Can Intensive Early Childhood Intervention Programs Eliminate Income-Based Cognitive and Achievement Gaps?*

Greg J. Duncan
University of California, Irvine

Aaron J. Sojourner
University of Minnesota and IZA

October 24, 2012

Abstract

How much of the income-based gaps in cognitive ability and academic achievement could be closed by a two-year, center-based early childhood education intervention? Data from the Infant Health and Development Program (IHDP), which randomly assigned treatment to low birth weight children from both higher- and low-income families between ages one and three, shows much larger impacts among low- than higher-income children. Projecting IHDP impacts to the U.S. population's IQ and achievement trajectories suggests that such a program offered to low-income children would essentially eliminate the income-based gap at age three and between a third and three-quarters of the age-five and age-eight gaps.

* Greg J. Duncan is Distinguished Professor at the School of Education at University of California, Irvine. Aaron J. Sojourner is Assistant Professor at the Carlson School of Management at the University of Minnesota. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A090467 to the University of California, Irvine (George Farkas, Ph.D., PI) and by the National Institute for Child Health and Human Development, through grant HD065704 to the University of California, Irvine (Greg Duncan and George Farkas, PIs). The opinions expressed are those of the authors and do not represent views of these sponsors. We are grateful to Jeanne Brooks-Gunn, Peg Burchinal, George Farkas, Pamela Klebanov, Katherine Magnuson, Judy Temple, Christine Weiland, Hiro Yoshikawa, participants in the Human Capital Research Collaborative and the editor and referees for their help and comments, and to Erik Ruzek and Juan Chaparro for research assistance. Access to the data is restricted. The IHDP data can be obtained through the Inter-university Consortium for Political and Social Research and the ECLS-B data through the U.S. Department of Education's Institute for Education Sciences. The authors will provide additional guidance on request.

I. Introduction

Early childhood education programs are seen by some as a way of improving the schooling readiness of poor children and enabling them to take full advantage of the benefits of K-12 educational investments (Knudsen et al. 2006; Ludwig and Sawhill 2007). But can any single program eliminate achievement gaps? The impacts of modern Head Start and Early Head Start programs directed at children growing up in low-income families are estimated to be modest at best, particularly when outcomes are assessed within a few years of program completion (Puma et al. 2010; Love et al. 2005).[1] Some short-term impact estimates for state pre-kindergarten programs, which are income-targeted in some states and universal in others, are more promising (Wong et al. 2008; Gormley et al. 2008) and mixed evidence of longer-run impacts for these programs is starting to emerge (Hill et al. 2012; Ladd et al. 2012).

Evaluations of the Abecedarian (Campbell et al. 2002), Perry Preschool (Schweinhart et al. 2005; Heckman et al. 2010), and Chicago Child-Parent Center (Reynolds et al. 2011) programs have often been cited as evidence of the long-run impacts and high benefits relative to costs of high-quality programs (Karoly 2001; Knudsen et al. 2006; Bartik 2011). Extracting broad policy lessons from these programs is difficult because all three programs were only offered to low-income and predominantly children of color and their mothers.

Scaled-up, government-funded programs might be offered universally rather than restricted to children from low-income families in the belief that they would benefit all children, that higher-income children generate positive peer effects for low-income children, or in order to generate the political support necessary for public funding. A universal program would close income-based gaps only if its impacts were much larger for low-income children than for higher-income children and if sufficient numbers of low-income families chose to enroll their children in the program.

The goal of this paper is to estimate the degree to which an intensive Abecedarian-type intervention, begun at birth or age one but lasting only until age three, would close income-based gaps in cognitive ability and school readiness. We consider both universal and targeted versions of such a program, with the targeted program restricting eligibility to children living in families with income within 180 percent of the poverty line.

To generate our estimates, we draw data from the Infant Health and Development Program (IHDP), which offered a package of services including free, full-day, Abecedarian-type early education to a randomly chosen subset of 985 children in eight sites scattered around the country (Gross et al. 1997). The IHDP provided seven to nine hours of daily child care and used a game-based curriculum that emphasized language development. Eligibility was not restricted by family income, race or ethnicity and a demographically heterogeneous set of children and families enrolled in the study. A high-quality evaluation design included random assignment of program services to treatment and control groups and assessment of intelligence quotient (IQ) during and up to 15 years after the completion of the program.[2] Published reports have shown very large impacts of the program on IQ during the program and generally smaller impacts, confined exclusively to the heavier babies, after it ended (Brooks-Gunn et al. 1994; Gross et al. 1997; McCarton et al. 1997; McCormick et al. 2006).

Apart from the convenience-based selection of the eight study sites, the main obstacle to generalizing from the IHDP to the larger population of U.S. children is that IHDP services were offered only to babies with birth weights below the low birth weight (LBW) threshold (≤ 2,500 grams). Research has shown that some low birth weight babies, particularly those with birth weights below 1,500 grams, exhibit developmental delays (Gross et al. 1997; Klebanov, Brooks-Gunn, and McCormick 1994a). This raises the question of whether program impacts for low birth weight children generalize to the larger population. As detailed below, we address generalizability issues by showing increasing program impacts throughout the birth weight range and by weighting the IHDP sample to reflect the demographic characteristics of U.S. children.

We find that the IHDP program boosted the cognitive ability of low-income children much more than the cognitive ability of higher-income children. Although early education by family income interactions have been reported in several published studies, our results have much greater internal validity since they are based on a demographically and geographically diverse sample, coupled with a well-implemented random-assignment design and strong program treatment. Population projections show that either a universal or an income-based targeted program would essentially eliminate income-based gaps in IQ at age three – at the end of the program. Despite considerable fadeout of program effects, our estimates suggest that income-based gaps in age-five IQ would be substantially reduced or even eliminated completely. Our increasingly imprecise estimates suggest that one-third to three-quarters of the gaps in age eight IQ and achievement would be eliminated.

II. Background

It is no secret that children from different socioeconomic strata start school with very different skills. A recent study by Duncan and Magnuson (2011) used data from the Early Childhood Longitudinal Study – Kindergarten Cohort to compare children in the bottom and top quintiles of socioeconomic status (SES). They found that low-SES children scored about 1.3 standard deviations lower than high-SES children in their kindergarten-entry reading and math skills and nearly two-thirds of a standard deviation lower in teacher ratings of attention skills. Moreover, they were one-fourth of a standard deviation worse in terms of teacher-reported antisocial behavior. None of these gaps shrank over the course of elementary school, and in the case of antisocial behavior, the SES-based gap nearly doubled. More than half of the SES gaps were found *within* schools, which suggests that the very different kinds of schools attended by poor and affluent children do not account for all of the gaps.

Less well known is the startling growth in the income-based gap on test scores across cohorts of children born since the 1950s (Reardon 2011). Among children born around 1950, test scores of low-income children (defined to be at the 10^{th} percentile of the family income distribution) lagged behind those of their better-off peers (defined to be the 90^{th} percentile) by a little over half a standard deviation. Fifty years later, this gap was twice as large.[3] Given the importance of achievement skills in determining educational success, it should come as no surprise that growth in the income-based gap in children's reading and math achievement has translated into a larger gap in schooling completed by children growing up in poor families compared with their more affluent peers (Duncan and Murnane 2011).

What might be done to close these gaps? Early childhood education (ECE) programs are seen by many as a way of improving the schooling readiness of children and enabling them to take full advantage of the benefits of K-12 educational investments (Knudsen et al. 2006). As with many other social programs, ECE services can be targeted toward low-income children or offered universally regardless of economic need (Scokpol 1991; Greenstein 1991; Barnett et al.

2004). The value of targeting ECE programs on low-income preschool children is ambiguous owing to competing hypotheses about differential program effects can be found in developmental research and theory. A compensatory hypothesis (Sameroff and Chandler 1975) predicts that children who are at risk because of economic disadvantage, low skills, difficult temperaments, etc. derive greater benefit from skill-building high-quality early education programs relative to children who are not at risk. This hypothesis provided the rationale for the initial and continued funding for programs such as Head Start and Early Head Start. However, some have argued for a Matthew effect hypothesis (for example, Stanovich 1986) in which children with greatest initial advantages will profit the most.

We know very little about the comparative effectiveness of infant/toddler and preschool programs for children from low- versus high-income families. The best-known programs (Abecedarian, Perry, and the Chicago Parent-Child Program) restricted eligibility to low-income and/or disadvantaged minority children. Recent evaluations of the national Head Start and Early Head Start programs are also constrained by income limits on eligibility for both programs, although both Loeb et al. (2007) and Magnuson et al. (2004) shows that associations between attending Head Start or center-based care and kindergarten test scores are somewhat stronger for subgroups with the lowest socioeconomic status. The Oklahoma Pre-K program offered services to children who qualified for the free or reduced-price lunch program as well as children who did not. Also in line with the compensatory hypothesis, Gormley et al.'s (2008) evaluation found considerably higher impacts for the former than the latter group.

As with the Oklahoma Pre-K sample, income-based eligibility criteria were not part of the Infant Health and Development Program, which offered a package of services including free, full-day, Abecedarian-type child care to a randomly chosen subset of the 985 mothers and children that it recruited. These children were born in hospitals in eight sites scattered around the country, leading to the enrollment of a very demographically heterogeneous set of children and parents. We build on the IHDP's diverse sample and experimental design to estimate the extent to which income-based gaps in school readiness would be closed by such a program.

III. Approach and Data

Approach. The Infant Health and Development Program was designed to deliver the center-based Abecedarian curriculum to an economically and ethnically diverse sample of one and two year olds in eight sites scattered around the country (McCarton et al. 1997). However, all infants recruited into the IHDP study were born LBW (≤ 2,500 grams = 5.51 pounds) and premature (gestational age at birth ≤ 37 weeks). The motivation for this restriction was that, while the Abecedarian curriculum had been shown to enhance cognitive outcomes for normal-birth-weight, socially-disadvantaged children, no empirical evidence existed about its effectiveness for LBW children. IHDP documentation notes that neonatologists favored the inclusion of only the very low birth weight infants (≤ 1500 g), who are known to be at greatest risk for developmental disabilities but that program developers also felt that it was important to assess impacts in a population where there was some evidence of effectiveness. To balance these two concerns, "it was decided to include infants weighing ≤ 2500 g with gestational age ≤ 37 weeks, but to stratify the sample into two weight groups. The 'lighter' group (≤ 2000 g), would make up two-thirds of the sample, and the 'heavier' group (2001 – 2500 g) would compose one-third of the sample" (Gross et al. 1997).[4] Program take-up was high, the curriculum appeared to be well implemented, attrition through age eight was modest, and a rigorous randomized controlled trial (RCT)-based design provides treatment estimates for a series of IQ and achievement measures both for the entire IHDP sample and for low- and higher-income subsamples.

