwidth of 0.2. Circles in figures represent the mean outcome by cell within intervals of 0.05 of distance to cutoff. The kernel used was Epanechnikov. The sample only includes boys.

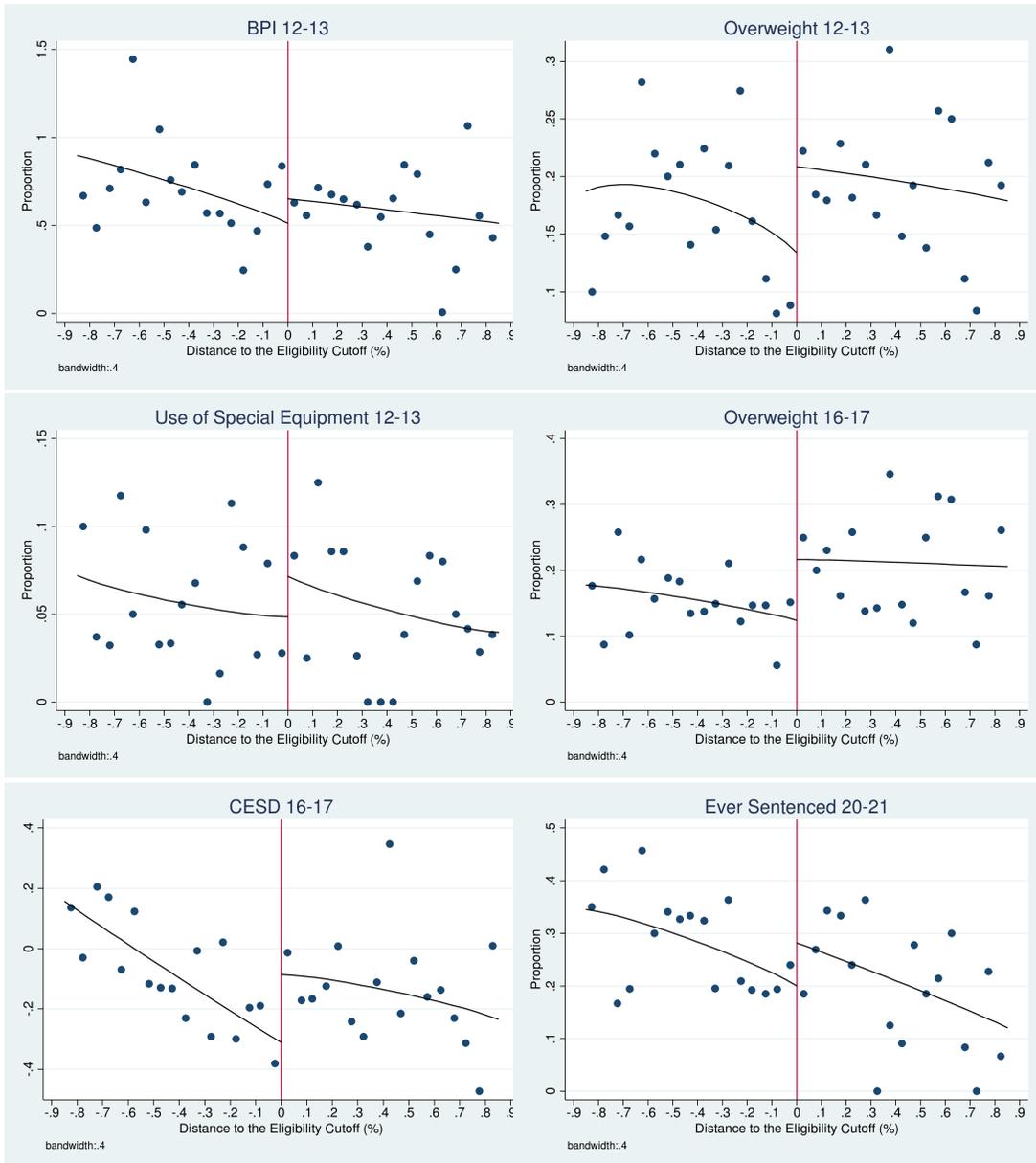


Figure C.7: Average outcomes by eligibility status, Bandwidth = 0.4.

Note: The continuous lines in Figure present local linear regression estimates of several outcomes on percentage distance to cutoff. This figure replicates estimates of figure 2, using a bandwidth of 0.4. Circles in figures represent the mean outcome by cell within intervals of 0.05 of distance to cutoff. The kernel used was Epanechnikov. The sample only includes boys.

