Early Intervention in Low Birth Weight Premature Infants: Results at 18 Years of Age for the Infant Health and Development Program

Marie C. McCormick, Jeanne Brooks-Gunn, Stephen L. Buka, Julie Goldman, Jennifer Yu, Mikhail Salganik, David T. Scott, Forrest C. Bennett, Libby L. Kay, Judy C. Bernbaum, Charles R. Bauer, Camilia Martin, Elizabeth R. Woods, Anne Martin and Patrick H. Casey

Pediatrics 2006;117;771-780
DOI: 10.1542/peds.2005-1316

ARTICLE

Early Intervention in Low Birth Weight Premature Infants: Results at 18 Years of Age for the Infant Health and Development Program

Marie C. McCormick, MD, ScD[a], Jeanne Brooks-Gunn, PhD[b], Stephen L. Buka, ScD, MS[a], Julie Goldman, MSSA, MS[a], Jennifer Yu, ScD, EdM[a], Mikhail Salganik, PhD[c], David T. Scott, PhD[d], Forrest C. Bennett, MD[e], Libby L. Kay, MSSW, LCSW[f], Judy C. Bernbaum, MD[g], Charles R. Bauer, MD[h], Camilia Martin, MD, MS[i], Elizabeth R. Woods, MD, MPH[j], Anne Martin, DrPH, MPH[b], Patrick H. Casey, MD[k]

Departments of [a]Society, Human Development and Health, and [c]Biostatistics, Harvard School of Public Health, Boston, Massachusetts; [b]National Center for Children and Families, Teachers College, and College of Physicians and Surgeons, Columbia University, New York, New York; Departments of [d]Psychiatry and Behavioral Sciences and [e]Pediatrics, University of Washington, Seattle, Washington; [f]Department of Psychiatry, University of Texas Southwestern Medical Center, Dallas, Texas; [g]Department of Pediatrics, University of Pennsylvania School of Medicine, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania; [h]Department of Pediatrics, University of Miami School of Medicine, Miami, Florida; [i]Department of Neonatology, Beth Israel Deaconess Medical Center, Boston, Massachusetts; [j]Children's Hospital Boston, Boston, Massachusetts; [k]Department of Pediatrics, University of Arkansas for Medical Sciences, Little Rock, Arkansas

The authors have indicated they have no financial relationships relevant to this article to disclose.

ABSTRACT

OBJECTIVE. To assess whether improvements in cognitive and behavioral development seen in preschool educational programs persist, we compared those in a multisite randomized trial of such a program over the first 3 years of life (INT) to those with follow-up only (FUO) at 18 months of age.

METHODS. This was a prospective follow-up of the Infant Health and Development Program at 8 sites heterogeneous for sociodemographic characteristics. Originally 985 children were randomized to the INT (n = 377) or FUO (n = 608) groups within 2 birth weight strata: heavier low birth weight (HLBW; 2001–2499 g) and lighter low birth weight (LLBW; ≤2000 g). Primary outcome measures were the Peabody Picture Vocabulary Test (PPVT-III), reading and mathematics subscales of the Woodcock-Johnson Tests of Achievement, youth self-report on the Total Behavior Problem Index, and high-risk behaviors on the Youth Risk Behavior Surveillance System (YRBSS). Secondary outcomes included Weschler full-scale IQ, caregiver report on the Total Behavior Problem Index, and caregiver and youth self-reported physical health using the Medical Outcome Study measure. Assessors were masked as to study status.

RESULTS. We assessed 636 youths at 18 years (64.6% of the 985, 72% of whom had not died or refused at prior assessments). After adjusting for cohort attrition, differences favoring the INT group were seen on the Woodcock-Johnson Tests of Achievement in math (5.1 points), YRBSS (−0.7 points), and the PPVT-III (3.8 points) in the HLBW youth. In the LLBW youth, the Woodcock-Johnson Tests of Achievement in reading was higher in the FUO than INT group (4.2).

CONCLUSIONS. The findings in the HLBW INT group provide support for preschool education to make long-term changes in a diverse group of children who are at developmental risk. The lack of observable benefit in the LLBW group raises questions about the biological and educational factors that foster or inhibit sustained effects of early educational intervention.

www.pediatrics.org/cgi/doi/10.1542/peds.2005-1316

doi:10.1542/peds.2005-1316

Key Words
early intervention, long-term results, low birth weight, randomized controlled trial

Abbreviations
IHDP—Infant Health and Development Program
LLBW—lighter low birth weight
HLBW—heavier low birth weight
INT—intervention
FUO—follow-up only
YRBSS—Youth Risk Behavior Surveillance System
BPI—Behavior Problem Index
WASI—Weschler Abbreviated Scale of Intelligence
PPVT-III—Peabody Picture Vocabulary Test-Version III
HSG—high school graduate
CI—confidence interval
LBW—low birth weight

Accepted for publication Aug 12, 2005

Address correspondence to Marie C. McCormick, MD, ScD, Department of Society, Human Development and Health, Harvard School of Public Health, 677 Huntington Ave, Boston, MA 02115. E-mail: mmccormi@hsph.harvard.edu

INTENSIVE EARLY EDUCATIONAL interventions have been documented to improve the cognitive outcomes[1] and, in some cases, reduce antisocial behavior early in the school experience.[2] In several single-site, randomized studies of center-based interventions, longer-term follow-up into adolescence and early adulthood has revealed persistence of early-intervention effects seen as greater school achievement,[3,4] less grade retention and special education,[3,5] and more high school completion[6] and participation in college.[4] In addition to better educational outcomes, early-intervention programs have also resulted in better social outcomes such as less adolescent parenting[4] and less delinquency and fewer arrests.[6–8] Parallel effects have been seen in follow-up of at least 1 home-visiting program[9] and a nonrandomized program conducted in several centers in 1 city, the Chicago Child-Parent Center Program.[10] However, all of these studies involve healthy, low-income, poor children generally from a single site. Thus, inferences about the generalizability of these results to groups more heterogeneous for socioeconomic status and health are limited.