Two factors in particular make it difficult to generalize from the IHDP to the U.S. population: All IHDP infants were born low birth weight and, although diverse, the demographic characteristics of IHDP families do not match closely to those of the general population. To address the low birth weight issue, we first present evidence showing that the developmental trajectories of the IQs of low birth weight children in the Early Childhood Longitudinal Study – Birth Cohort (ECLS-B) roughly parallel those of normal birth weight children. Then, we show that the IHDP sample includes a substantial numbers of children near the 2,500-gram low birth weight threshold and, most importantly, that IHDP program impacts are, if anything, increasing throughout the entire range of birth weights. This leads us to base our estimates only on the IHDP babies with birth weights above 2,000 grams, conventionally designated as high low birth weight (HLBW).

To correct for demographic misalignment, we develop and apply a set of ratio estimating weights to the IHDP sample based on ECLS-B joint distributions of race/ethnicity, income, maternal education and marital status. The procedures for obtaining estimates are as follows. Define D as an indicator that a child's family income is below 180 percent of the poverty threshold.[5] Define T as a treatment indicator. If we could assign treatment at random in the nationally-representative ECLS population, we would estimate:

(1) $Y_E = a_0 - a_1 D + a_2 T + a_3 D{*}T + e$

where Y_E is IQ or achievement measured in the ECLS-B at various ages during or after the end of the hypothetical program. The "E" subscript denotes an ECLS-B-based estimate. A minus sign precedes the expected negative coefficient on D (a_1) so that the sign on the outcome gap between groups is positive. As shown in Table 1, a_1 measures the gap between average outcome levels for children from higher-income families compared to those from low-income families in the absence of treatment. The effect of universal (offered to the entire population of one- to three-year olds) and targeted (offered only to one- to three-year-old children with family income below 180 percent of the poverty line) programs on outcomes and on the income-based gap could be estimated as described in Table 1. However, T=0 in the ECLS-B's national sample, so we cannot identify the portion of the gap closed by a program (C_a). We can only estimate a_1, the observed "raw" gap absent treatment.

[Table 1 here]

To estimate C, we use the weighted IHDP HLBW sample to generate estimates of b_1, b_2 and b_3 from:

(2) $Y_I = b_0 - b_1 D + b_2 T + b_3 D{*}T + (\text{Site Dummies}) + u$

where Y_I is IQ or achievement measured in the IHDP at various ages during or after the end of the program and the "I" subscript denotes an IHDP-based estimate. This gives analogous estimates of the percentage of gap (C_b) that would be closed by a targeted or universal program, which can be seen by replacing all "a" terms in Table 1 with their analogous "b" terms.

Then, $a_1{*}C_b$ estimates the magnitude of the closure that would be achieved by applying IHDP-based treatment effects to the weighted IHDP-based gap and $a_1{*}(1-C_b)$ estimates the magnitude of the residual gap. As a measure of the gap, one might instead prefer the observed ECLS-B gap rather the weighted IHDP-based gap and ask how much of this gap would a given program close. This amounts to replacing the “a”s with “b”s in the numerators of Table 1 ratios but leaving a_1 in the denominators. Call these gap-closing estimates C_m (Table 2).[6]

[Table 2 here]

Data. As described in the “Approach” section, the Infant Health and Development Program was an eight-site randomized clinical trial designed to evaluate the efficacy of a comprehensive early intervention program for low birth weight premature infants. Infants weighing 2,500 g or less at birth, regardless of parental income status, were screened for eligibility if their post-conceptional age between January and October 1985 was 37 weeks or less and if they were born in one of eight participating medical institutions. Following hospital discharge, a total of 985 infants were randomly assigned either to a comprehensive early childhood intervention group or to a control group that was offered only a package of free medical services explained below.

Children in the treatment group received weekly home visits through 12 months of age, which consisted of a curriculum of child development and parenting education, mental health counseling and support, and referral to social services within the community. Despite these services, there were no significant treatment impacts at age 12 months on either the children or their home environments (Bradley et al. 1994). Home visits continued on a biweekly basis between ages one and three.

Between ages one and three, children in the treatment group were also entitled to attend the free, high-quality IHDP-run child development center located in each city. The curriculum was based on the one used in the Abecedarian Preschool program (Campbell et al. 2002). Free transportation was made available to encourage take-up. Infants in both the treatment and control groups also participated in a pediatric follow-up program of periodic medical, developmental, and familial assessments from 40 weeks of conceptional age (when they would have been born if they had been full term) to 36 months of age, corrected for prematurity.

A frequency distribution of the birth weights of the 985 infants is shown in Figure 1. Most of the infants weighed between 1,500 and 2,500 grams. For reasons detailed below, our analysis will concentrate on the 362 heavier low birth weight children in the 2,001-2,500 gram range.

[Figure 1 here]

We draw our data from a variety of sources – maternal-report questionnaires, home visits, and laboratory tests. Assessment ages for the IHDP are one, two, three, five and eight. The IHDP provides the following cognitive measures: the Bayley IQ mental subscale at ages one and two; the Stanford-Binet IQ mental subscale at age three; the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) Full Scale IQ at age five; and the Wechsler Intelligence Scale for Children (WISC) at age eight. We also study math and reading achievement at age eight as measured by Woodcock-Johnson tests. To preserve comparability with national norms, we standardize all individual outcomes into z-scores that have mean zero and standard deviation one, using the national norms provided by the tests' original developers. All are high-quality, well-validated measures.

One of our analytic goals is to estimate differential treatment effects by income. Our indicator of low-income status is based on whether family income as reported by the mother when the child was 12 months old was below 180 percent of the poverty line (Leventhal and Brooks-Gunn 2001). Some 10.2 percent of mothers failed to report income in that interview. We assume these income data are missing at random conditional on observables and use multiple imputation to make inference (Little and Rubin 1987).[7]

To assess sensitivity of the results to this assumption, we also relax it and estimate bounds on key parameters considering all possible values for the missing indicators (Horowitz and Manski,2000; Horowitz et al. 2003). Given the random assignment nature of the IHDP treatment, we use baseline measures in some specifications to control for small demographic differences in the treatment and control groups and improve the precision of the experimental estimates of treatment effects. These baseline measures include maternal-report data on race/ethnicity (with indicators of African-American and Hispanic status) and maternal education level (four categories – less than high school, high school graduate - no college, high school graduate - some college, and college graduate). At the child level, we control for child's sex, birth weight in grams, gestational age at birth in weeks, a neonatal health index, maternal age in years at child's birth, and a set of site dummies. These variables have no missing observations.

Response rates were high in the early waves of IHDP data collection, but lower for the longer-run follow-ups. For the sample of high low birth weight children used in our analyses, response rates for IQ tests were 91.1 percent, 88.9 percent, 90.6 percent, 81.4 percent and 85.9 percent at age one, two, three, five, and eight, respectively. Outcome data are assumed missing at random conditional on covariates and, for each outcome, cases with missing outcome data are dropped.[8]

As described in the online appendix, the more familiar ECLS-B has followed to kindergarten entry a large, nationally representative sample of children born in 2001. We use the ECLS-B-provided weights to make the sample nationally-representative adjusting both for differential sampling probabilities and for differential nonresponse. The ECLS-B provides the following cognitive measures: a reduced-item form version of the Bayley (Bayley Short Form-Research Edition) designed to produce equivalent scores to the original Bayley Scales of Infant Development mental subscale at age 24 months; the Peabody Picture Vocabulary Test and ECLS-B developed literacy and math assessments at age 48 months; and ECLS-B developed reading and math assessments at kindergarten entry. All of these tests have been normed to the general population. We also utilized the following demographic measures from the ECLS-B: maternal education in years, maternal race, ethnicity, and marital status, household size, and child birth weight.

Population weights. Although the IHDP sample is economically and ethnically diverse, it was not designed so that its demographic characteristics matched those of any larger population. To correct for this, we construct a set of weights based on the relative frequency of observations that fell into cells defined jointly by family income (below or above 180 percent of the poverty line), race/ethnicity (African-American, Hispanic, or other [mostly white/non-Hispanic]), maternal schooling (no college or at least some college) and marital status (married or not) in the IHDP HLBW subsample and the ECLS-B sample. All of these demographic characteristics were measured at nine months in the ECLS-B and at birth in the IHDP, except for family income status, which was measured at 12 months in the IHDP. The details of these procedures and comparisons of the unweighted and weighted IHDP sample are provided in the online appendix.

IV. Results

Trajectories of low and normal birth weight babies. Since the ECLS-B is nationally representative, oversampled low birth weight births and measured cognitive ability repeatedly up to the point of kindergarten entry, it is well suited for providing data on the developmental trajectories of low and normal birth weight babies. Using the full, weighted ECLS-B sample to standardize its IQ measures to have mean zero and standard deviation of one, Table 3 shows IQ scores at various ages for normal birth weight (more than 2,500 grams), high low birth weight (2,000-2,500 grams), and very low birth weight (less than 2,000 gram) babies.

[Table 3 here]

Age 24 months is the first point at which IHDP's evaluation measured impacts of its center-based ECE services. In the ECLS-B, measured IQs of HLBW babies at 24 months are about one-sixth of a standard deviation below those of normal birth weight babies; the gap for low-low birth weight (LLBW) babies is about twice as large. In the case of HLBW babies, the IQ and, at 48 and 60 months, achievement gaps are within 0.10 sd of the 24-month IQ gap, suggesting roughly parallel trajectories. In the case of LLBW babies, the 24-month gap is larger than any of the later gaps, and math gaps tend to be consistently larger than reading gaps.[9] This adds to our confidence that that result from the IHDP's HLBW babies may generalize and our wariness that results from the IHDP's LLBW babies may not.

Marginal treatment effects by birth weight. We next examined IHDP treatment effects by age and birth weight for any indication that treatment effects declined with birth weight, which would raise concerns that IHDP treatment effects might not generalize to normal birth weight babies. If anything, the opposite was true. Marginal treatment effects with 95 percent confidence bands are shown in Figure 2 for standardized IQ measures taken at ages two, three, five and eight and age-eight math and reading achievement.[10] We fit linear through fourth-order polynomials and found that the results were generally quite similar. For the sake of conciseness we show only linear interactions here and fourth-order polynomials as Appendix Figure 1.

[Figure 2 here]

Substantial and statistically significant treatment effects on IQ are apparent for most birth weights in the middle (age two) and the very end (age three) of the center-based IHDP treatment. In all cases, treatment effects in the 2,000-2,500 gram range (the definition of high LBW) are at least as large as treatment effects at lighter birth weights. At ages five and eight (three and five years after the end of the program, respectively), treatment/control group differences are less apparent, although in all cases the treatment group advantages are at least as large for the HLBW babies as for the lighter-birth weight babies. The consistently rising marginal treatment effects for all outcomes across the 2,000 to 2,500 grams range suggest that patterns of treatment effects for the HLBW children can provide a useful basis for generalizing to at least those normal birth weight babies who are at the lighter end of the birth weight spectrum assuming the smooth pattern continues.