D Data Construction

This Appendix includes the details about the construction of the sample used.

Table D.1: Sample selection in the CNSLY.

	Number of children	Dropped
Original sample	11,495	
Dropped:		
No information on HS participation	9,210	2,285
Missing information: income, family size and state of residence at age 4	7,236	1,974
Missing information: income, family size 0-2 and birth weight	6,381	855
Children observed at ages 12-13 or 16-17 or 20-21 to whom at least one outcome is observed	5,433	948
Total sample	5,433	
Sample around cutoff at age 4 (15%-185%)	2,875	
After dropping children not used in first stage estimation due to multicollinearity	2,833	
By age		
12-13	2,550	
16-17	2,416	
20-21	1,977	
At all ages	1,595	

Of the 2285 children dropped from the sample because of lack Head Start information, (i) 507 are dropped because the mother was not interviewed after 1986 (of these 398 are children of the mothers in the military subsample which was discontinued in 1985), (ii) 617 are children not interviewed after 1986 (of these 80 died before the CNLSY began), (iii) 362 children were interviewed after 1986 but the last time they were interviewed they were too young (less than 3 years old) for the Head Start question to be applied, (iv) 630 children are dropped for the sample simply because they have missing information in Head Start variable in the raw data, and (v) finally, 169 children are not included in the sample because despite having information about Head Start participation, they do not have information about the time spent in program.

Table D.2: Missing information before age 3.

	(1)	(2)
Sample	Male	Female
1[Eligible at 4]	0.022 [0.241]	0.187 [0.227]
Observations	1,203	1,160
Control Mean	0.035	0.079
SD	0.185	0.270
Marginal Effect	0.003	0.022

Note: The table reports results of probit estimates of an indicator for whether the child has missing information on any of the pre-age 3 controls that we add to specification of table 4, that is, log of average family income and family size for ages 0-2 or child's birth weight. The controls excluded from the table are: cubic in log family income and family size at age 4, an interaction between these two variables, a dummy indicating the presence of a father figure in the household at age 4, race and age dummies, and dummies for year and state of residence at age 4.

The sample used in estimation includes only children ages 12-13. Robust standard errors are reported in brackets clustered at state-year at age four level. * significant at 10%; ** significant at 5%; *** significant at 1%.

E Cost-benefit Analysis

We here the details of our simple cost-benefit calculation. As we showed in Table 1, eligibility to Head Start is associated with an increase in the probability of participation at the margin of 19-22% among males (we use 21% for the calculations presented here). Thus, when comparing costs and benefits we take into account that eligibility does not imply participation and vice-versa. The calculations are simpler to present if we focus on a single cohort of participants, and we choose the cohort of children entering Head Start in 1992 (born in 1988), as it is the most contemporaneous year that allows an analysis to all three age groups we studied. As a comparison, we also consider the cost-effectiveness of the version of the program available in 2003, computing the costs in that year but assuming the same expected benefits associated to the 1992-cohort.⁴⁸

We computed the cost of attendance per child in 1992 using the federal appropriation and the number of enrolled children obtained from the Head Start Fact Sheets. These figures were \$1,439,903,924 for the former and 621,078 for the later. We then discount the annual cost of attendance at age four to age 0 (using an interest rate of 4%). Since by expanding eligibility cutoffs \$1 boys will be 21% more likely to participate, the expected cost of participation is \$416.

Concerning the potential benefits, we start by focusing on health. We showed that Head Start is associated with improvements in mental health among adolescents, a reduction in the incidence of a chronic condition that requires the use of special equipment (a brace, crutches, a wheelchair, special shoes, a helmet, a special bed, a breathing mask, an air filter, or a catheter), and a reduction in the probability of being overweight. We provide an estimate for the lower bound of the benefits associated with an improvement in health conditions, by considering only the savings associated with reduction in obesity (for which we can get more reliable figures).

Childhood and adolescence obesity increases the risk of developing diseases such as high cholesterol, hypertension, respiratory ailments, orthopedic problems, depression, and type 2 diabetes. The cost of obesity is estimated to be US\$2741 per year by Cawley and Meyerhoefer, 2012 (their calculations use 2005 USD, which is equivalent to US\$3011 in 2009 USD). We assume an equal cost across all obese individuals. Since obese adolescents have 70 percent chance of becoming overweight or obese adults⁴⁹ we also estimate the present value of costs for a child who is obese throughout adolescence and young adulthood, in particular between ages 13 and 30. Therefore, even if we exclude the days lost of work and college due to conditions

⁴⁸All monetary values are measured in 2009 dollars.