The Infant Health and Development Program (IHDP) was a multisite, randomized, controlled trial of an educational intervention until 3 years of age for low birth weight (LBW) preterm infants. Infants were sampled from 2 birth weight strata (≤2000 g and 2001–2500 g) to assure an oversample of those at higher developmental risk (the lighter LBW [LLBW] group) and continuity with studies of poor, normal birth weight children (the heavier LBW [HLBW] group) at 8 sites, and as a result, the sample for this study is quite heterogeneous for socioeconomic status and health status, particularly neurologic factors that might influence school outcomes. The intervention consisted of home visits every week for the first year and every other week in the second and third year, daily center-based education beginning at 12 months corrected for duration of gestation continuing until 36 months of age, with the addition of parent support groups. All of the children received intensive pediatric follow-up care. The program has been described in detail elsewhere.[11–13]

The results of the trial through 8 years of age have been published previously[14–17] and will be summarized briefly here. The assessment of outcomes was conducted in 3 prior phases. At 36 months of age corrected for duration of gestation, at the end of phase I (the trial phase), the IQs of those who received the intervention (INT) were substantially higher than those who received follow-up services only (FUO): a 14-point difference for the HLBW group and 7 points for the LLBW group. Likewise, lower behavior problem scores were seen in the INT group. By age 5 (phase II, the preschool period) and age 8 (phase III, the early school period), no differences in behavior scores were observed. At both ages 5 and 8, the INT in the HLBW group still experienced a significant advantage in IQ scores (4 points higher), as well as mathematics achievement testing (4 points higher). No differences in IQ or achievement between INT and FUO groups were seen for the children in the LLBW group.

The IHDP is based on the design and curriculum of the Abecedarian Project,[18] and the results in IQ scores and achievement testing parallel results seen in the Abecedarian Project[18] at 8 years of age. More recent data from that project reveal persisting differences in IQ and achievement up to 21 years of age,[19] prompting us to extend the follow-up of the IHDP to 18 years with a fourth wave of data collection (phase IV). The purpose of this effort is to ascertain whether such persistent differences would be observed in the IHDP, especially in the HLBW group. In line with other studies of normal birth weight children, we hypothesized that we would see persistent IQ and achievement differences and less risky behavior favoring the INT group.

METHODS

Study Population

Detailed information on the recruitment of the study population and follow-up through 8 years of age has been published previously[11–17,19] and will be summarized briefly. To be eligible for the trial, an infant must have been born in 1 of the participating hospitals at each of the 8 sites at a birth weight ≤2500 g, a gestational age of ≤37 weeks, and reside in a catchment area defined by distance from the early educational center. Infants were enrolled between October 1984 and August 1985 in 2 birth weight strata: ≤2000 g, and 2001–2500 g. Within birth weight strata, infants were randomly assigned to the INT group (n = 377) or FUO group (n = 608). Although the sites were heterogeneous for a variety of characteristics, the randomization procedure resulted in comparability between the INT and FUO groups at study entry.[14]

Infants in both arms of the study received periodic medical, developmental, and social service assessments with referral for community services when indicated. Those in the INT group received an educational program delivered through home visits (weekly during the first year and every other week during the second and third years of life), a daily center-based program beginning at 12 months corrected for duration of gestation, and parent support groups coinciding with the start of the center-based program.[11] These parent support groups met every other month. At the end of the INT period at 3 years of age, the children received whatever community education programs were available at the site. Although the availability of such services varied greatly among the sites, there were no differences between the INT and FUO groups in receipt of such services.[15] As noted above, major assessment points for the intervention occurred at 3, 5, and 8 years of age. Completion rates at these ages

were 92%, 82%, and 89%, respectively. The lower rate at age 5 reflects the lack of funding for staff to assess children who were not close to any study site.

Data Collection at 18 Years of Age (Phase IV)

Measures

The phase IV assessment at 18 years of age was designed to obtain information in the areas of academic achievement, behavior, cognitive abilities, and physical health. To the extent possible, instruments were selected to provide continuity with previous assessment periods and to have known references or to have been used in large population studies.

Academic Achievement

Academic achievement was assessed using the Woodcock-Johnson Tests of Achievement-Revised.[20] These tests yielded letter-word identification, passage comprehension, calculation, and applied problems subtests of the standard scores. Two scores were calculated, the Woodcock-Johnson Tests of Achievement in reading and in mathematics, referenced to a mean of 100 and an SD of 15, with higher scores indicative of better achievement.

In addition, the youth's mother or other caregiver was asked about his/her progress in school. Specifically, these questions deal with completion of school, grade repetition, classification for special education, and school dropout.

Behavior

Two aspects of behavior were assessed: risk behaviors and more general behavioral problems. Risk behaviors were assessed using youth self-report on questions taken from the Youth Risk Behavior Surveillance System (YRBSS).[21] Items from this survey were selected to characterize the youth along dimensions affected by early intervention in early studies: conduct problems, suicidal ideation/attempts, smoking, alcohol and marijuana use, and risky sexual activity (see the Appendix for the scoring algorithm). More general behavioral problems were ascertained using the Behavior Problem Index (BPI),[22] completed by both the youth and the primary caregiver (the latter for comparison with earlier assessments). For both the YRBSS and the BPI, higher scores indicate more problems.

Finally, we asked the caregiver and the youth specific questions about youth involvement with the justice system. These questions included items on trouble with the police, arrests, and time in jail. A positive event was coded if either reported it.

Cognitive Development

Two measures of cognitive development were administered: the Weschler Abbreviated Scale of Intelligence (WASI)[23] and the Peabody Picture Vocabulary Test-Version III (PPVT-III).[24] The latter, a measure of receptive vocabulary, has been used in this sample at each assessment point since 3 years of age and was designated the primary outcome because of this continuity over time. Both tests are referenced to a mean of 100 and an SD of 15. The WASI provides 4 subscales (vocabulary, block design, similarities, and matrix reasoning) and can generate a verbal IQ and performance IQ, as well as a full IQ score, with only the last being considered here.

Physical Health

Physical health was characterized by youth self-report and caregiver report on the 36-Physical Component Summary measure.[25] This scale reflects limitations in activities of daily living because of health, as well as positive health. It is referenced to a mean score of 50 and an SD of 10, with higher scores indicative of better health.

Of these measures, math and reading achievement tests, the risk behavior scale, youth self-report on the BPI, and the PPVT-III were considered the primary outcomes for the analysis. However, the results for the WASI, caregiver-reported behavior problem score, and self-report and caregiver report on physical health are provided for continuity with earlier reports.