Treatment main effects and interactions. Proceeding on the assumption that the ECLS-B-weighted sample of HLBW babies in the IHDP sample can be used to estimate treatment interactions with income, we present estimates from ordinary least squares (OLS) regression models for all of the available IQ (Table 4) and achievement (Table 5) measures in the IHDP. Each table shows coefficients from three models: i) T only; ii) T and D entered additively, and iii) T and D main effects and their interaction. In all cases we include controls for site, the child's sex, gestational age at birth, birth weight and neonatal health index. Results for the main coefficients of interest are summarized in Table 4, with complete details in Appendix Tables 3-5.

[Tables 4 and 5 here]

Average treatment effect estimates on IQ for this sample of HLWB babies are given in Model A of Table 4. Consistent with Figures 2-4, large treatment effects emerge by age two, peak at age three and, while point estimates continue to be positive, become statistically insignificant by age five. Treatment impacts on ages eight achievement are insignificant as well. Model 2 adds a “low income” dummy variable to the model. Estimates show that the age two and three IQs of children reared in low-income families score close to one standard deviation below those of higher-income children, while age five and eight IQs are half to two-thirds of a standard deviation lower for children reared in low-income families. All of these differences are statistically significant. Table 5 shows that achievement differences are similar to the IQ differences, with a range between one half and one standard deviation between low and high income children across different ages and subjects.

Treatment effect differences between low and higher-income children are estimated by the coefficient on the Treatment by Low income interaction variable in Model C. In this specification, the coefficient on the “Treatment” dummy represents the estimated program impact for children from higher-income families. Treatment impacts on IQ are estimated to be much larger for children from low-income families than for children from higher-income families – by 0.87 sd at age two and 1.32 sd at age three. These differences persist two years after the end of the program, as is evident from the statistically significant +0.86 sd interaction coefficient at age five. The point estimate for the treatment-by-income interaction (+0.57 sd for age eight IQ) is substantial in size but statistically insignificant. Interaction coefficients on reading and math achievement at age eight are both substantial and statistically significant. The magnitude of these interaction coefficients suggests that an IHDP-type program may well eliminate quite a bit of the income-based IQ and achievement gaps.

The "Treatment" variable in Model C provides an estimate of the program's treatment effect on higher income children. Point estimates for all age five and eight IQ and achievement measures are negative although always statistically insignificant. So while the sign of the point estimate might suggest that the program hurt the cognitive development of higher income children relative to their high-income counterparts in the control group, these estimates are imprecise and 90 percent confidence intervals include positive effects as well. We return to this point in our discussion section.[11]

Gap closing. We now apply ECLS-B-based population weights to the IHDP data on treatment effect interactions to estimate the extent to which income-based IQ and achievement gaps would be closed by an IHDP-type early education intervention between ages one and three offered either universally or targeted only to low-income children (Table 6 and Figure 3).[12] The second row of Table 6 shows that at age two, after one year of the early childhood education curriculum, the 0.82 sd higher treatment effect for low- relative to high-income children closed 75 percent of the 1.35 sd gap in the case of a universal program and 117 percent of the gap if the program was offered only to low-income children. A 95 percent confidence interval for the share of this age-two gap closed by a universal program ranges from 34 percent to 116 percent, while the confidence interval for the share closed by a targeted program ranges from 78 percent to 158 percent. The extent to which income-based gaps at age three would be closed by an Abecedarian-type program is somewhat greater than at age two in the case of a universal program and slightly less in the case of a targeted program.

[Table 6 and Figure 3 here]

A key question motivating our efforts is whether a very high-quality infant and toddler program might be able to close the school readiness gap between low and higher income children. The age-five IQ results presented in Table 6 suggest that virtually all of the income-based gaps would be closed by a universal program. Although it does not include zero, the large confidence interval for the universal program estimate suggests caution against over-interpreting this point estimate. Surprisingly, only 72 percent of the income-based gap in age-five IQ would be closed with a targeted program. The counterintuitive reduced effectiveness for targeted relative to universal programs comes from the negative point estimate of age-five IQ impacts for children with incomes above 180 percent of the poverty line. It is important to bear in mind that the negative treatment impacts estimated for the age five IQs of higher-income children was not statistically significant and that the confidence intervals on gap reductions for targeted and universal program overlap considerably.

While not all of the impact estimates at age eight are statistically significant, the pattern suggests that a universal program would reduce income-based gaps by more than half, while a targeted program would reduce gaps by one-third to about one-half. Here again, the overlap between the confidence intervals of target and universal program is considerable and none of the underlying negative treatment impacts estimates for the IQ and achievement of higher-income children was statistically significant. While we can usually reject the null hypothesis of no gap closing, we cannot reject the hypothesis of similar gap closings for targeted and universal programs.

The results presented in Tables 6-8 rest on the assumption that the data on low-income status is missing at random, which helps to provide point identification and is necessary to justify use of multiple imputation. However, if this identifying condition was not valid, it is possible that results would change substantially. Estimates can be more sensitive to missing covariate data than to missing outcome data.[13] To assess the sensitivity of the results to the missing-at-random assumption, we relax it and study the set of parameter values consistent with the model and the observed data, considering all possible combinations of values for the missing low-income indicators. We estimate the model for each outcome and for each possible combination of missing low-income indicator values. Across all possible combinations, the minimum and maximum estimated value of each parameter is recorded, providing point estimates of the lower and upper bounds on each parameter.[14] These are reported in Table 7. By and large their ranges are consistent with the picture provided by our previous analysis.

[Table 7 here]

V. Discussion

Our paper has sought to estimate how much an intensive two-year center-based Abecedarian-type intervention begun at age one would close income-based gaps in cognitive ability and school readiness. The analysis suggests that at age three – at the end of the program – income-based gaps would be essentially eliminated with either a universal or income-based targeted program. Income-based gaps in age five IQ were also substantially reduced (in the case of a targeted program) or completely eliminated (for a universal program). Our increasingly imprecise estimates suggest that one-third to three-quarters of the gaps in age eight IQ and achievement would be eliminated.

These results make two contributions. First, they inform the debate over targeted versus universal ECE programs by taking advantage of a well-implemented, intensive early education treatment administered to a demographically and geographically diverse sample of children. Results from its random-assignment evaluation design show how much more the IQs of low- than higher-income children profit from such treatments. Second, we use demographic methods to project the impacts of these results to the national population of young children living in low- and higher-income families.

While it is certainly encouraging to see that school readiness gaps between high and low-income children might be reduced or even eliminated with an intensive early education program, several cautions are in order. First, prudent policy planning should be based on a comparison of benefits and costs of competing programs, as well as evidence that scale-up does not compromise program impacts. In contrast with results from the current paper and others based on model programs targeting children from low-income families, recent evaluations of the Early Head Start (Love et al. 2005) and Head Start (Puma et al. 2010) programs have not produced evidence of large impacts on low-income children in the short-run, although there is evidence of substantial long-run effects from Head Start programs (Ludwig and Miller 2007; Deming 2009).

Early Head Start (EHS) is the federal program closest to the IHDP in design. Both EHS and the IHDP offer families a mix of home visiting and center-based care for children up to age three. Why the difference in effects between EHS and IHDP? One possibility is that the difference in effects derives from differences in program intensity and quality. For instance, according to calculations based on Love et al. (2005), the average Early Head Start participant received 437 hours of center-based care. In contrast, the average member of the IHDP treatment group received 260 days of center-based care, or 2,080 hours if attending for eight hours a day. Moreover, the IHDP went to great lengths to ensure that care standards were uniformly high and the curriculum was well implemented, while the quality of Early Head Start programs is more variable (Love et al. 2005; Vogel et al. 2011).

As with any high-quality center-based program for infants and toddlers, the low (three to one) staff-to-student ratio and other services offered in their Abecedarian-type treatment made IHDP services relatively expensive.[15] Evidence from state pre-K programs suggests that relatively high-quality care can be taken to scale. Many state programs targeting primarily three and four year olds have been implemented at large scales in recent years. Early reports show positive short-run achievement effects of some (Wong et al. 2008), especially for low-income children (Bartik 2011), and there is some emerging evidence of positive effects on age-eight achievement (Hill et al. 2012; Ladd et al. 2012).

We do not know whether the IHDP treatment could be scaled up in a general way, although the curriculum of the IHDP itself replicated the curriculum used in Abecedarian and was successfully implemented in eight sites scattered around the country. If the program were scaled up to a national program, the current study suggests the intensive, high-quality services it would provide could make a large, persistent positive impacts on low-income children's cognitive skill and academic achievement and reduce, if not eliminate, the early skills gap between America's children from low and higher-income families.

A second cautionary note is that success in closing income-based gaps may not generalize directly to success in closing gaps defined by race or ethnicity. Reardon (2011) shows that trends in the racial gap in tests scores are quite different than trends in income-based test-score gaps. When we repeated our OLS regressions of IHDP treatment impacts on IQ and achievement differences between blacks and whites and between Hispanics and non-Hispanic whites, we did not find the kinds of consistent impact patterns favoring minority children as we did for low-income children.[16] This is an important issue for future research.

Third, the evolving patterns of impacts found for low- and higher-income children depend upon the quality of care for control-group children during and after the program. If current patterns of care quality for low and high-income children differ from those in the late 1980s, then the patterns of impacts and gap closings found here may not generalize to the current day. However, Leventhal et al. (2000) reports that, in the full IHDP sample at age five, average preschool attendance was 24.1 and 23.0 hours/week in the control and treatment groups, respectively. Using the ECLS-B's nationally-representative sample of four year olds in 2005, Jacobson Chernoff et al. (2007) reports that 57.5 percent of children report attending center-based care. If these children attended an average of 40 hours per week, this would yield a population average of 23 hours/week attendance, very close to the IHDP, age-five average. Furthermore, center-based care participation does not vary by birth weight in the ECLS-B. Taken together, this suggests that patterns of center-based care use are similar in these two samples.

Fourth, while treatment effects on the IQs of higher-income children are positive and significant during and at the end of the program, point estimates of IQ and achievement at age five and eight for these children are not statistically significant at conventional levels. The point estimates are noisy but negative, which produces a larger point estimate of the fraction of gap closed for the universal than the targeted program. However, the two programs are estimated to have similar confidence intervals for fraction of gap closed and this seems a more appropriate way to interpret the results. The estimates provide strong evidence that a targeted program would close a large share of the gap and that a universal program would produce similar results, since the program appears to have no significant effect for children from higher-income families.[17]

Fifth, unlike Abecedarian, Perry or Child-Parent Centers, the IHDP treatment delivered services in a center serving a heterogeneous group of children. If peers matter, then this is part of the treatment effect. A targeted program that delivers care in a setting with a more homogeneous group of children may produce different results.

More than two decades ago Lisbeth Schorr wrote of the promise of early childhood intervention programs in a book titled *Within Our Reach* (Schorr 1989). More recent work has echoed this theme (for example, Ludwig and Sawhill 2007; Kirp 2007). At that time, she could only speculate on whether income-based achievement gaps might be closed with intervention programs. Although based on an experiment involving low birth weight children, our analysis provides more concrete evidence supporting these conjectures about the potential of early childhood interventions to close achievement gaps.