⁴⁹See http://aspe.hhs.gov/health/reports/child_obesity/. Access on January 3, 2012.

related to obesity, the discounted value of direct health expenses with diseases associated to obesity between the ages of 13 and 30 adds up to \$23,808 (see Panel "Alternative 1" of Table E.1). Using the more conservative measure presented in the previous paragraph, the present value of savings associated with a reduced probability of being obese at just age 13 adds up to \$1808 (see Panel "Alternative 2" of Table E.1).

To compute the savings associated to reduction in criminality we use the direct expenditures in criminal justice for 2007, which are part of *Justice Expenditure and Employment Statistical Extracts for 2007*.⁵⁰ We use the direct expenditures per year for criminal corrections, which include correctional supervision (adults supervised in the community on probation or parole and those incarcerated in state or federal prisons and local jails). The total expenditure in correctional supervision adds up to 72 billion dollars (in 2009 USD). This figure ignores direct judicial and police expenditures, which together with correctional supervision add up to \$219 billion dollars.⁵¹ Note that we obtain the savings associated with reduction in criminal activity in just one year, in which the individual is taken to be at 20 years old.

To obtain the cost per criminal offence we use the total number of offenders supervised by correctional authorities in 2007: 7,267,500 offenders. The average cost by offender including only the costs associated with supervision is about \$9,867 per year (\$4,683 after discount to age 0 - see Panel "Alternative 2"). If we also include court and policy costs, then the average annual cost per offender reaches \$30,269 (\$14,364 after discounted for age 0). The marginal individual due to expanding eligibility by \$1 is 11% less likely to have been sentenced of any criminal charge by age 20, therefore the expected saving in justice functions due to reduced criminal activity ranges between \$515 and \$1580, depending on whether we use a narrower or wider definition of expenditures.

The net benefits of the policy are present in Table E.1. Even conservative cost-benefit calculation shows that expanding Head Start is cost effective. If one considers only the static returns on health (the reduction on the health costs associated with childhood obesity at age 13), relaxing the eligibility thresholds implies that the marginal child by age 13 will be 10% less likely to be obese, which represents an expected saving of \$181. By age 20 she will be 11% less likely to be under correctional supervision, and the expected discounted saving adds up to \$515. Thus,

⁵⁰This is a national level data set from which we obtain disaggregated expenditures according to different activities, thus we assume uniform costs across all types of criminal activities and locations. We choose to measure expenses associated with criminal activity in 2007 because we are interested in the analysis for the cohort of children born in 1988, who are 20 years old in 2008, and there is no data available for 2008.

⁵¹Our figures number ignore victims costs, which are accounted in the cost-benefit analysis of Perry Preschool Program performed in Heckman et al. (2009).

the program reaches the break-even point, even without accounting for its benefits on mental health, reduction in probability of presenting a chronic condition, and future days lost of work due to worse health (the present value of net savings add up to \$280). These calculations are presented in Panel "Alternative 2". When we account for the savings associated with (1) improvement in health through reducing the likelihood of obesity during adolescence and early adulthood and (2) for a broader measure of criminal expenses, which includes court and police protection costs, the present value of net savings associated to expanding eligibility to Head Start reaches \$3545 (see Panel "Alternative 1" in table E.1). These calculations suggest that the internal rate of return of the program is at least 4%.

The current version of Head Start costs about three times more per child than the version available in 1992 (see column for FY2003). Assuming that the pattern of benefits does not vary across the 1992 and the 2003 cohorts, the net benefits associated with Head Start through expanded eligibility add up to \$2832 under the broader definition of benefits (Panel "Alternative 1"). Under the more strict definition included in Panel "Alternative 2" the net benefits are negative and add up to \$433.

References

- [1] Cawley J and Meyerhoefer C., 2012, "The medical care costs of obesity: an instrumental variables approach", *Journal of Health Economics*, 2012, 31:219-30.

Table E.1: Cost-benefit analysis.