Baseline Measures

As in previous analyses of the IHDP, a standard set of prerandomization variables was used to compare participants and nonparticipants in this assessment and to adjust the outcomes for differences across sites and birth weight strata. These variables included birth weight in grams, maternal age in years, race/ethnicity (black, Hispanic and white/other), gender of child (male/female), maternal educational attainment at the time of the birth of the child (less than high school graduate [HSG], HSG, and more than HSG), and Neonatal Health Index,[26] a measure of neonatal length of stay in the hospital relative to birth weight, which serves as an indication of the severity of the neonatal course. This index is referenced to a mean of 100 and SD of 16. Baseline variables also included study site: Arkansas (Little Rock, AK), Einstein (South Bronx, NY), Harvard (Boston, MA), Miami (Miami, FL), Pennsylvania (Philadelphia, PA), Texas (Dallas, TX), Washington (Seattle, WA), and Yale (New Haven, CT). Gestational age at birth was also available.

Assessments at Age 8

In addition, we compared participants and nonparticipants for the cognitive, achievement, and behavioral assessments done at 8 years of age. These measures included the Wechsler full-scale IQ score,[27] achievement scores in math and reading,[20] the PPVT-Revised,[28] and behavior problem scores.[29]

Analysis

Participants and nonparticipants at the 18-year assessment were compared on the prerandomization variables and 8-year outcomes noted above. A subject was considered a participant if any of the outcomes of interest

were observed at the 18-year visit and as a nonparticipant otherwise. The comparison of the differences between participants and nonparticipants in the INT and FUO groups were evaluated by using the *t* statistic for continuous measures and χ^2 statistics for categorical measures.

To compare the INT and FUO groups on the outcomes of interest at age 18, separate linear (analysis of covariance) models were developed for each outcome variable within each of 2 birth weight strata using the S-plus software described by Venables and Ripley.[30] The same set of 7 baseline variables measured at enrollment as noted above was used in each model of the comparison of the INT versus FUO. Multivariate models were used to estimate the differences between the INT and FUO groups to adjust for differences in sample sizes among the sites. [14] The linear models were used for estimation of population marginal mean values and differences (INT versus FUO) for each outcome. The estimated marginal means were obtained by substituting into the linear model the average values of the covariates for a particular birth weight stratum. Nominal *P* values derived from the linear models were used for primary outcome comparisons.

We evaluated the robustness of these conventional estimates of the difference between the INT and FUO groups by comparing them with the results of the analysis of covariance where each of the respondents was weighted by the inverse of the outcome-dependent estimated propensity score[31] to assess the effect of attrition. We assumed that the propensity score (which, in our analysis, is the outcome-dependent probability of observing an outcome of interest at the 18-year visit) depended on the values of the full-scale IQ and child behavior checklist at the 8-year visit and baseline characteristics of a subject. Thus, for example, subjects from Harvard with low IQs at age 8 were less likely to participate at age 18, and the average participant with these characteristics would be assigned higher weights (more influence) in the propensity-adjusted analyses of the difference between the INT and FUO groups using this approach. Under the modeling assumptions, this method removes nonresponse bias.

Human Subjects

The study was approved by the institutional review boards of all participating institutions. In addition, certification of confidentiality was obtained for this phase. Written informed consent was obtained separately for all of the participants: study subject and caregiver.

RESULTS

Comparison of Participants and Nonparticipants at 18 Years of Age

At age 18, 636 (64.6%) of the 985 of the participants in the original IHDP trial were assessed: 67.4% of the INT group and 62.8% of the FUO group (odds ratio: 1.22; $P = .15$; 95% confidence interval [CI]: 0.93 to 1.60). This represents 72% of those eligible to be followed (ie, those who had not died or refused additional participation by 8 years of age). Refusal rates were low (2.7% of the 338 in the INT group and 4.6% of the 540 in the FUO group seen at age 8). There was 1 death (in the INT group), which occurred after the assessment was scheduled but before it could be completed. Because the family wished to participate in this round and provided both caregiver and sibling responses, this individual was considered in the participant category. The major reason for lack of response was the inability of the site teams to locate the subjects, or subjects repeatedly missed appointments. As seen in Table 1, those seen at 18 years of age did not differ from the original sample in the baseline characteristics except for maternal educational attainment at the time of enrollment, race/ethnicity, and study site. In the INT group, mothers of lower educational attainment and Hispanic ethnicity were less likely to participate in phase IV. These factors also influenced participation in the HLBW INT group with the additional lower participation among males (41.5% among participants vs 62.5% among nonparticipants; $P = .028$). Within the LLBW INT group, only race/ethnicity was a significant predictor. Race/ethnicity and male gender were also associated with participation status in the HLBW FUO group. The effect of study site was seen in all of the groups overall and within birth weight strata. The Einstein, Harvard, and Washington sites had relatively lower follow-up rates, and Arkansas, Pennsylvania, and Yale sites were relatively higher. When participants and nonparticipants were compared for the assessments at age 8 (Table 2), participants in the INT group had a somewhat higher IQ, PPVT-III, and reading and math achievement and lower behavior problem (ie, fewer problems) scores than nonparticipants. In the FUO group, participants had fewer behavior problems than nonparticipants.

Comparison of INT and FUO Groups on Primary and Secondary Outcomes

Because the initial trial was designed and powered to permit comparisons within birth weight strata, and because subjects were sampled differently in each birth weight stratum, we have displayed the results separately for the HLBW and LLBW groups (Table 3). As in the previous assessments at 5 and 8 years of age, no statistically significant differences were seen between INT and FUO groups in the group weighing ≤2000 g in comparisons adjusted for the baseline variables only (Table 3 line "a" for each outcome), although a difference in reading achievement scores favoring the FUO group achieved $P = .08$. When the results were also adjusted for attrition (Table 3, line "b"), the difference in reading favoring the FUO group was strengthened.