References

Bailey, Martha, and Susan Dynarski. 2011. "Inequality in Postsecondary Education." In *Whither Opportunity: Rising Inequality, Schools, and Children's Life Chances*, ed. Greg J. Duncan and Richard J. Murnane, 117-132. New York: Russell Sage Foundation.

Barnett, Steven, Kirsty Brown, and Rima Shore. 2004. “The Universal versus Targeted Debate: Should the United States Have Preschool For All?” *Preschool Policy Matters*. Issue 6. New Brunswick: National Institute for Early Education Research.

Bartik, Timothy. 2011. *Investing in Kids: Early Childhood Programs and Local Economic Development*. Kalamazoo: W.E. Upjohn Institute for Employment Research.

Bradley, Robert, Leanne Whiteside, Daniel Mundfrom, Patrick Casey, Bettye Caldwell, and Kathleen Barrett. 1994. “Impact of the Infant Health and Development Program on the Home Environments of Infants Born Prematurely and with Low Birthweight.” *Journal of Educational Psychology* 86(4): 531-41.

Brooks-Gunn, Jeanne, Cecilia McCarton, Patrick Casey, Marie McCormick, et al. 1994. “Early Intervention in Low-Birth-Weight Premature Infants: Results Through Age 5 Years from the Infant Health and Development Program." *Journal of American Medical Association* 272(16): 1257-1262.

Campbell, Frances, Craig Ramey, Elizabeth Pungello, Joseph Sparling, and Shari Miller-Johnson. 2002. “Early Childhood Education: Young Adult Outcomes from the Abecedarian Project.” *Applied Developmental Science* 6(1): 42-57.

Deming, David. 2009. "Early Childhood Intervention and Life-Cycle Skill Development: Evidence from Head Start." *American Economic Journal: Applied Economics* 1(3): 111-34.

Duncan, Greg, and Katherine Magnuson. 2011. "The Nature and Impact of Early Achievement Skills, Attention Skills, and Behavior Problems." In *Whither Opportunity: Rising Inequality, Schools, and Children's Life Chances*, ed. Greg J. Duncan and Richard J. Murnane, 47-69. New York: Russell Sage Foundation.

__________, and Richard Murnane. 2011. "Introduction." In *Whither Opportunity: Rising Inequality, Schools, and Children's Life Chances*, ed. Greg J. Duncan and Richard J. Murnane, 3-23. New York: Russell Sage Foundation.

Gormley, William, Deborah Phillips, and Ted Gayer. 2008. "Preschool Programs Can Boost School Readiness." *Science* 320(5884): 1723-4.

Greenstein, Robert. 1991. "Universal and Targeted Approaches to Relieving Poverty: An Alternative View." In *The Urban Underclass*, ed. Christopher Jencks and Paul Peterson, 437-459. Washington, DC: Brookings Institution Press.

Gross, Ruth, Donna Spiker, and Christine Haynes, eds. 1997. *Helping Low Birth Weight, Premature Babies: The Infant Health and Development Program*. Stanford: Stanford University Press.

Heckman, James, Seong Moon, Rodrigo Pinto, Peter Savelyev, and Adam Yavitz. 2010. "The Rate of Return to the HighScope Perry Preschool Program." *Journal of Public Economics* 94(1): 114-128.

Hill, Carolyn, William Gormley, and Shirley Adelstein. 2012. “Do the Short-Term Effects of a Strong Preschool Program Persist?” Working Paper #18, Center for Research on Children in the U.S. Washington D.C: Georgetown University.

Horowitz, Joel, and Charles Manski. 2000. “Nonparametric Analysis of Randomized Experiments with Missing Covariate and Outcome Data.” *Journal of the American Statistical Association* 95(449): 77-84.

__________, Charles Manski, Maria Ponomareva, and Jörg Stoye. 2003. “Computation of Bounds on Population Parameters When the Data Are Incomplete.” *Reliable Computing* 9(6): 419-440.

Jacobson Chernoff, Jodi, Kristin Flanagan, Cameron McPhee, and Jennifer Park. 2007. “Preschool: First Findings from the Third Follow-up of the Early Childhood Longitudinal Study, Birth Cohort (ECLS-B).” NCES 2008-025, National Center for Education Statistics. Washington, D.C: Institute of Education Sciences, U.S. Department of Education.

Jefferis, Barbara, Chris Power, and Clyde Hertzman. 2002. “Birth Weight, Childhood Socioeconomic Environment, and Cognitive Development in the 1958 British Birth Cohort Study.” *British Medical Journal* 325(7359): 1-6.

Karoly, Lynn. 2001. “Investing in the Future: Reducing Poverty Through Human Capital Investments.” In *Understanding Poverty*, ed. Sheldon Danziger and Robert Haveman, 314-346. Cambridge: Harvard University Press.

Kirp, David. 2007. *The Sandbox Investment: The Universal Preschool Movement and Kids-First Politics*. Cambridge: Harvard University Press.

Klebanov, Pamela, Jeanne Brooks-Gunn, and Marie McCormick. 1994a. “Classroom Behavior of Very Low Birth Weight Elementary School Children." *Pediatrics* 94(5): 700-708.

__________, Jeanne Brooks-Gunn, and Marie McCormick. 1994b. “School Achievement and Failure in Very Low Birth Weight Children.” *Journal of Developmental and Behavioral Pediatrics* 15(4): 248-256.

Knudsen, Eric, James Heckman, Judy Cameron, and Jack Shonkoff. 2006. “Economic, Neurobiological, and Behavioral Perspectives on Building America’s Future Workforce.” *Proceedings of the National Academy of Sciences* 103(27): 10155-62.

Ladd, Helen, Clara Muschkin, and Kenneth Dodge. 2012. “From Birth to School: Early Childhood Initiatives and Third Grade Outcomes in North Carolina.” Working Paper. Durham: Duke University.

Leventhal, Tama, and Jeanne Brooks-Gunn. 2001. “Changing Neighborhoods and Child Well-Being: Understanding How Children May Be Affected in the Coming Century.” *Advances in Life Course Research* 6: 263-301.

__________, Jeanne Brooks-Gunn, Marie McCormick, and Cecelia McCarton. 2000. “Patterns of Service Use in Preschool Children: Correlates, Consequences, and the Role of Early Intervention.” *Child Development* 71(3): 802-819.

Little, Roderick, and Donald Rubin. 1987. *Statistical Analysis with Missing Data*. New York: Wiley and Sons.

Loeb, Susanna, Margaret Bridges, Daphna Bassok, Bruce Fuller, and Russell Rumberger. 2007. “How Much is Too Much? The Influence of Preschool Centers on Children’s Social and Cognitive Development.” *Economics of Education Review* 26(1): 52–66.

Love, John, Ellen Kisker, Christine Ross, Jill Constantine, Kimberly Boller, Rachel Chazan-Cohen, et al. 2005. "The Effectiveness of Early Head Start for 3-Year-Old Children and Their Parents: Lessons for Policy and Programs." *Developmental Psychology* 41(6): 885-901.

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

___________, and Isabel Sawhill. 2007. "Success by Ten: Intervening Early, Often and Effectively in the Education of Children." Discussion Paper 2007-2, Hamilton Project. Washington D.C: The Brookings Institution.

Magnuson, Katherine, Marcia Meyers, Christopher Ruhm, and Jane Waldfogel. 2004. "Inequality in Preschool Education and School Readiness." *American Educational Research Journal* 41(1): 115-157.

McCarton, Cecilia, Jeanne Brooks-Gunn, Ina Wallace, Charles Bauer, Forrest Bennett, Judy Bernbaum, Sue Broyles, Patrick Casey, Marie McCormick, David Scott, Jon Tyson, James Tonsela, and Curtis Meinen. 1997. "Results at Age 8 Years of Early Intervention for Low-Birth-Weight Premature Infants: the Infant Health and Development Program." *Journal of the American Medical Association* 277(2): 126-132.

McCormick, Marie, Jeanne Brooks-Gunn, Stephen Buka, Julie Goldman, Jennifer Yu, Mikhail Salganik, David Scott, Forrest Bennett, Libby Kay, Judy Bernbaum, et al. 2006. "Early Intervention in Low Birth Weight Premature Infants: Results at 18 Years of Age for the Infant Health and Development Program," *Pediatrics* 117(3): 771-780.

McDonald, Alison. 1964. “Intelligence in Children of Very Low Birth Weight.” *British Journal of Preventive & Social Medicine* 18(2): 59-74.

Puma, Michael, Stephen Bell, Ronna Cook, Camilla Heid, Gary Shapiro, Pam Broene, Frank Jenkins, Philip Fletcher, Liz Quinn, Janet Friedman, Monica Rohacek, Gina Adams, and Elizabeth Spier. 2010. “Head Start Impact Study Technical Report.” Washington, D.C.: U.S. Department of Health and Human Services.

Reardon, Sean. 2011. “The Widening Academic-Achievement Gap between the Rich and the Poor: New Evidence and Possible Explanations.” In *Whither Opportunity: Rising Inequality, Schools, and Children's Life Chances*, ed. Greg J. Duncan and Richard J. Murnane, 91-116. New York: Russell Sage Foundation.

Reynolds, Arthur, Judy Temple, Barry White, Suh-Ruu Ou, and Dylan Robertson. 2011. “Age 26 Cost–Benefit Analysis of the Child Parent Center Early Education Program.” *Child Development* 82(1): 379-404.

Sameroff, Arnold, and Michael Chandler. 1975. “Reproductive risk and the continuum of caretaking casualty.” In *Review of Child Development Research, Volume* 4, ed. Frances. D. Horowitz. Chicago: University of Chicago Press.

Schweinhart, Lawrence, Jeanne Montie, Zongping Xiang, William Barnett, Clive Belfield, and Milagros Nores. 2005. *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40*. Ypsilanti: High/Scope Press.

Schorr, Lisbeth. 1989. *Within Our Reach: Breaking the Cycle of Disadvantage.* New York: Anchor Press.

Skocpol, Theda. 1991. “Targeting Within Universalism.” In *The Urban Underclass*, ed. Christopher Jencks and Paul Peterson, 411 – 436. Washington, DC: Brookings Institution Press.

Stanovich, Keith. 1986. “Matthew Effects in Reading: Some Consequences of Individual Differences in the Acquisition of Literacy.” *Reading Research Quarterly* 21(4): 360–407.

Vogel, Cheri, Kimberly Boller, Yange Xue, Randall Blair, Nikki Aikes, Andrew Burwick, Yevgeny Shrago, Barbara Carlson, Laura Kalb, Linda Mendenko, Judith Cannon, Sean Harrington, and Jillian Stein. 2011. “Learning As We Go: A First Snapshot of Early Head Start Programs, Staff, Families, and Children.” OPRE 2011-7, Office of Planning, Research, and Evaluation, Administration for Children and Families. Washington, DC: US Department of Health and Human Services.