	FY 1992	FY 2003
COSTS		
Total appropriation	1,439,903,924	5,718,484,287
Nb children	621,078	909,608
Cost per child	2,318	6,287
Annual cost of attendance at age 4 (discounted for age 0, interest rate = 4%)	1981.77	5373.95
BENEFITS - Alternative 1		
1) CRIME		
Total cost of corrections in 2007 (see JEEE, 2007)	219,930,610,736	
Adults under correctional supervision 2007 (see BJS, 2010)	7,267,500	
Cost/correction	30,262	
Cost at age 20 discounted for age 0 (interest rate = 4%)	14,364	
2) OBESITY		
Per Capita Cost (Cawley and Meyerhoeferd, 2012)	3,011	
Direct costs of obesity between ages 13-30 (discounted for age 0)	23,808	
NET SAVINGS		
11% less likely to be sentenced at age 20-21	1580.01	
10% less likely to be obese at 13-30	2380.79	
21% more likely to participate in HS	416.17	1128.53
Total savings (benefits - cost)	3544.62	2832.27
BENEFITS - Alternative 2		
1) CRIME		
Total cost of corrections in 2007 (see JEEE, 2007)	71,709,452,250	
Adults under correctional supervision 2007 (see BJS, 2010)	7,267,500	
Cost/correction	9867.14	
Cost at age 20 discounted for age 0 (interest rate = 4%)	4683.36	
2) OBESITY		
Per Capita Cost (Cawley and Meyerhoeferd, 2012)	3,011	
Cost at age 13 discounted for age 0 (interest rate = 4%)	1,808	
NET SAVINGS		
11% less likely to be sentenced at age 20-21	515.17	
10% less likely to be obese at 12-13	180.83	
21% more likely to participate in HS	416.17	1128.53
Total savings (benefits - cost)	279.83	-432.53

Note: All monetary values are in 2009 dollars. Values discounted to age 0 are calculated using an interest rate of 4%. Costs related to criminal activity are obtained from the Justice Expenditure and Employment Extracts for 2007. The figures for "total justice system" include the costs with police protection, judicial and legal expenses and corrections. The number of adults under correctional supervision is obtained from U.S. Department of Justice, 2010.

F Eligibility to Head Start

According to the Head Start Act, Sec. 645(a)(2)(A) "children from low-income families shall be eligible for participation in programs assisted under this subchapter (*Head Start*) if their families' incomes are below the poverty line, or if their families are eligible or, in the absence of child care, would potentially be eligible for public assistance"⁵². Alternatively, grantees may enroll up to 10% of children from "over-income" families.⁵³ See table B.1 in Appendix B for a summary of Head Start's legislation since the program was launched in 1965. The eligibility criteria have been unchanged throughout the period of analysis (See www.eric.ed.gov and Zigler and Valentine, 1979.)

A low-income family is a family whose income before taxes is below the poverty line or a family that is receiving public assistance, even if the family's income exceeds the poverty line. The U.S. Department of Health and Human Services considers public assistance as AFDC/TANF and SSI (see 45 CFR Part 1305.2). In section 3 of the main text we explain why we did not impute SSI eligibility.

The income period of time to be considered for eligibility is the 12 months immediately preceding the month in which application or reapplication for enrollment of a child in a Head Start program is made, or the calendar year immediately preceding the calendar year in which the application or reapplication is made. We use income from the last calendar year since it is the income measure available in NLSY79.

The relevant income measure includes (1) money wages or salary before deductions; (2) net income from self-employment; (3) payments from Social Security or railroad retirement; (4) unemployment compensation, strike benefits, workers' compensation, veterans benefits, public assistance (AFDC/TANF, SSI, Emergency Assistance money payments, and non-Federally funded General Assistance or General Relief money payments); (5) training stipends; (6) alimony, child support, and military family allotments or other regular support from an absent family member or someone not living in the household; (7) private pensions, government employee pensions, and regular insurance or annuity payments; (8) college or university scholarships, grants, fellowships, and assistantships; (9) dividends, interest, net rental income, net royalties, and periodic receipts from estates or trusts; (10) net gambling or lottery winnings. Relatively to this definition, the main measure of

⁵²See Title VI, Subtitle A, Chapter 8, Subchapter B, of the Omnibus Budget Reconciliation Act of 1981, Public Law 97-35 (42 USC 9840) and its amends (http://www.law.cornell.edu/uscode/html/uscode42/usc_sec_42_00009840---000-.html).

⁵³Indian Tribes meeting the conditions specified in 45 CFR 1305.4(b)(3) are excepted from this limitation (see 45 CFR Part 1305 - source 57 FR 46725, Oct. 9, 1992, as amended at 61 FR 57226, Nov. 5, 1996).

income used throughout the paper includes Food Stamps.