Although the use of the covariates was considered

TABLE 1 **Comparison of Participants and Nonparticipants in Phase IV (18 Years) According to Baseline Characteristics and Study Arm**

Characteristic	INT		FUO	
	Participant	Nonparticipant	Participant	Nonparticipant
Total	254	123	382	226
% of total	67.4	32.6	62.8	37.2
Birth weight, g				
Mean (SD)	1817 (441)	1824 (436)	1795 (465)	1758 (473)
P	.89		.35	
≤2000 g, %	63.0	61.0	64.1	62.8
>2000 g, %	37.0	39.0	35.9	37.2
P	.79		.81	
Gestational age, wk				
Mean (SD)	33.1 (2.4)	33.0 (2.8)	33.1 (2.7)	32.9 (2.9)
P	.72		.37	
NHI				
Mean (SD)	100.6 (15.1)	100.8 (17.7)	99.4 (15.1)	100.0 (17.1)
P	.93		.66	
Maternal age, y				
Mean (SD)	24.6 (5.8)	24.7 (6.2)	25.1 (6.2)	24.6 (6.0)
P	.88		.34	
Maternal education				
<High school, %	36.6	56.1	35.9	42.0
High school, %	30.3	22.0	28.3	25.7
>High school, %	33.1	22.0	35.9	32.3
P	<.01		.32	
Race/ethnicity				
Black, %	50.8	58.5	53.7	49.1
Hispanic, %	7.5	14.6	9.4	14.2
White/other, %	41.7	26.8	36.9	36.7
P	<.01		.18	
Male, %	47.2	56.9	47.1	49.1
P	.10		.70	
Study site (% of total)				
Arkansas	14.6	8.9	17.5	5.8
Einstein	11.0	14.6	13.1	18.6
Harvard	8.7	18.7	11.5	21.7
Miami	11.8	11.4	9.9	8.0
Pennsylvania	15.4	7.3	12.0	3.1
Texas	10.2	18.7	12.3	18.1
Washington	13.0	14.6	10.2	18.1
Yale	15.4	5.7	13.4	6.6
P	<.01		<.001	

NHI indicates Neonatal Health Index.[27]

appropriate because of the differences in the original sample sizes among sites, the effect of this adjustment was minor. For example, the unadjusted mean PPVT-III scores for the LLBW INT and FUO groups were 94.1 and 96.1, respectively, compared with 93.7 and 96.1 in Table 3.

In contrast, among those >2000 g, there were statistically significant elevations in PPVT-III and math achievement scores favoring the INT group in the results adjusted for the baseline covariates. The INT group experienced 5- and 6-point higher reading and math achievement scores, respectively, with the latter at $P = .01$. No statistically significant difference was seen in self-reported behavior problems or the risk behavior scale from the YRBSS in these analyses. When the comparisons were adjusted for attrition, the difference in reading achievement diminished but that for the math achievement remained statistically significant. A difference was seen favoring the INT group on the YRBSS score as having fewer risky behaviors ($P = .05$). When the INT and FUO groups were compared on the measures of secondary outcomes, no statistically significant differences were seen, although the difference favoring the INT group on the full-scale IQ score among the HLBW group (Table 4) had a *P* value of .07.

Fig 1 provides a summary of the published outcomes, where comparable, from all 4 of the assessment periods. In the HLBW group, the 14-point difference seen at age 3 diminished to ~4 points, but this difference has remained stable and consistent over time. Likewise, the differences in PPVT scores favoring the INT have persisted after age 3. Changes in the achievement scores in

TABLE 2 **Comparison of Participants and Nonparticipants in Phase IV (18 years) According to Cognitive, Behavioral, and Achievement Scores at Age 8**

Variable	INT		FUO	
	Participant	Nonparticipant	Participant	Nonparticipant
Full-scale IQ				
Mean (SD)	87.5 (29.2)	79.1 (23.3)	86.5 (27.8)	83.4 (26.6)
P		.02		.22
PPVT-III				
Mean (SD)	86.1 (23.9)	80.9 (19.6)	86.2 (23.1)	83.7 (22.3)
P		.07		.26
W-J reading				
Mean (SD)	99.6 (20.5)	93.1 (22.2)	98.1 (21.2)	95.9 (21.1)
P		.01		.30
W-J math				
Mean (SD)	97.7 (23.1)	90.3 (23.3)	96.4 (21.9)	93.5 (22.1)
P		.01		.16
CBCL total				
Mean (SD)	30.6 (20.7)	33.7 (19.7)	30.1 (18.5)	34.1 (20.8)
P		.23		.02

W-J indicates Woodcock-Johnson Achievement Test[19]; CBCL, Child Behavior Checklist.[30]

this group were generally parallel while exhibiting the expected decline in achievement scores with aging into the late teen years. The figure also reveals that results from the current phase suggesting that small differences favoring the FUO group in PPVT, IQ, and math achievement scores also emerged early and persisted.

To examine whether the results for those <2000 g at birth might reflect a heavier weighting of children with significant neurologic complications, we reran the analyses for those children who had an IQ >85 at age 1 year. Restriction of the analysis to this group led to estimate differences in reading scores (−3.1 [95% CI: −7.1 to 0.09], $P = .13$, adjusted for baseline covariates; −2.8 [95% CI: −6.9 to 1.2], $P = .17$, for the propensity-weighted estimates) that were not statistically significant.

We also explored differences in specific behavioral and school outcomes. Among the INT LLBW group, 16% had been arrested compared with 20% of the FUO group (with $P = .11$, adjusted for baseline covariates and attrition). The comparable figures for the HLBW group was 21% and 26%, respectively ($P = .56$). Approximately 12% to 13% of both birth weight groups had already been in jail with no differences between the INT and FUO groups. Among the LLBW group, 10% of the INT group and 8% of the FUO group were classified as having dropped out of school ($P = .40$); in the HLBW group, it was 11% and 12%, respectively ($P = .58$). With regard to ever being classified for special education, 30% of INT group and 26% of the FUO group in the LLBW stratum had been so classified ($P = .39$) and 17% and 24% of the HLBW stratum ($P = .12$).

DISCUSSION

The results of this phase of the IHDP suggest a persistent benefit of the intervention for the subset of HLBW participants and absence or even reversal of any intervention effect for the youth born weighing ≤2000 g. These results parallel those seen at 8 years of age.[16] As expected, IQ and achievement scores were lower and behavior problem scores higher than expected population references, reflective of patterns seen among those who were LBW and/or premature at birth.[32]

Our rates of juvenile arrest at age 18 are comparable with those the seen in the Chicago study,[10] but the difference between the INT and FUO groups is smaller, 4% to 5% as compared with 8% in the Chicago study, where the difference did achieve statistical significance. Likewise, our incarceration rates are similar to those reported in the Abecedarian intervention group[4] but somewhat lower than the rates in the control group. The differences in incarceration rates also fail to achieve statistical significance. Comparisons with the Perry Preschool Project cannot be made, because they report the percentage with ≥5 arrests at rates substantially higher than arrest records in other studies.

