Enhancing the Outcomes of Low-Birth-Weight, Premature Infants

A Multisite, Randomized Trial

The Infant Health and Development Program

The Infant Health and Development Program is an eight-site clinical trial designed to evaluate the efficacy of a comprehensive early intervention in reducing the developmental and health problems of low-birth-weight (≤2500 g) premature (≤37 weeks) infants. Nine hundred eighty-five infants, stratified by site and weight (≤2000 g or 2001 to 2500 g), were randomly assigned to receive an educational curriculum focused on child development, as well as family support and pediatric follow-up, or only pediatric follow-up. At corrected age 36 months, the intervention group had significantly higher mean IQ scores than the follow-up group (mean difference in the heavier group was 13.2 and in the lighter group 6.6), significantly fewer maternally reported behavior problems, and a small, but statistically significant, increase in maternally reported minor illnesses for the lighter-weight group only, with no difference in serious health conditions.

For editorial comment see p 3069.

OVER THE last decade, the survival rate for low-birth-weight (LBW) infants has increased markedly, raising questions about their subsequent health and development. A number of studies have found that LBW infants are at increased risk for developmental delay and for a variety of medical complications in infancy compared with their normal-birth-weight counterparts. At later ages, LBW children tend to have lower scores on tests of cognitive functioning, and are more prone to difficulties in behavioral adjustment and are at risk for having learning problems and poor academic achievement, even when cognitive test scores are normal. The risk for cognitive deficits is present throughout the full spectrum of birth weights less than or equal to 2500 g, although the risk increases as birth weight decreases. The likelihood of adverse developmental and scholastic outcomes also is greater in the face of socioeconomic disadvantage—itself a risk factor for low birth weight and prematurity—and places many LBW premature infants at dual risk from both biologic and environmental factors.

A number of intervention studies have shown improved outcomes for LBW infants. However, most have been conducted at single sites with a small number of subjects and have assessed short-term benefits. The most persuasive evidence of the efficacy of early intervention comes from single site studies of normal-birth-weight infants and preschoolers from socially disadvantaged families. The applicability of interventions designed for normal-birth-weight children, however, has not been tested for LBW premature children, a population that may have biologic constraints that limit their responsiveness to such interventions. Furthermore, many such programs include a group care component that might result in increased exposure to acute infectious conditions, the effect of which is unknown on this vulnerable population. Current legislation, Public Law 99-457, focusing on provision of interventions for handicapped and at-risk children, highlights the relevance and the immediacy of this issue.
Table 1.—Targeted and Actual Enrollment for Primary Analysis Group

<table>
<thead>
<tr>
<th>Study and Weight Group</th>
<th>Targeted No. of Patients (Per Site)</th>
<th>Actual No. of Patients†</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Arkansas</td>
<td>Einstein</td>
</tr>
<tr>
<td>Follow-up</td>
<td>90</td>
<td>80</td>
</tr>
<tr>
<td>Heavier</td>
<td>60</td>
<td>50</td>
</tr>
<tr>
<td>Lighter</td>
<td>45</td>
<td>48</td>
</tr>
<tr>
<td>Intervention</td>
<td>45</td>
<td>48</td>
</tr>
<tr>
<td>Heavier</td>
<td>15</td>
<td>21</td>
</tr>
<tr>
<td>Lighter</td>
<td>30</td>
<td>27</td>
</tr>
<tr>
<td>Total</td>
<td>135</td>
<td>128</td>
</tr>
<tr>
<td>Heavier</td>
<td>45</td>
<td>51</td>
</tr>
<tr>
<td>Lighter</td>
<td>90</td>
<td>77</td>
</tr>
</tbody>
</table>

*Follow-up indicates the group receiving the pediatric follow-up services but not the intervention services; intervention, the group receiving the pediatric follow-up and the intervention services; heavier, infants weighing 2001 to 2500 g at birth; lighter, infants weighing less than or equal to 2000 g at birth.
†For full names of sites, see list of participating universities.

The Infant Health and Development Program is the first multisite, randomized clinical trial designed to evaluate the efficacy of combining early child development and family support services with pediatric follow-up in reducing developmental, behavioral, and other health problems among LBW premature infants (birth weight, ≤2500 g; gestational age, ≤37 weeks). The intervention protocol and the specific curricula used in the Infant Health and Development Program were adapted for LBW infants from two longitudinal studies of successful early intervention with socially disadvantaged normal-birthweight children.

PATIENTS AND METHODS

Eight medical institutions that serve diverse demographic populations in different geographical locations were selected through a national competitive review.