Income does not include capital gains; any assets drawn down as withdrawals from a bank, the sale of property, a house or a car; or tax refunds, gifts, loans, lump-sum inheritances, one-time insurance payments, or compensation for injury. Also excluded are noncash benefits, such as the employer-paid or union-paid portion of health insurance or other employee fringe benefits; food or housing received in lieu of wages; the value of food and fuel produced and consumed on farms; the imputed value of rent from owner-occupied non-farm or farm housing; and such Federal non-cash benefit programs as Medicare, Medicaid, food stamps, school lunches, and housing assistance, and certain disability payments made to disabled children of Vietnam veterans as prescribed by the Secretary of Veterans Affairs.⁵⁴

The Department of Health and Human Services considers that the income period of time to be considered for eligibility is the 12 months immediately preceding the month in which application or reapplication for enrollment of a child in a Head Start program is made, or the calendar year immediately preceding the calendar year in which the application or reapplication is made, whichever more accurately reflects the family's current needs. We use income from last calendar year because it is the income measure available in NLSY79. As of our knowledge D.H.H.S. does not issue any specific definition of "family unity" and therefore we use NLSY79's definition.

To check the veracity of declared income, centers are required to verify the following proofs: Individual Income Tax Form 1040, W-2 forms, pay stubs, pay envelopes, written statements from employers, or documentation showing current status as recipients of public assistance, and should keep a signed statement by an employee identifying which of these documents was examined and stating that the child is eligible to participate in the program. Some centers do not keep an accurate register.

Given that there are two routes of eligibility to Head Start for each child we perform two separate comparisons (the child is eligible if she is in a poor family or if she is in an AFDC/TANF eligible family):

1. Impute child's poverty status: the child is in a poor family if the annual family gross income is below or equal to the Federal Poverty Guideline for each year of data available.
2. Impute child's family AFDC/TANF eligibility. See "AFDC Eligibility Requirements" below for a detailed description.

⁵⁴See <http://eclkc.ohs.acf.hhs.gov/hslc> for the Head Start Program Definition of income and Federal Poverty Guidelines.

To obtain the maximum level of income that would allow Head Start eligibility we perform several comparisons:

1. If the child's family does not verify the categorical requirements to be entitled to AFDC/TANF, the maximum gross income that would have allowed Head Start eligibility is the Federal Poverty Guideline.
2. If the child's family is categorically eligible to AFDC/TANF, several scenarios may emerge:
 - (a) if the family is not receiving income from AFDC, or if this information is missing, then two income tests must be verified in order to become AFDC income eligible:
 - i. if the *gross income test* is not valid, the maximum level of income that would allow Head Start eligibility is

$$\text{MAX}(m \times 12 \times \text{Need Standard}, \text{Federal Poverty Guideline})$$

where m is 1.5 for the years of 1982, 1983 and 1984, and 1.85 from then onwards. We use the Need Standard in the state of residence and year at age four and the Federal Poverty Guideline of the year in which the child turned four years old.

- ii. if the *gross income test* is verified, then the relevant cutoff point will be given by

$$\text{MAX}(\text{MIN}(m \times 12 \times \text{NS}, 12 \times \text{NS} + \text{Annual Deductions}), \text{Federal Poverty Guideline})$$

where NS is the Need Standard in the state of residence at age four.

- (b) since 1982, if the family is currently receiving income from AFDC/TANF only the *gross income test* is performed and the maximum level of income above which the family no longer is income eligible is given by

$$\text{MAX}(m \times 12 \times \text{Need Standard}, \text{Federal Poverty Guideline})$$

- (c) the gross income test had not been implemented as of 1979, 1980 and 1981, and the cutoff is given by

$$\text{MAX}(12 \times \text{NS} + \text{Annual Deductions}, \text{Federal Poverty Guideline}).$$

AFDC Eligibility Requirements⁵⁵

Eligibility for AFDC requires that household contains at least one child less than eighteen years old, and has sufficiently low income and assets levels. Additionally, children in two-parents families may be eligible to AFDC-UP (Unemployed Parent), which requires that parents satisfy a work history requirement and work less than 100 hours per month while on welfare (the Family Support Act of 1988 mandates that states set up AFDC-UP programs, but it allows states to limit benefits to six months per year). There are two income tests that an applicant family must pass in order to become AFDC income eligible (U.S. Congress, 1994):

- the gross income test: a gross income limit for AFDC eligibility of 150 percent of the state's standard was imposed by The Omnibus Reconciliation Act (OBRA) of 1981, and raised to 185 percent by The Deficit Reduction Act of 1984;
- the countable income test: it requires that family income after some disregards must be less than the state's need standard. The countable income is the gross income subtracted of work related expenses, child care expenses, child support disregards up to a maximum.⁵⁶

Eligibility is re-assessed annually, and for those who are already recipients of AFDC/TANF only the first income test is required. To impute the threshold for AFDC/TANF income eligibility for each child we merge the need standard, child support disregards, child care expenses and work related expenses information with the child-level data from the CNLSY by state of residence and family size for each year.