Comparison of our educational outcomes with other studies is complicated by the fact that our sample has not yet established a post–high school educational pattern in terms of enrollment in higher education as reported in Abecedarian,[4] and even final high school graduation rates are unknown. Our dropout rate is substantially lower than the ~50% seen in the Chicago study,[10] as is the difference between intervention and control groups. Our rates of special education and differences between INT and FUO in the HLBW group are similar to those in the Chicago study, where the difference was highly significant.

In this analysis, we have not examined outcomes by prerandomization variables other than birth weight. This decision reflects the results at age 8 where no difference in effect size is seen at different levels of maternal education in contrast to a hypothesized greater effect of the intervention for children from more disadvantaged backgrounds.[16]

The persistence of an intervention effect in the HLBW group provides substantial reinforcement to the emerging literature on longer-term effects of early educational intervention.[4–9] First, the IHDP was implemented in 8 sites and is modeled on the Abecedarian Project,[4] arguing for replicability of the results. Second, the IHDP has been characterized as methodologically rigorous with strict randomization and masked assessments at major outcome points at all of the follow-up phases and has avoided some of the concerns associated with other projects.[8] In contrast to previous studies, which involved largely poor, healthy children, the IHDP expands the results of early educational intervention in 2 ways. First, the sample is heterogeneous with regard to health, especially neurodevelopmental status, by virtue of it being a sample of preterm, LBW children. Although not all LBW children sustain neurologic injury,[32] being LBW

TABLE 3 **Major Outcomes at Age 18 Years for Youth in the IHDP**

Variable	*n*		Marginal Mean Scores		Difference INT vs FUO Estimated Value (95% CI)	*P*
	INT	FUO	INT	FUO		
Birth weight <2000 g						
PPVT-III						
a	154	234	93.7	96.1	−2.5 (−5.5 to 0.5)	.10
b	152	232	93.0	95.5	−2.4 (−5.4 to 0.5)	.10
Woodcock Johnson						
Reading						
a	153	233	92.5	96.2	−3.7 (−7.9 to 0.4)	.08
b	151	231	91.4	95.6	−4.2 (−8.4 to 0.0)	.05
Math						
a	153	234	87.3	89.4	−2.0 (−5.2 to 1.1)	.21
b	151	231	86.0	88.7	−2.7 (−5.9 to 0.6)	.11
TBPI-Y						
a	145	226	9.8	9.8	0 (−1.2 to 1.2)	.98
b	144	224	9.9	10.0	0 (−1.2 to 1.2)	.99
YRBSS						
a	148	231	2.4	2.2	0.2 (−0.3 to 0.7)	.41
b	147	229	2.4	2.1	0.3 (−0.2 to 0.8)	.29
Birth weight >2000 g						
PPVT-III						
a	91	132	99.6	94.5	5.1 (1.2 to 9.0)	.01
b	89	132	98.8	95.0	3.8 (−0.3 to 7.8)	.07
Woodcock Johnson						
Reading						
a	90	130	100.4	95.0	5.3 (−0.4 to 11.1)	.07
b	88	130	98.9	95.7	3.3 (−2.5 to 9.1)	.27
Math						
a	91	130	95.5	89.4	6.1 (1.6 to 10.6)	.01
b	89	130	94.9	89.8	5.1 (0.6 to 9.7)	.03
TBPI-Y						
a	85	126	10.2	9.3	0.9 (−0.6 to 2.3)	.23
b	83	126	10.4	9.3	1.1 (−0.4 to 2.6)	.15
YRBSS						
a	89	127	2.4	2.9	−0.5 (−1.3 to 0.3)	.19
b	87	127	2.2	3.0	−0.7 (−1.5 to 0.0)	.05

TBPI-Y indicates Total Behavior Problem Index-Youth Self-Report; a, adjusted for all the variables in Table 1 except gestational age; b, adjusted for variables in Table 1 except for gestational age and attrition.

confers increased risk of neurodevelopmental disability that may make academic achievement difficult. For example, in unpublished analyses, children with cerebral palsy in the IHDP experienced no advantage in cognitive function at age 3 from the intervention. Second, the IHDP was more heterogeneous with regard to socioeconomic status than previous studies. We have documented that children of mothers with some college or more education did not benefit from the intervention immediately at the end of the trial.[33] Thus, although the effects sizes are modest, they reinforce the existing literature because they reflect a more heterogeneous study population biased against finding a difference. It should be noted that the pattern of persistence seen in Fig 1, including diminishing performance on achievement testing with age, closely parallels results for the Abecedarian Project.[18]

As in the 8-year assessment, questions might be raised about the significance of a 4-point difference in achievement or the PPVT-III. However, it should be noted that the difference in mean scores also indicates a shift in the entire distribution in favor of the INT group. In the past, we have argued that such a shift would decrease the percentage of children classified as intellectually deficient.[16] Others have also argued the importance of such differences and the potential to alter the distribution of those who will do poorly in school.[34]

The lack of persistence of an effect in the LLBW children is also relevant to this emerging literature in that it raises questions about greater attention to the factors that foster or inhibit sustained effects. The LLBW groups did initially benefit from the intervention as indicated by a 6.6-point difference in IQ scores favoring the INT group at 3 years of age.[14] Moreover, the lack of sustained effect was not because of the experience of very premature or very small infants, because the same effect size is seen among those ≤1500 g and even ≤1000 g in birth weight.[35] However, differences between the INT and FUO groups among the LLBW children disappear by age 5 years.[15] One analysis suggests that the intensity of participation may be a factor in that sus-