Wong, Vivian, Thomas Cook, Steven Barnett, and Kwanghee Jung. 2008. “An Effectiveness-Based Evaluation of Five State Pre-Kindergarten Programs.” *Journal of Policy Analysis and Management* 27(1): 122-154.

Table 1

Construction of hypothetical gap closing estimates from the ECLS-B

	No program (T=0)	Universal (T=1)	Targeted (T=D)
Predicted outcome among Higher income (D=0)	a_0	a_0+a_2	a_0
Predicted outcome among Low income (D=1)	a_0-a_1	$a_0-a_1+a_2+a_3$	$a_0-a_1+a_2+a_3$
Predicted Gap	a_1	a_1-a_3	$a_1-a_2-a_3$
Portion of gap closed (C_a)		a_3/a_1	$(a_2+a_3)/a_1$

Table 2

Gap closing estimates based on the IHDP and mixed IHDP/ECLS-B

Measures of portion of gap closed	Universal (T=1)	Targeted (T=D)
C_b	b_3/b_1	$(b_2+b_3)/b_1$
C_m	b_3/a_1	$(b_2+b_3)/a_1$

Table 3

Means and standard deviations for ECLS-B measures of cognitive ability and achievement trajectories, for normal birth weight, HLBW and LLBW children

	Normal birth weight (Greater than 2,500 grams)	HLBW (2,000-2,500 grams)	LLBW (Less than 2,000 grams)
IQ or achievement measure			
IQ at 24 months	0.02 (1.00)	-0.17 (1.01)	-0.36 (0.96)
N	6534	1026	1355
PPVT 48 months	0.02 (1.00)	-0.19 (0.98)	-0.24 (1.02)
	6125	962	1320
Reading 48 months	0.02 (1.00)	-0.22 (0.91)	-0.13 (1.16)
	6059	945	1283
Math 48 months	0.02 (1.00)	-0.27 (0.97)	-0.25 (1.04)
	6059	940	1267
Reading 60 months	0.01 (1.00)	-0.13 (1.01)	-0.10 (1.00)
	4941	715	1032
Math 60 months	0.02 (0.99)	-0.23 (1.02)	-0.25 (1.10)
	4941	716	1037
Demographic characteristics			
Maternal education in years	12.86 (2.86)	12.63 (2.69)	12.56 (2.77)
N	7693	1177	1749
Income / 180 percent poverty	1.57 (1.45)	1.33 (1.3)	1.32 (1.3)
N	7729	1189	1769

Note: Standard deviations are given in parentheses. All IQ scores are standardized based on the ECLS-B's weighted national norms to have mean equal to 0 and standard deviation equal to 1.

Table 4

Treatment effects on IQ z-score by low-income status using IHDP HLBW sample with ECLS-B weights.

Outcome (sample size)		Model A	Model B	Model C
Age 1 IQ	Treatment	0.109	0.112	0.065
(n=330)		(0.132)	(0.133)	(0.177)
	Low income		-0.037	-0.072
			(0.122)	(0.171)
	Treatment x			0.097
	(Low income)			(0.253)
Age 2 IQ	Treatment	0.793***	0.878***	0.433*
(n=322)		(0.160)	(0.223)	(0.219)
	Low income		-0.875***	-1.181***
			(0.244)	(0.270)
	Treatment x			0.872**
	(Low income)			(0.280)
Age 3 IQ	Treatment	0.903***	1.001***	0.323
(n=328)		(0.147)	(0.181)	(0.210)
	Low income		-1.017***	-1.482***
			(0.192)	(0.240)
	Treatment x			1.319***
	(Low income)			(0.308)
Age 5 IQ	Treatment	0.102	0.148	-0.264
(n=295)		(0.116)	(0.166)	(0.201)
	Low income		-0.509*	-0.820***
			(0.246)	(0.231)
	Treatment x			0.861***
	(Low income)			(0.201)
Age 8 IQ	Treatment	0.156	0.224	-0.067
(n=311)		(0.158)	(0.169)	(0.323)
	Low income		-0.595**	-0.806***
			(0.185)	(0.196)
	Treatment x			0.572
	(Low income)			(0.361)

Coefficient significance (within site correlation corrected standard errors): *0.10 **0.05 ***0.01.

All models also condition on child gender, birth weight, gestational age at birth, neonatal health index and site indicators. Estimates in appendix.

Table 5

Treatment effects on achievement z-score by low-income status using IHDP HLBW sample with ECLS-B weights

Outcome (sample size)		Model A	Model B	Model C
Age 8 Reading	Treatment	-0.116	-0.041	-0.456
(n=308)		(0.209)	(0.261)	(0.267)
	Low income		-0.643***	-0.936***
			(0.156)	(0.123)
	Treatment x			0.804***
	(Low income)			(0.184)
Age 8 Math	Treatment	0.120	0.187	-0.137
(n=312)		(0.149)	(0.206)	(0.197)
	Low income		-0.594*	-0.830**
			(0.257)	(0.281)
	Treatment x			0.636**
	(Low income)			(0.224)

Coefficient significance (within site correlation corrected standard errors): *0.10 **0.05 ***0.01.

All models also condition on child gender, birth weight, gestational age at birth, neonatal health index and site indicators. Estimates in appendix.

Table 6

Estimated impacts of IHDP treatment effects on high/low income IQ and achievement gaps.

Age and outcomes	High-low income gap		IHDP treatment effect for non low-income subsample	IHDP treatment* Low-income interaction	Percent Gap closed from universal program	Percent Gap closed from targeted program
	IHDP mean difference	b1	b2	b3	Cb	Cb
IQ at 9 or 12 months	0.24* (0.14)	0.01 (0.18)	0.11 (0.18)	0.04 (0.25)	159 (100545)	999 (182168)
IQ at age 2	1.35*** (0.18)	1.09** (0.28)	0.46* (0.23)	0.82** (0.30)	74.8** (22.2)	117.2*** (20.8)
IQ at age 3	1.76*** (0.17)	1.42*** (0.24)	0.34 (0.21)	1.28*** (0.31)	89.4*** (15.2)	113.4*** (8.4)
IQ at age 5	1.08*** (0.14)	0.76** (0.21)	-0.22 (0.22)	0.77** (0.22)	101.2* (43.7)	71.8*** (15.8)
IQ at age 8	1.03*** (0.15)	0.77*** (0.17)	-0.08 (0.29)	0.52 (0.34)	67.1 (40.9)	56.7** (15.9)
Reading at age 8 or grade 3	1.08*** (0.18)	0.82*** (0.13)	-0.44 (0.27)	0.74** (0.23)	89.7* (37.4)	36.6 (28.6)
Math at age 8 or grade 3	0.96*** (0.19)	0.76** (0.24)	-0.20 (0.19)	0.64** (0.22)	83.6** (30.6)	56.8** (18.6)

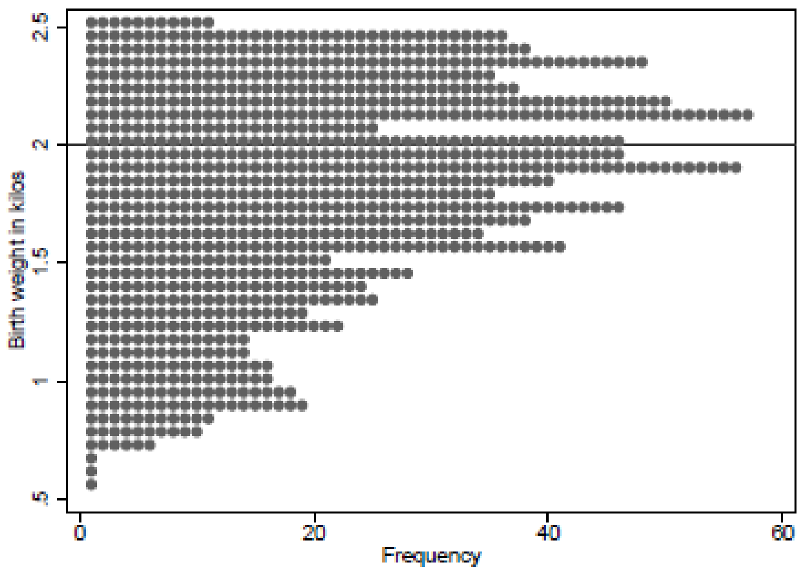
Notes: “Low income” is defined as having family income < 180 percent of the U.S. poverty line. Estimates in the table come from applying ECLS-B-based weights to IHDP data.

Table 7

Bounds on key parameters from partial identification analysis over all possible combinations of values for observations with missing low-income status

Parameter	High-low income gap: b1		IHDP treatment effect for non low-income subsample: b2		IHDP treatment* Low-income interaction: b3		Percent of gap closed from universal program: Cb		Percent of gap closed from targeted program: Cb	
	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Max.
Age-1 IQ	-0.03	0.01	-0.12	0.18	-0.08	0.41	-3.7M	8.9M	-3.3M	8.2M
Age-2 IQ	1.05	1.13	0.09	0.55	0.66	1.31	60.7	122.1	108.9	131.1
Age-3 IQ	1.29	1.45	-0.04	0.48	1.06	1.75	79.5	125.7	110.1	129.0
Age-5 IQ	0.71	0.80	-0.48	-0.16	0.65	1.12	88.2	146.2	65.1	85.0
Age-8 IQ	0.70	0.82	-0.36	0.01	0.37	0.86	50.1	110.8	50.4	70.8
Age-8 Reading	0.74	0.89	-0.89	-0.34	0.57	1.26	73.8	158.9	29.4	56.5
Age-8 Math	0.70	0.81	-0.59	-0.13	0.51	1.15	70.0	153.1	50.5	75.3

Note: confidence intervals on bounds are 95 percent percentile confidence intervals. The estimated bounds on percent of gap closed for age-1 IQ are measured in millions of percentage points (M). All other estimates in the last four columns are in percentage points.



Note: High (low) low birth weight is above (below) 2 kilograms.

Figure 1

Dotplot of birth weight distribution in IHDP sample.