Federal AFDC law requires that all income received by an AFDC recipient or applicant must be counted against the AFDC grant except that income explicitly excluded by definition or deduction. The disregards can be computed as follows. Prior to 1981 there was no allowance for work related expenses and child care expenses were capped at 160 dollars per month. The OBRA of 1981 continued to cap the deduction for child care costs at 160 dollars per month and set the work incentive disregard for work expenses at 75 dollars per month. These allowances were increased by the Family Support Act of 1988 that raised work expenses disregards to 90 dollars per month and the child care expenses to 200 dollars per each child under two years old and 175 dollars for month per each child two years or older. This was effective from October 1, 1989, but as our income values are annual we used

⁵⁵See Hoynes, 1996, for a description of AFDC eligibility rules.

⁵⁶Details on all state-specific values can be found in the Welfare Rules Database of the Urban Institute.

it from 1990 onwards. In 1996, work related expenses were subsequently raised to 100 dollars, 200 dollars in 1997 and 250 dollars per month since 1998. Between 1997 and 1999, child care expenses were set at 200 dollars per each child either she was under or older than two years old. Additionally, the Deficit Reduction Act of 1984 established a monthly disregard of 50 dollars of child support received by family, that is valid from 1985 (inclusive) onwards. As the last age in which we impute program eligibility is 7 years old and the youngest child in our sample was born in 1996, 2003 is the last year in which eligibility status should be imputed.⁵⁷

Since NLSY79 does not contain systematic collection of child care and work related expenses we assume that families fully deduct the full disregard of child care expenses for all children under 6 years old and no disregard for older children (as is imposed by AFDC requirements), and deduct the full amount of work related expenses if the mother or her spouse is working.

Need standard, work related expenses, child care expenses and child support disregards are defined in monthly levels but we convert them into annual values to be comparable with the annual gross income measure available in the NLSY79.

Treatment of Earned Income Tax Credit (EITC) has changed over time. Prior to 1981, EITC was counted only when received, however the OBRA of 1981 requires to assume that working AFDC recipients received a monthly EITC if they appeared eligible for it and regardless of when or if the credit was actually available. The 1984 legislation returned to prior law policy with respect to the EITC: it was to be counted only when actually received. The Family Support Act of 1988 required to disregard the EITC in determining eligibility for and benefits under the AFDC program. As EITC information in NLSY79 started to be recovered with the 2000 wave we ignore the EITC in our analysis.

Our treatment of the data regarding stepparent's or mother's partner income was as follows. The OBRA of 1981 required that a portion of the stepparent income to count as part of the income, however, as NLSY79 total income does not include mother's partner income, we do not include it in the definition of income, but as long as child's mother is married, her husband's income is included in the definition of family income (regardless of whether she is married or not with the child's natural father). Also, if mother is cohabiting, her partner will not be included in the family size variable.

When determining AFDC/TANF eligibility we took into account the program categorical requirements with respect to the family structure. Eligibility under the traditional AFDC program requires that a child resides in a female-headed household, which we considered as a family where a father-figure is missing. How-

⁵⁷NLSY79 and CNLSY surveys were not conducted in odd years after 1994, and income, family size, child's mother cohabiting status and state of residence were not imputed for these years.

ever, children in two-parents households may still be eligible under the AFDC-Unemployed Parent program in those states in which the program is available⁵⁸. Eligibility for AFDC-UP is limited to those families in which the principal wage earner is unemployed but has a history of work. We consider that the principal wage earner has a "history of work" if the the father (or step-father) worked on average less than 100 hours per month in the previous calendar year.⁵⁹ We do not perform the assets test required by AFDC, as information on assets is only available after 1985 in NLSY79.