TABLE 4 **Secondary Outcomes at Age 18 Years of Youth in the IHDP**

Variable	*n*		Marginal Mean Scores		Difference INT vs FUO Estimated Value (95% CI)	P
	INT	FUO	INT	FUO		
Birth weight <2000 g						
WASI full-scale IQ						
a	154	236	89.4	91.2	−1.8 (−4.5 to 0.9)	.19
b	152	233	88.5	90.7	−2.2 (−5.0 to 0.6)	.12
TBPI-CAR						
a	144	234	9.2	9.1	0.2 (−1.2 to 1.6)	.81
b	142	230	9.9	9.3	0.5 (−0.9 to 2.0)	.46
PCS-CAR						
a	155	238	53.6	54.0	−0.4 (−1.6 to 0.8)	.51
b	153	235	53.3	54.0	−0.6 (−1.9 to 0.6)	.31
PCS-YA						
a	149	234	53.4	54.2	−0.8 (−2.3 to 0.7)	.31
b	148	232	53.4	54.1	−0.8 (−2.3 to 0.8)	.33
Birth weight >2000 g						
WASI full-scale IQ						
a	92	132	94.6	91.3	3.3 (−0.3 to 6.9)	.07
b	90	132	93.7	92.0	1.7 (−2.0 to 5.3)	.38
TBPI-CAR						
a	82	124	8.5	9.3	−0.9 (−2.7 to 0.9)	.34
b	80	123	8.9	9.6	−0.7 (−2.6 to 1.1)	.45
PCS-CAR						
a	86	131	55.5	54.9	0.6 (−1.0 to 2.1)	.47
b	84	130	55.5	55.0	0.6 (−1.0 to 2.1)	.48
PCS-YA						
a	90	129	55.1	54.3	0.9 (−0.9 to 2.7)	.34
b	88	129	55.3	54.6	0.7 (−1.1 to 2.5)	.46

TBPI-CAR indicates Total Behavior Problem Index by caretaker report; PCS-CAR, Physical Health Scale by caretaker report; PCS-YA, Physical Health Scale by youth report; a, adjusted for all the variables in Table 1 except gestational age; b, adjusted for variables in Table 1 except for gestational age and attrition.

tained effects at age 8 were seen among those attending >400 days of center-based education (out of a maximum of 500) during the intervention period.[36] In addition, we speculate that earlier entry into special education may also play a role and needs additional investigation.

Several limitations affect the interpretation of our study. The first is the lower response rate than previous rounds of assessment in the IHDP, particularly with loss among more disadvantaged children and those study participants with lower previous cognitive development. Our adjustment of the results for these factors suggests that the losses to the cohort do not significantly affect the results, but differences may have occurred in unmeasured factors.

Second, the age of assessment, 18 years, may not be optimal in identifying economically important outcomes, such as higher educational attainment, definitive occupational status, and eventual income. In the Perry Preschool Project, economically significant differences are seen at 27 and now 40 years of age.[6,8] Even at 21 years of age, the Abecedarian Project[4] did not detect significant differences in employment but, in results parallel to ours, did find differences in risk behaviors. They did, however, find an intervention advantage in the proportion enrolled in a 4-year college or university.

Third, because our study sample is not all disadvantaged, the power to detect significant differences in some areas is limited. As noted earlier, arrest rates in the Perry Preschool Project[8] are substantially higher than other longer-term follow-up studies of early education. With all rates of adverse outcomes, such as arrest or special education classification, differences favoring the INT group, especially the HLBW youth, are either similar or about half the effect size. Thus, we would have needed a substantially larger sample, in the order of the 1500 in the Chicago study,[10] to achieve statistical significance. This finding also suggests caution in extrapolating cost-effectiveness arguments of early educational intervention based on reductions in costs among populations with less risk of poor educational attainment and criminal behavior.

Despite these limitations, this phase of the IHDP provides important support for the efficacy of early educational interventions in the longer-term outcomes of children. It also presents significant challenges in identifying those who would most benefit and those who may need continuing support to achieve academically. However, the results provide support for extending such educational opportunities to a broader spectrum of children than included in previous studies.