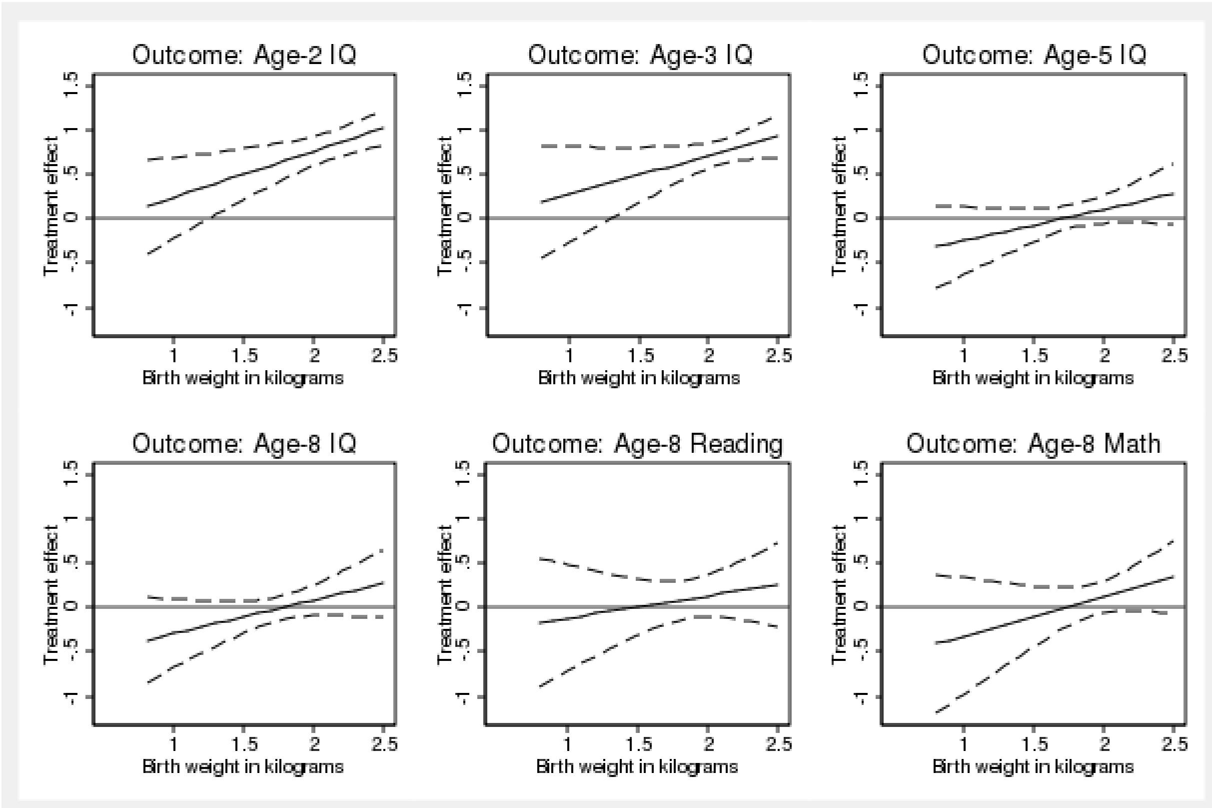


Figure 2

Average marginal effects on IQ and achievement z-scores of treatment interacted with birth weight in IHDP sample.

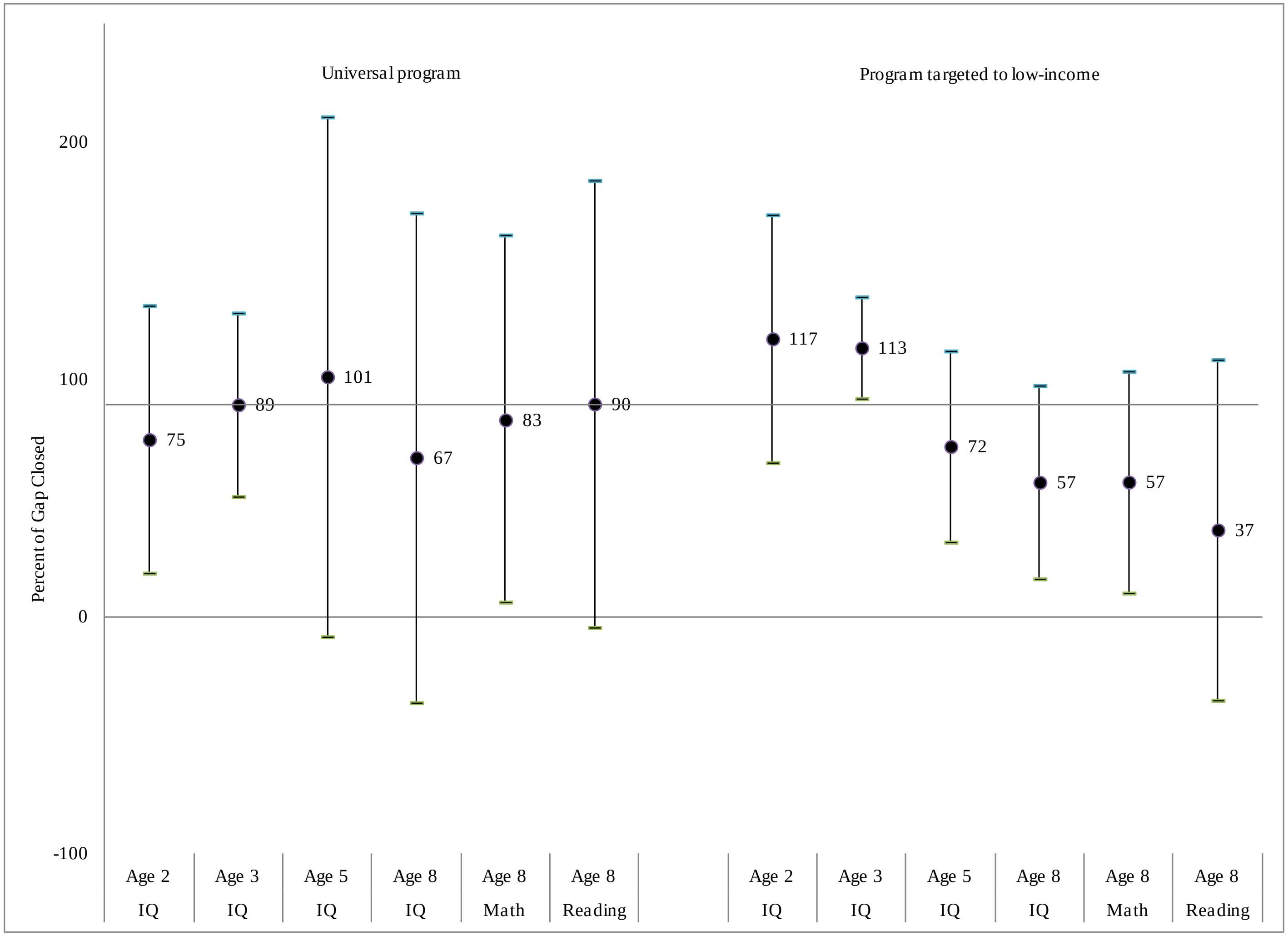


Figure 3

Percent of Cognitive and Achievement Gap Closed by Universal and Targeted IHDP by Age of Child

1. Longer-run impact estimates for Head Start children enrolled two or more decades ago are considerably bigger (Ludwig and Miller 2007; Deming 2009), although employ different identification strategies. Also, the use of center-based care for children in control groups has likely increased considerably, rendering it difficult to generalize from the experiences of older cohorts.

2. Owing to concern over possible attrition bias at the time of the age-18 follow-up, we confine our analysis to cognitive and achievement impacts through age eight.

3. Reardon concludes that the increasing correlation between income and achievement was more important than growth in income inequality for growth in the income-based achievement gap. Another possibility is that the gap, measured in contemporaneous standard deviation units, grew mechanically from a secular decrease in population achievement variance, although growing income-based gaps are also observed for college graduation (Bailey and Dynarski 2011) and years of completed schooling (Duncan and Murnane 2011). In any case, the black-white achievement gap moved in the opposite direction, shrinking by about half over the same period.

4. This stratification into higher and lower birth weight groups was the only baseline interaction specified *ex ante* by the study designers. Based on the medical, developmental, and neurobiological evidence available, they recognized that treatment effects would likely vary between these two strata.

5. Many programs directed at low-income children use a 180 percent-of-poverty income threshold. In addition, it produces two ample-sized income groups in our data.

6. Suppose $Y_1 = cY_2 + d$ so that the ECLS IQ scores (Y_1) are a linear function of IHDP IQ scores (Y_2). Then C_b is less biased than C_m unless c=1, in which case they are the same.

7. For each of the 37 children with a missing income/needs ratio, we impute a low-income indicator using a probit model conditional on a set of fully observed pre-assignment characteristics: maternal age, race, education, number of living children, and previous number of LBW, premature children at time of study child's birth; study child's weight, gestational age, neonatal health index, and parity order at birth; and study site indicators. Identification assumes that low-income status is missing at random conditional on these covariates. Each case has ten imputed replicates. The low-income status of most of these cases appears quite certain on the basis of the baseline observables used for imputation. Consider the frequency of number of replicates imputed low-income among the 37 cases.

Number of replicates imputed low-income	0	1	2	3	4	5	6	7	8	9	10	Total
Frequency	0	1	0	0	2	2	3	6	3	11	9	37

Twenty-one cases (57 percent) have either all or all but one of their 10 replicates imputed to the same low-income status. In only seven cases (19 percent) is the number of replicates imputed as low-income within 2 of the number of replicates imputed as not low-income.

8. We concentrate on impacts through age 8 because the only later follow-up, conducted when the children were 18 years old, successfully interviewed on 61.9 percent of eligible respondents. Results (available on request for IQ and achievement) are very similar to those found at age eight.

9. Developmental trajectories of the lighter and heavier LBW babies differed within the IHDP sample as well. Klebanov, Brooks-Gunn, and McCormick (1994b) study school achievement outcomes among different birth weight strata using a sample that includes both normal and low birth weight children. They find only small differences between the normal birth weight (NBW) (> 2, 500 g) and heavier LBW (1,501 -2,500 g) strata. Differences become pronounced in

comparisons to very LBW (1,001-1, 500 g) and extremely LBW children (< 1,000 g). Since the HLBW sample used below includes only the top half of the heavier LBW range they consider (2,001-2,500 g), differences with normal birth weight children should be even more muted. In another paper, Klebanov, Brooks-Gunn, and McCormick (1994a) compared strata's elementary-school classroom behavior and find even fewer differences between NBW and HLBW children. As part of a broader literature on the cognitive development of low birth weight children, McDonald (1964) tested the IQs of over 1,066 children aged six to nine who weighed less than four pounds at birth. He writes: "When compared with a national sample (of Britain and Wales) matched on social class, the mean I.Q. of 98.4 found in the sample was lower than the expected mean of about 103 in Britain at the present time. But when the 107 children with cerebral palsy, blindness, or deafness were excluded and in addition eleven (1.8 per cent.) children with I.Q.s below 50, which may be considered to be pathologically low, the mean was 102.4. There was thus no evidence that, when children with these handicaps were excluded, the mean I.Q. differed from that of the general population." Jefferis et al. (2002) draw on data from the 1958 British birth cohort to develop evidence on whether normal birth weight and low birth weight children experience roughly parallel developmental trends in math achievement. Of particular relevance to our study, they examine this question separately for children of higher and lower social class, corresponding roughly to children of higher and lower income families in our study. Among children of higher social class, those born LBW experience the same changes in achievement as those born normal birth weight. Among children of lower social class, those born LBW experience similar changes in achievement as those born normal birth weight. The levels of achievement are generally lower for LBW versus normal BW children. However, trends are similar.