References

- [1] Hoynes, H., 1996, Welfare transfers in two-parents families: labor supply and welfare participation under AFDC-UP, *Econometrica*, vol.64(2), March 1996, 295-332.
- [2] Zigler, E., and Valentine, J. (Eds.). Project Head Start: A Legacy of the War on Poverty. New York: The Free Press/Macmillan, 1979.

⁵⁸In 1988, the Family Support Act required all States, effective October 1, 1990, to provide AFDC-UP (except American Samoa, Puerto Rico, Guam, and the Virgin Islands until funding ceilings for AFDC benefits in these areas are removed). The two-parent program reverted to optional status for all States after September 30, 1998.

⁵⁹Since 1971, Federal regulations have specified that an AFDC parent must work fewer than 100 hours in a month to be classified as unemployed, unless hours are of a temporary nature for intermittent work and the individual met the 100-hour rule in the two preceding months and is expected to meet it the following month. See U.S. Congress, 1994, for the specific requirements.

G Heterogenous Effects

Figure G.1 displays the range of income cutoffs and family size in our data over which we are able to identify the effects of Head Start. This figure shows that relatively to the standard regression discontinuity, we can exploit additional variation. This "continuum of discontinuities" allows us to go beyond the traditional regression discontinuity design and identify treatment effects for individuals over a wide range of values for income and family size (the two main running variables). Black, Galdo and Smith (2005) also recognize the potential of multiple discontinuities to identify heterogeneous effects of the program. Figure G.1 also includes the number of children per family size and intervals of income cutoffs of \$571. This figure shows that β in equation (3) is a weighted average of effects for children around different cutoffs, where the largest weights correspond to children living in families with 2 to 6 members and with cutoffs varying between \$US10,000-20,000.

Therefore, we also consider models where β varies explicitly across individuals:

$$Y_i = \alpha + \beta_i HS_i + g(Z_i, X_i) + \varepsilon_i. \quad (5)$$

We implement this estimator as follows. In our case, the income cutoffs vary mainly with family size (besides state, year and mother's marital status), so there is a threshold for \bar{Z}_f for each family size f . Therefore, it is possible to estimate effect of the program around each cutoff $\bar{Z}_f, f = \{2, \dots, 15\}$, that is,

$$E(\beta_f | HS(\bar{Z}_f - \delta) - HS(\bar{Z}_f + \delta) = 1, Z_i = \bar{Z}_f).$$

The effect estimated under homogeneous effects is a weighted average of the effect around each threshold

$$\beta = \frac{\sum_{f=2}^{15} \beta_f \mathbf{1}[|Z_{if} - \bar{Z}_f| \leq h]}{\sum_{f=2}^{15} \mathbf{1}[|Z_{if} - \bar{Z}_f| \leq h]}$$

where h is the bandwidth to trim the sample around each cutoff. Therefore, we recover β_f by estimating the following model:

$$Y_i = \alpha + h^*(Z_i, X_i) \times E_i + g(Z_i, X_i) + \varepsilon_i \quad (6)$$

For simplicity, we model $h^*(Z_i, X_i)$ as:

$$\begin{aligned} h^*(Z_i, X_i) = & \beta_0 + \beta_1 \ln Y_{4i} + \beta_2 (\ln Y_{4i})^2 + \beta_3 (\ln Y_{4i})^3 + \beta_4 FS_{4i} + \beta_5 FS_{4i}^2 \\ & + \beta_6 FS_{4i}^3 + \beta_7 \ln Y_{4i} FS_{4i} \end{aligned}$$

where Y_{4i} is child's family income at age 4 and FS_{4i} is the family size. Potentially we would like to estimate $h^*(Z_i, X_i)$ using a more flexible specification, but our sample size forces us to be parsimonious. Even with such parsimonious specification our estimates of this function in the Section 5 are quite imprecise, and results should be seen as suggestive and illustrative of the potential of this approach. We report estimates of $\beta_j, j = 0, \dots, 7$, as well as estimates of the average partial effect of Head Start. We also display graphical representations of $h^*(Z_i, X_i)$.