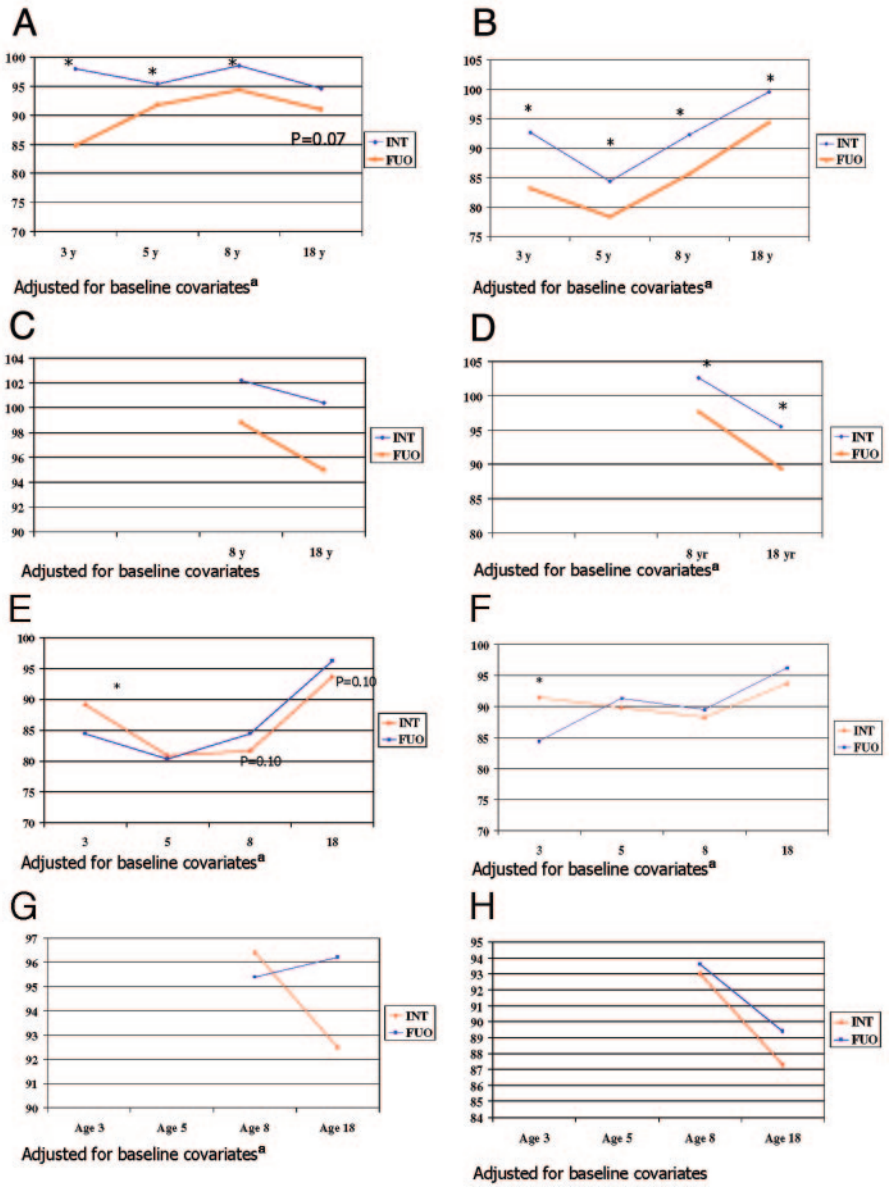

FIGURE 1

Cognitive and achievement scores at different assessment points in the IHDP for the heavier birth weight group. A, IHDP: IQ results over time—HLBW group (adjusted for baseline covariates; [a] $P < .05$). B, IHDP: PPVT results over time—LLBW group (adjusted for baseline covariates; [a] $P < .05$). C, IHDP: Woodcock-Johnson Achievement Test reading results over time—HLBW group (at 18, $P = .10$; adjusted for baseline covariates, $P = .05$). D, IHDP: Woodcock-Johnson Achievement Test math results over time—HLBW group (adjusted for baseline covariates; [a] $P < .05$). E, IDHP: PPVT results over time—LLBW group (adjusted for baseline covariates; [a] $P < .05$). F, IDHP: IQ results over time—LLBW group (adjusted for baseline covariates; [a] $P < .05$). G, IDHP: Woodcock-Johnson Achievement Test reading results over time—LLBW group (at 18, $P = .10$; adjusted for baseline covariates). H, IDHP: Woodcock-Johnson Achievement Test math results over time—LLBW group (adjusted for baseline covariates). Baseline covariates: birth weight, maternal age at youth's birth, race/ethnicity, child gender, maternal educational attainment at youth's birth, Neonatal Health Index,[26] and study site. Values for 3- and 5-year assessments are from ref 16; values for 8-year assessments are from ref 16.

APPENDIX

The YRBSS questionnaire has no established scoring algorithm. We identified questions in 6 risk areas (antisocial behavior, suicidal ideation/attempt, smoking, alcohol usage, and marijuana usage) considered most relevant to the study. Because there were different numbers of items in each of these categories, a simple sum across the questions would have weighted some areas more than others. Thus, to obtain an interval score for risk behaviors, we developed a 3-level score for each area, assigning the numeric values of 0, 1, and 2 to the low, medium, and high scores. Cut points were assigned so that ~15% of the group fell into the highest-risk group, the majority (>60%) in the low-risk group, and the remainder in the moderate-risk group. We summed these area-dependent scores to produce an overall score of risky behavior. This approach parallels that of Flisher et al,[37] in a report of behavior in a community sample of children and adolescents.[1] Of the 636 subjects, data for this analysis was missing on 5%. The resulting score had a mean of 2.4 in our study population, with an SD of 2.6 and a range of 0–11. Initial bivariate comparisons of INT and FUO groups were done, therefore, using nonparametric techniques. Adjusting for baseline covariates and attrition, however, relied on parametric statistics as described in "Methods."

ACKNOWLEDGMENTS

This study was supported through grant 039543 from the Robert Wood Johnson Foundation. The funder provided funding for the data collection and reviewed both the proposal and the final results.

For previous participants in the IHDP, see refs 14–16.