10. Despite frequent home visits during the first year of life, there was virtually no average treatment effect on age one IQ scores, perhaps in part because of unreliability inherent in measuring cognitive ability at that age.

11. While not of direct interest in this study, it is worth noting that the treatment effect estimates for children from low-income families obtained from this model are:

Outcome	Coefficient	SE	p-value
Age 2 IQ	1.290	(0.263)	0.002
Age 3 IQ	1.630	(0.262)	0.000
Age 5 IQ	0.586	(0.192)	0.018
Age 8 IQ	0.500	(0.163)	0.018
Age 8 Read.	0.350	(0.257)	0.216
Age 8 Math	0.500	(0.263)	0.099

12. We ignore the age-12-month IQ as there is no evidence of treatment impacts prior to the start of the Abecedarian-type curriculum at age 12 months. ECLS-B estimates of the income-based gaps in age-four PPVT, reading and math, and age-five reading are all about .65. The age-five math gap is .71.

13. To get some intuition for why, note that for an estimator of $\beta = (X'X)^{-1}(X'Y)$, missing Y data enter only the numerator while missing X data enter both the numerator and denominator.

14. Standard errors on the bound estimates can be computed by bootstrap but is computationally very intensive.

15. Gross et al. (1997) report that running IHDP's Miami child care site for the final year (children in their third year of life) cost $15,146 per child. Converting to 2006 dollars using CPI and multiplying by two years puts the cost of the childcare treatment at just over $60,000. This does not include the modest home-visiting portion of the treatment that was offered when children were between ages one and three. Because the costs may be lower in a program that has run for many years, they study the costs of other similar, nonexperimental childcare programs available to children with developmental disabilities. If we apply the same adjustments, the two-year estimates total about $48,000. For a similar program that served children free of disabilities and which did not provide transportation, the two-year cost would be about $29,000.

16. Supplemental analyses failed to point to a clear reason for this. It's not take-up, since black children in the treatment group attended 61 percent of days versus 50 percent for non-blacks. Treatment effects are large for children from low-income families and null for children from higher-income families among both blacks and non-blacks considered separately. Nor does it appear driven by negative correlation between site quality and fraction black. Interactions with race and site show broadly similar pattern across sites, with larger positive effects for whites and smaller effects for blacks. For some yet-to-be discovered reason, negative point estimates for the effect for higher-income blacks offset the positive treatment effect for low-income blacks, diminishing the estimated effect of a black-targeted program.

17. If a universal program were implemented, word about null impacts for higher-income households might lead few of these households to enroll their children in the program. This would eliminate the "negative effect" and imply an upward bias in our estimates of the effectiveness of a universal program for closing gaps in school readiness. There is little evidence

---

of differential take-up of IHDP program services during its operation; while low-income families took up the IHDP's center-based services on a slightly higher fraction of possible days than did higher-income families, the difference is small and far from significant. Nevertheless, these considerations lead us to be more cautious about our universal than targeted estimates. Nonetheless, we present the results, which were included in our *ex ante* study plan.

Online Appendix for Duncan and Sojourner (10/24/2012)

The Early Childhood Longitudinal Study – Birth Cohort (ECLS-B) has followed to kindergarten entry a large, nationally representative sample of children born in 2001. Birth certificates constituted the sampling frame. Children with low birth weight, twin births, and American Indian/Native Alaskan, Asian/Pacific Islander, and Chinese children were oversampled. The ECLS-B provides detailed information on children's development, health, and learning experiences during the years leading up to school entry. Data were collected from sample children and their parents, as well as childcare providers and teachers. Children's cognitive abilities were assessed at 9 months, 24 months, 48 months (preschool), and kindergarten entry using a combination of existing measures and measures developed specifically by the Educational Testing Service.

The base year (nine-month) ECLS-B sample is comprised of about 10,700 infants. Baseline data were collected on a rolling basis between the fall of 2001 and 2002. The two-year-old data collection took place between the fall of 2002 and fall of 2003, with an approximate completed sample size of 9,800 children. The 48-month data collection took place between the fall of 2005 and spring of 2006, with an approximate completed sample size of 8,900. The kindergarten data collection took place during the school year in which the child attended kindergarten, with an approximate completed sample size of 8,000. The weighted response rate for the nine-month data collection, based on the percentage of completed parent interviews out of the total number of eligible cases, was 74.1%. The weighted response for the two-year data collection, based on the percentage of completed parent interviews out of the total number of children eligible to participate at two years, was 93.1%, with 94.0% of children having at least some child assessment data. The weighted response rate for the preschool data collection was 90.8%. The weighted response rate for the kindergarten data collection was 90.0%.

Although the IHDP sample is economically and ethnically diverse, it was not designed so that its demographic characteristics matched those of any larger population. To correct for this, we construct a set of weights based on the relative frequency of observations that fell into cells defined jointly by family income (below or above 180% of the poverty line), race/ethnicity (African-American, Hispanic, or other [mostly white/non-Hispanic]), maternal schooling (no college or at least some college) and marital status (married or not) in the IHDP HLBW subsample and the ECLS-B sample. Low counts led us to combine categories of some college and college graduate for Hispanics. All of these demographic characteristics were measured at nine months in the ECLS-B and at birth in the IHDP, except for family income status, which was measured at 12 months in the IHDP.

Cell frequencies for the two samples and the ECLS/IHDP frequency ratios are given in Appendix Table 1. Relative to the ECLS-B, the IHDP sample overrepresents African-Americans and, to a lesser extent, mothers with low levels of education. The ECLS/IHDP frequency ratios constitute the weights that we apply to most of our IHDP-based analyses. Demographers have long used this kind of sub-classification or post-stratification estimator to calculate mean outcomes in a population adjusted for differences in covariate distributions between a sample and a population (Kitagawa 1964; Cochran 1968). We use it to identify and estimate population average treatment effects given subpopulation treatment effects in an experimental sample and data on differences in the covariate distributions between the experimental sample and population. Generalization from sample treatment effects to population treatment effects is an active area of research in econometrics (Imbens 2004; DiNardo and Lee 2011) and statistics (Hedges and O'Muircheartaigh 2011; Tipton 2012). The crucial assumption is that heterogeneity in treatment effects is characterized by the combination of maternal ethnicity, marital status, and education. Tipton refers to this condition as unconfounded sample selection and states that the set of covariates used for ratio weighting, “must include all covariates that both explain variation in the potential treatment effects and differ in distribution in the population and the experiment. This is a smaller set of variables than those associated with potential outcomes, which is the requirement for matching in observational studies.” Because of the multiple imputation of low-income status in the IHDP, the ratios are allowed to vary across replicates as appropriate.

[Appendix Table 1 here]

Appendix Table 2 presents three sets of descriptive statistics for the IHDP high low birth weight sample. The first describes the unweighted IHDP HLBW sample and shows it to be a relatively disadvantaged group of children both in terms of family characteristics such as income as well as cognitive ability, with the exception of IQ at age one and age-8 achievement. Measuring the cognitive ability of infants is difficult, leading us to place relative little weight on the 9 and 12-month data available in the two data sets. The second set of results presents analogous descriptive statistics after weighting IHDP cases to match the ECLS-B sample on maternal education, ethnicity, and marital status. This is our analytic sample and includes both treatment and control groups. Mean cognitive skills and family income are both higher than in the unweighted IHDP sample.

[Appendix Table 2 here]

The third set of results presents the clearest evidence of how well the weighted IHDP compares to the national population. In this case, descriptive statistics are based only on the IHDP control group cases, which, like the national population, has not been given the IHDP treatment. In the national population, IQ z-scores and standardized math and reading achievement would be mean 0 and standard deviation 1. Up to age 3, IQ means for IHDP controls deviate substantially from national norms. However, at later ages, the reweighted sample means matches closely national norms. At age 5, the IQ z-score mean is -0.03 and standard deviation 1.09. At age 8, they are 0.04 and 1.11. The high age-8 achievement means suggest that the reweighted sample over-represents higher-achieving students.

References

Cochran, William. 1968. “The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies.” *Biometrics* 24(2): 295-313.

DiNardo, John, and David Lee. 2011 “Program Evaluation and Research Designs.” In *Handbook of Labor Economics* 4(A), ed. Orley Ashenfelter and David Card, 463-536. Amsterdam: Elsevier.

Hedges, Larry, and Colm O’Muircheartaigh. 2011. “Improving Generalizations from Designed Experiments.” Northwestern University working paper.

Imbens, Guido. 2004. “Nonparametric Estimation of Average Treatment Effects Under Exogeneity: a Review.” *Review of Economics and Statistics* 86(1): 4-29.

Kitagawa, Evelyn. 1964. “Standardized Comparisons in Population Research.” *Demography* 1(1): 296-315.

Tipton, Elizabeth. 2012. “Improving Generalizations from Experiments Using Propensity Score Subclassification: Assumptions, Properties, and Contexts.” *Journal of Educational and Behavioral Statistic*s. Forthcoming.

Appendix Table 1

Joint distributions over maternal marital status, ethnicity, and education for the ECLS-B sample and the IHDP HLBW sample, and the ratio of their cell densities

Family income status	Ethnicity	Maternal Marital Status & Education: Married No college	Married Any college	Unmarried No college	Unmarried Any college
		IHDP - HLBW sample (n=362)			
Low income	White/other	0.066	0.025	0.044	0.011
	African-Am.	0.052	0.030	0.251	0.033
	Hispanic	0.022	0.006	0.052	
Not low-inc.	White/other	0.036	0.160	0.014	0.008
	African-Am.	0.019	0.017	0.019	0.011
	Hispanic	0.006	0.008	0.006	
Missing Income	White/other	0.014	-	0.022	0.003
	African-Am.	0.006	0.003	0.033	0.006
	Hispanic	0.003	-	0.014	
		ECLS-B sample (n=8840)			
Low income	White/other	0.072	0.053	0.069	0.024
	African-Am.	0.011	0.009	0.062	0.016
	Hispanic	0.061	0.018	0.077	
Not low-inc.	White/other	0.064	0.311	0.024	0.022
	African-Am.	0.005	0.018	0.006	0.010
	Hispanic	0.018	0.031	0.018	
		Mean (standard deviation) ECLS:IHDP ratio across 10 imputed replicates			
Low income	White/other	0.92 (0.02)	2.14	1.17 (0.06)	1.99 (0.22)
	African-Am.	0.18 (0.00)	0.28 (0.01)	0.22 (0.00)	0.43 (0.02)
	Hispanic	2.48 (0.09)	3.32	1.21 (0.18)	
Not low-inc.	White/other	1.66 (0.08)	1.94	1.16 (0.16)	2.21 (0.32)
	African-Am.	0.25 (0.00)	0.96 (0.07)	0.25 (0.02)	0.81 (0.11)
	Hispanic	3.09 (0.39)	3.69	3.58 (0.88)	

Note: Some cells are collapsed owing to small case counts.