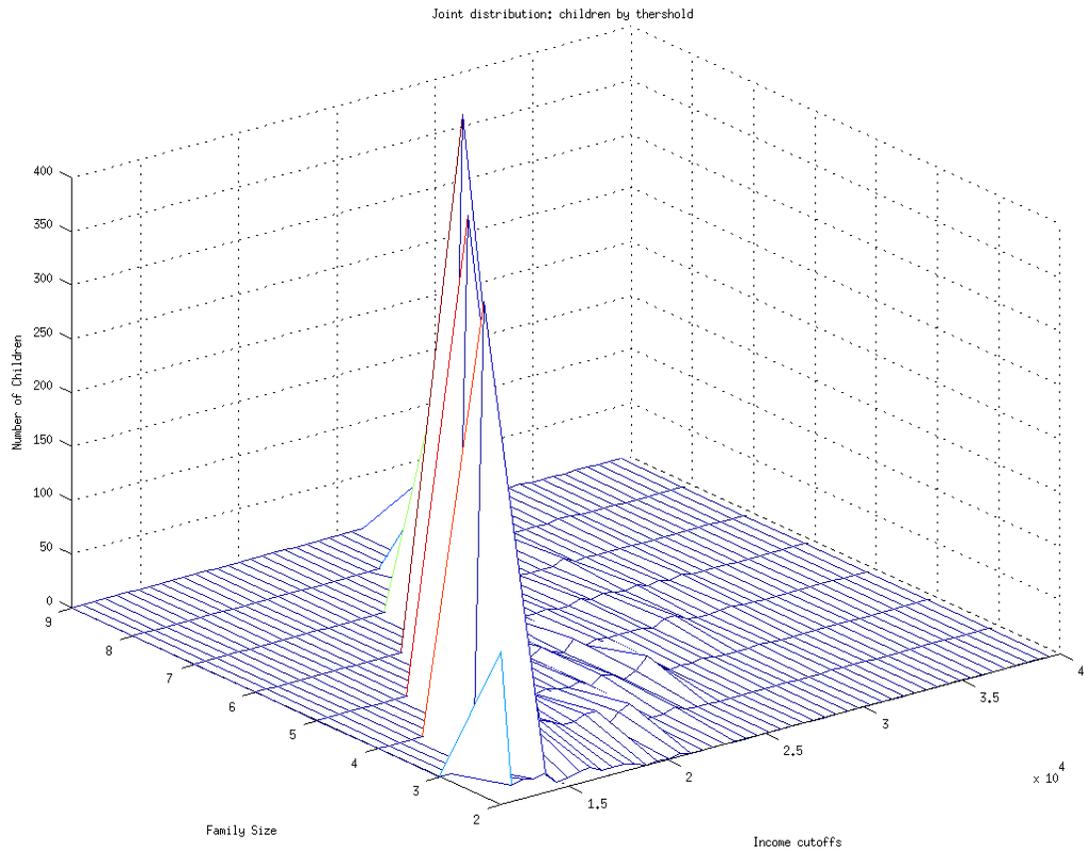


Figure G.1: Density of children around each threshold.

Note: Only children whose family income is at most twice the relevant threshold at age four are included. The sample used is further restricted to children living in families with at most nine members and with a cutoff at most of \$40,000.

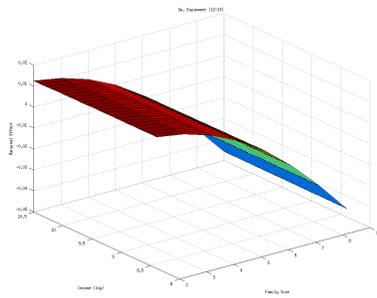
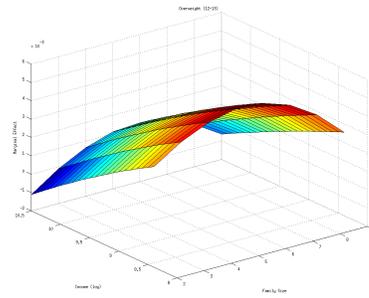
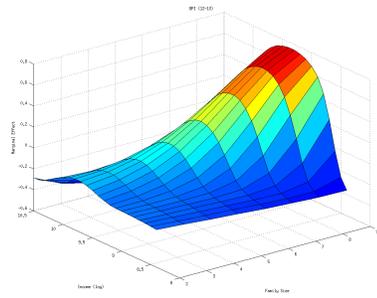


Figure G.2: Ages 12-13.

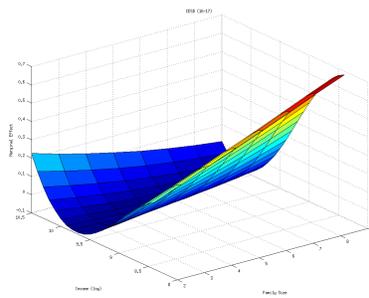


Figure G.3: Ages 16-17