REFERENCES

1. Barnett WS. Long-term effects of early childhood programs on cognitive and school outcomes. *Future Child.* 1995;5:25–50
2. Yoshikawa H. Long-term effects of early childhood programs on social outcomes and delinquency. *Future Child.* 1995;5: 51–75
3. Campbell FA, Ramey CT. Cognitive and school outcomes for high-risk African American students at middle adolescence; positive effects of early intervention. *Am Educ Res J.* 1995;32: 743–772
4. Campbell FA, Ramey CT, Pungello E, et al. Early childhood education: Young adult outcomes from the Abecedarian Project. *App Dev Sci.* 2002;6:42–57
5. Gray SW, Ramsey B, Claus R. *From 3 to 20: The Early Training Project*. Baltimore, MD: University Part Press; 1982
6. Schweinhart LJ, Barnes HV, Weikart DP, et al. *Significant Benefits: The High/Scope Perry Preschool Study Through Age 27.* Monographs of the High/Scope Educational Research Foundation, No. 10. Ypsilanti, MI: High/Scope Educational Research Foundation; 1993
7. Lally JR, Mangione PL, Honig AS. The Syracuse University Family Development Research Project: long-range impact of an early intervention with low-income children and their families. In: Powell DR, ed. *Parent Education as Early Childhood Interventions: Emerging Directions in Theory, Research and Practice.* Norwood, NJ: Ablex; 1988
8. Schweinhart LJ, Montie J, Xiang Z, Barnett WS, Bellfield CR, Nores M. *Lifetime Effects: The High/Scope Perry Preschool Study Through Age 40.* Monographs of the High/Scope Educational Research Foundation No. 14. Ypsilanti, MI: High/Scope Press. 2006
9. Olds D, Henderson CR, Cole R, et al. Long-term effects of nurse home visitation on children's criminal and antisocial behavior: 15-year follow-up of a randomized controlled trial. *JAMA.* 1998;280:1238–1244
10. Reynolds AJ, Temple JA, Robertson DL, et al. Long-term effects of an early child childhood intervention on educational achievement and juvenile arrest. *JAMA.* 2001;285:2339–2346
11. Gross RT, Spiker D, Haynes C, eds. *Helping Low Birth Weight Premature Babies*. Palo Alto, CA: Stanford University Press; 1997
12. Ramey CT, Bryant DM, Wasik BH, et al. Infant Health and Development Program for low birth weight, premature infants: program elements, family participation, and child intelligence. *Pediatrics.* 1992;89:454–465
13. Liaw FR, Brooks-Gunn J. Patterns of low birth weight children's cognitive development and their determinants. *Dev Psychol.* 1993;29:1024–1035
14. The Infant Health and Development Program. Enhancing the outcomes of low-birth-weight, premature infants: a multisite, randomized trial. *JAMA.* 1990;263:3035–3042
15. Brooks-Gunn J, McCarton CM, Casey PH, et al. Early intervention in low-birth-weight premature infants: results through age 5 years from the Infant Health and Development Program. *JAMA.* 1994;272:1257–1262
16. McCarton CM, Brooks-Gunn J, Wallace IF, et al. Results at age 8 years of early intervention for low-birth-weight premature infants. The Infant Health and Development Program. *JAMA.* 1997;277:126–132
17. McCormick MC, McCarton C, Brooks-Gunn J, et al. The Infant Health and Development Program: interim summary. *J Dev Behav Pediatr.* 1998;19:359–370
18. Campbell FA, Pungello EP, Miller-Johnson S, et al. The development of cognitive and academic abilities: growth curves from an early childhood educational experiment. *Dev Psychol.* 2001;37:231–242
19. Constantine, WL, Haynes CW, Spiker D, et al. Recruitment and retention in a clinical trial for low birth weight, premature infants. *J Dev Behav Pediatr.* 1993;14:1–7
20. Woodcock RW, Johnson MB. *Manual for the Woodcock-Johnson Tests of Achievement-Revised*. Allen, TX: RCL Enterprises; 1990
21. Kolbe LJ, Kann L, Collins JL. Overview of the Youth Risk Behavior Surveillance System. *Public Health Rep.* 1993;108 (supp 1):2–10
22. Peterson JL, Zill N. Marital disruption, parent-child relationships and behavior problems in children. *Marriage Fam.* 1986; 48:295–307
23. Psychological Corporation. *Wechsler Abbreviated Scale of Intelligence*. San Antonio, TX: Psychological Corporation; 1999
24. Dunn LM, Dunn LM. *Peabody Picture Vocabulary Test*, 3rd ed. Circle Pines, MN: American Guidance Service, Inc; 1997
25. Ware JE, Kosinski M. *SF-36 Physical & Mental Health Summary Scales: A Manual for Users of Version 1*. 2nd ed. Lincoln, RI: QualityMetric Incorporated; 2001
26. Scott DT, Bauer CR, Kraemer HC, et al. The Neonatal Health Index. Chapter 25. In Gross RT, Spiker D, Haynes CW, eds. *Helping Low Birth Weight, Premature Babies. The Infant Health and Development Program*. Stanford, CA: Stanford University Press; 1977
27. Wechsler D. *Manual for the Wechsler Intelligence Scale for Children*. 3rd ed. San Antonio, TX: Psychological Corporation; 1991
28. Dunn LM, Dunn LM. *Manual for Peabody Picture Vocabulary Test-Revised*. Circle Pines, MI: American Guidance Service; 1981
29. Achenbach TM, Edelbrock C. *Manual for the Child Behavior Checklist and Revised Child Behavior Profile*. Burlington, VT: Queen City Printers; 1983
30. Venables WN, Ripley BD. *Modern Applied Statistics With S*. 4th ed. New York, NY: Springer; 2002
31. Little JA, Rubin D. *Statistical Analysis With Missing Data*. Baltimore, MD: Wiley; 2002
32. Hack M, Klein NK, Taylor HG. Long-term developmental outcomes of low birth weight infants. *Future Child.* 1995;5: 176–196
33. Brooks-Gunn J, Gross RT, Kraemer HC, et al. Enhancing the cognitive outcomes of low birth weight premature infants: For whom is the intervention most effective? *Pediatrics.* 1992;84: 1209–1215
34. Rock DA, Stenner AJ. Assessment issues in the testing of children at school entry. *Future Child.* 2005;15;13–32
35. McCormick MC, McCarton C, Tonascia J, et al. Early educational intervention for very low birth weight infants: Results from the Infant Health and Development Program. *J Pediatr.* 1993;123:527–533
36. Hill JL, Brooks-Gunn J, Waldfogel J. Sustained effects of high participation in an early intervention for low-birth-weight premature infants. *Dev Psychol.* 2003;39:730–744
37. Flisher AJ, Kramer RA, Hoven CW, et al. Risk behavior in a community sample of children and adolescents. *J Am Acad Child Adolesc Psychiatr.* 2000;39:881–887

Early Intervention in Low Birth Weight Premature Infants: Results at 18 Years of Age for the Infant Health and Development Program

Marie C. McCormick, Jeanne Brooks-Gunn, Stephen L. Buka, Julie Goldman, Jennifer Yu, Mikhail Salganik, David T. Scott, Forrest C. Bennett, Libby L. Kay, Judy C. Bernbaum, Charles R. Bauer, Camilia Martin, Elizabeth R. Woods, Anne Martin and Patrick H. Casey

Pediatrics 2006;117;771-780
DOI: 10.1542/peds.2005-1316

PEDIATRICS®
go
Advanced Search
OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

This item requires a subscription.
If you have an individual print subscription to *Pediatrics*, online access is included.

Full Text
Early Intervention in Low Birth Weight Premature Infants: Results at 18 Years of...
McCormick et al. *Pediatrics.*2006; 117: 771-780.

To view this item, select one of the options below:

Sign In: Please use your AAP ID Number or Temporary Activation Code for your User Name.

User Name
Password
Sign In
☐ Remember my user name & password.
Forgot your user name or password?

- Sign in without cookies.
- Can't get past this page?
- Help with Cookies.
- Need to Activate?

Purchase Short-Term Access

This Article
- Abstract FREE
- Full Text (PDF)
- P³Rs: Submit a response
- Alert me when this article is cited
- Alert me when P³Rs are posted
- Alert me if a correction is posted

Services
- E-mail this article to a friend
- Similar articles in this journal
- Similar articles in ISI Web of Science
- Similar articles in PubMed
- Alert me to new issues of the journal
- Add to My File Cabinet
- Download to citation manager
- Search for citing articles in: ISI Web of Science (4)

Google Scholar

- Pay per Article - You may access **this article** (from the computer you are currently using) for 2 days for US$12.00.
- Pay per Issue - You may access **this issue** (from the computer you are currently using) for 2 days for US$25.00.
- SitePass - You may access **all content in *Pediatrics* Online** (from the computer you are currently using) for 2 days for US$45.00.
- Regain Access - You can regain access to a recent Pay per Article**,** Pay per Issue **or** SitePass purchase if your access period has not yet expired.

Sign Up

- Become a Member - Join the American Academy of Pediatrics.
- Subscription Information
- Order Online - Purchase your individual subscription online. [regain access to your recent online order]

- **Articles by McCormick, M. C.**
- **Articles by Casey, P. H.**
- **Articles citing this Article**

PubMed

- **PubMed Citation**
- **Articles by McCormick, M. C.**
- **Articles by Casey, P. H.**

Related Collections

- **Premature & Newborn**

American Academy of Pediatrics
DEDICATED TO THE HEALTH OF ALL CHILDREN™