Appendix Table 2

Summary statistics for IHDP HLBW sample

IHDP HLBW Sample: Weighted:		All No		All ECLS-B		Control only ECLS-B	
Variable	N	Mean	SD	Mean	SD	Mean	SD
Age 1 IQ z-score	330	0.78	1.04	0.88	1.01	0.86	0.99
Age 2 IQ z-score	322	-0.04	1.36	0.23	1.42	-0.03	1.39
Age 3 IQ z-score	328	-0.68	1.32	-0.29	1.44	-0.60	1.42
Age 5 IQ z-score	295	-0.43	1.17	0.03	1.18	-0.03	1.09
Age 8 IQ z-score	311	-0.38	1.16	0.09	1.16	0.04	1.11
Age 8 reading achvmt.	308	0.02	1.36	0.42	1.36	0.46	1.35
Age 8 math achvmt.	312	-0.03	1.41	0.27	1.40	0.25	1.38
White/other	362	0.40		0.63		0.61	
African-American	362	0.48		0.13		0.14	
Hispanic	362	0.12		0.23		0.24	
Mother without any college	362	0.67		0.47		0.41	
Mother with any college	362	0.33		0.53		0.59	
Mother married	362	0.47		0.66		0.70	
Mother not married	362	0.53		0.34		0.30	
Income/needs ratio at 12 months	325	1.86	1.88	2.65	2.19	2.70	2.15
Income/needs < 1.8 indicator	325	0.66		0.47		0.45	
Male	362	0.52		0.53		0.55	
Birth weight, kilograms	362	2.26	0.14	2.26	0.14	2.26	0.14
Gestational age at birth	362	34.9	1.54	34.7	1.57	34.9	1.48
Neonatal health index	362	99.0	14.8	96.9	14.0	97.0	14.1
Maternal age at birth	362	24.6	6.07	26.9	6.09	27.3	6.28
Number prior premature, LBW births	362	0.31	0.62	0.22	0.54	0.20	0.50
Children living	362	1.96	1.12	1.85	1.07	1.90	1.14
Parity/birth order	362	1.99	1.18	1.87	1.09	1.89	1.13

Appendix Table 3

Treatment effects on IQ z-score at ages 1, 2 and 3 by low-income status using IHDP HLBW sample with ECLS-B weights

DV:		Age 1 IQ			Age 2 IQ			Age 3 IQ	
Model:	A	B	C	A	B	C	A	B	C
Treatment	0.109	0.112	0.065	0.793***	0.878***	0.433*	0.903***	1.001***	0.323
	(0.132)	(0.133)	(0.177)	(0.160)	(0.223)	(0.219)	(0.147)	(0.181)	(0.210)
Low income		-0.037	-0.072		-0.875***	-1.181***		-1.017***	-1.482***
		(0.122)	(0.171)		(0.244)	(0.270)		(0.192)	(0.240)
Treatment x			0.097			0.872**			1.319***
(Low income)			(0.253)			(0.280)			(0.308)
I(Male)	-0.127	-0.131	-0.136	-0.210	-0.275	-0.333	-0.016	-0.097	-0.173
	(0.212)	(0.216)	(0.216)	(0.209)	(0.220)	(0.222)	(0.054)	(0.092)	(0.106)
Gest. age at	-0.046*	-0.045*	-0.046*	0.035	0.059	0.051	0.031	0.064	0.048
birth	(0.023)	(0.022)	(0.022)	(0.101)	(0.096)	(0.091)	(0.104)	(0.100)	(0.092)
Birth weight	1.268***	1.271***	1.283***	-0.017	0.018	0.100	0.151	0.195	0.318
(kg)	(0.360)	(0.361)	(0.361)	(0.758)	(0.819)	(0.875)	(0.526)	(0.672)	(0.692)
Neonatal Health	0.008*	0.008*	0.008*	0.001	0.004	0.002	-0.003	0.001	-0.002
Index	(0.004)	(0.004)	(0.004)	(0.008)	(0.007)	(0.008)	(0.006)	(0.006)	(0.005)
Obs.	330	330	330	322	322	322	328	328	328

Coefficient significance (within site correlation corrected standard errors): $*$: 10% $**$: 5% $***$: 1%.

All models include site dummies

.

Appendix Table 4

Treatment effects on IQ z-score at ages 5, 8 and 18 by low-income status using IHDP HLBW sample with ECLS-B weights

DV:		Age 5 IQ			Age 8 IQ	
Model:	A	B	C	A	B	C
Treatment	0.102	0.148	-0.264	0.156	0.224	-0.067
	(0.116)	(0.166)	(0.201)	(0.158)	(0.169)	(0.323)
Low income		-0.509*	-0.820***		-0.595**	-0.806***
		(0.246)	(0.231)		(0.185)	(0.196)
Treatment x (Low income)			0.861***			0.572
			(0.201)			(0.361)
I(Male)	-0.060	-0.112	-0.158	0.025	-0.027	-0.063
	(0.117)	(0.139)	(0.140)	(0.131)	(0.153)	(0.163)
Gest. age at birth	0.033	0.051	0.042	0.083	0.102	0.095
	(0.073)	(0.068)	(0.066)	(0.059)	(0.060)	(0.063)
Birth weight, kg	0.582	0.620	0.630	0.364	0.400	0.431
	(0.629)	(0.690)	(0.724)	(0.652)	(0.643)	(0.696)
Neonatal Health Index	-0.005	-0.003	-0.006	-0.010*	-0.008	-0.009*
	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.005)
Obs.	295	295	295	311	311	311

Coefficient significance (within-site-correlation corrected standard errors): ∗: 10% ∗∗: 5% ∗∗∗: 1%.

All models include site dummies.

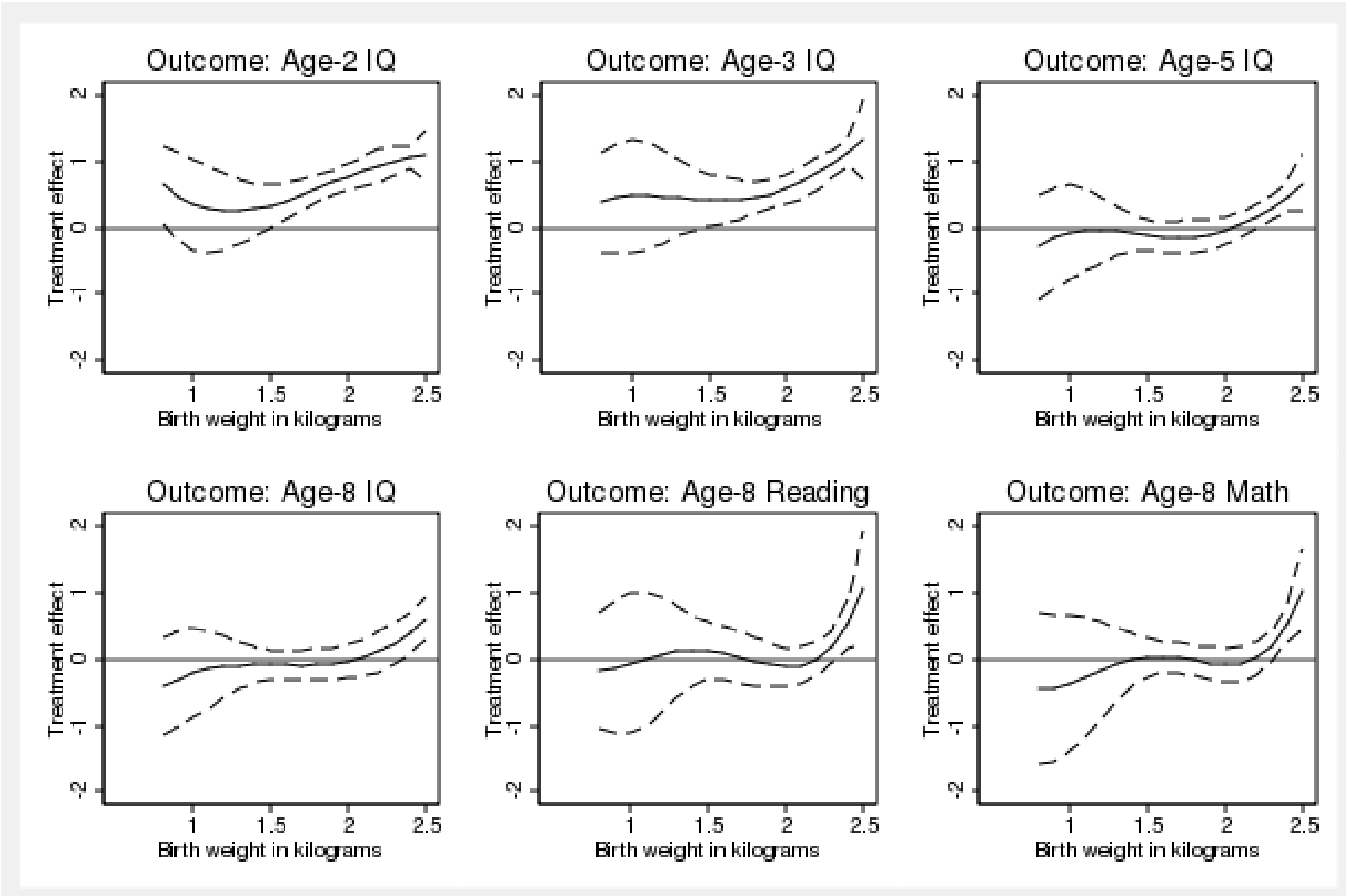
Appendix Table 5

Treatment effects on age 8 Woodcock-Johnson achievement by low-income status using IHDP HLBW sample with ECLS-B weights

DV:	Reading			Math		
Model:	A	B	C	A	B	C
Treatment	-0.116	-0.041	-0.456	0.120	0.187	-0.137
	(0.209)	(0.261)	(0.267)	(0.149)	(0.206)	(0.197)
Low income		-0.643***	-0.936***		-0.594*	-0.830**
		(0.156)	(0.123)		(0.257)	(0.281)
Treatment x (Low income)			0.804***			0.636**
			(0.184)			(0.224)
I(Male)	-0.393**	-0.455***	-0.498***	-0.026	-0.078	-0.118
	(0.120)	(0.129)	(0.139)	(0.152)	(0.181)	(0.181)
Gest. age at birth	0.101	0.123*	0.111	0.142*	0.162*	0.153*
	(0.065)	(0.061)	(0.063)	(0.071)	(0.076)	(0.078)
Birth weight, kg	-0.419	-0.381	-0.350	0.376	0.411	0.445
	(0.661)	(0.634)	(0.688)	(0.585)	(0.565)	(0.587)
Neonatal Health Index	-0.009	-0.007	-0.008	-0.007	-0.005	-0.006
	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)	(0.007)
Obs.	308	308	308	312	312	312

Coefficient significance (within site correlation corrected standard errors): ∗: 10% ∗∗: 5% ∗∗∗: 1%.

All models include site dummies.



Appendix Figure 1

Average marginal effects on IQ and achievement z-scores of treatment interacted with fourth-order polynomial of birth weight in IHDP sample.