The Sample

Enrollment Criteria.—A total of 4551 inborn infants who would reach 40 weeks' postconceptual age between January 7, 1985, and October 9, 1985, and whose birth weights were less than or equal to 2500 g were screened at birth for eligibility. Of these, 3249 infants were excluded before randomization by protocol criteria related primarily to (1) residence; (2) gestational age greater than 37 weeks, as assessed by a modification of the Ballard Examination; or (3) hospital discharge before or after the designated recruitment period. (Details are in appendix 1, which is deposited in National Auxiliary Publication Service [NAPS] document 04773.) Unhealthy infants were included unless they had an illness or neurological deficit so severe as to preclude participation in the intervention program; only 61 such infants were excluded. Thus, the sample includes the majority of LBW premature infants who survived the neonatal hospitalization and lived in the catchment area, ie, within 45 minutes transportation time of each center. Of the 1302 who met the eligibility criteria, the parents of 274 (21.1%) refused consent to be randomized. Among the 1028 infants who had the consent of their parents and were randomized, 43 were withdrawn before participating in the study. The remaining 985 infants constituted the primary analysis group on which the findings of this report are based.

The Target Sample Size.—The research design included stratification by eight sites and two birth-weight groups (infants weighing 2001 to 2500 g, designated heavier, and those weighing ≤2000 g, designated lighter). One third of the sample came from the heavier and two thirds from the lighter group. To minimize the cost of the study, subjects within each weight group were allocated such that one third were randomized to the intervention group and two thirds to the follow-up group. The targeted overall sample size of 985 was based on an estimated effect size (ES = difference between treatment group means, expressed in SDs) of 0.5. For a single outcome, a power of 99% (P < .05, two tailed) was required in the total group and the lighter group alone.

Based on our research design, the targeted number of patients at each of the eight sites was 135. The targeted and actual numbers of enrolled patients are shown in Table 1. The differences between these numbers reflect the effects of randomization allocation and the shortfall of subjects at some sites. The actual enrollment remained adequate for a power of at least 99%.

Recruitment and Randomization

Immediately after hospital discharge, patients were randomized by the National Study Office using an adaptive randomization method that monitored for a 2:1 balance and for absence of bias between the two study groups in each site and birth-weight stratum. In the randomization, balance was monitored for birth weight, gender, maternal education (less than high school graduate; high school graduate; some college, or more), maternal race (black, Hispanic, and white other), primary language spoken in the home, and infant participation in another study. Further details of the randomization process will be presented in a separate article.

Program Description

The program was initiated on discharge from the neonatal nursery and continued until 36 months of age, corrected for prematurity. Infants in the intervention and the follow-up groups participated in the same pediatric follow-up, which comprised medical, developmental, and social assessments, with referral for pediatric care and other services as indicated. The services exclusively for the intervention group consisted of three components: home visits,* child attendance at a child development center, and parent group meetings. All services were provided free to the families.

Home Visits.—The protocol specified weekly home visits for the first year, and biweekly visits thereafter. The home visitor provided health and developmental information and family support and implemented two specific curricula. One curriculum emphasized cognitive, linguistic, and social development via a program of games and activities for the parent to use with the child. The second curriculum involved a systematic approach to help parents manage self-identified problems.

Child Development Centers.—Beginning at 12 months and continuing until corrected age 36 months, the intervention children attended the center 5 days a week. The teaching staff continued to implement the curriculum of
The mother's report of behavior problems was selected as the measure of behavioral competence at 36 months, using the Child Behavior Checklist for Ages 2 to 3 years. 

Employing the authors' "Total Problem Raw Score," higher scores indicate more reported behavior problems.

Health Status.—Health status was regarded as multidimensional. The dimensions evaluated in this study were (1) morbidity, defined as the presence or absence of health conditions; (2) functional status, defined by limitations in activities of daily living due to health problems and by alterations in physical growth; and (3) maternal perception of the child's health. As no single comprehensive measure existed, six different measures were selected to assess these three health areas.

Morbidity.—Two measures were developed by the Infant Health and Development Program to ascertain (1) the overall morbidity experienced by the child and (2) the seriousness of the reported health problems. At each assessment, mothers were asked to recall and report the incidence of each illness and condition that occurred since the last assessment. Their verbatim reports were assigned International Classifi-
cation of Diseases ICD—Ninth Revision codes at the National Study Office, and all reports corresponding to the same code (using the first three digits) were counted as one illness or condition per assessment period.

- Mother's Report: Morbidity Index. This index is a summation over the three years of the number of hospitalizations, outpatient surgeries, injuries not resulting in hospitalization or outpatient surgery, and different (code) illnesses and conditions.

- Mother's Report: Serious Morbidity Index. This index consists of the number of years (0 to 3) in which there was one or more of the following hospitalizations, outpatient surgeries, prolonged or recurrent illnesses totaling 30 days or longer in a given year, and injuries and briefer illnesses that were predefined as serious by a panel of pediatricians.

   **Functional Status.**—The Functional Status II(R) Scale, used at 36 months, gauged the mother's view of the limitations in the child's basic daily activities as a result of health problems. Higher scores indicate a more favorable functional status. Growth was assessed by length and body mass index at 36 months.

   **Maternal Perception or Rating of Child Health Status.**—The General Health Ratings Index from the Rand Corporation Health Insurance Study was used at 36 months. Higher scores indicate better perceived health.

All health data were collected by the clinic staff, except the 36-month growth measures.

**Analytic Strategies**

**Primary Analysis.**—To accommodate the possibility that the effect of the intervention would differ among sites, a procedure developed by Fleiss was adapted. In brief, the procedure first tests for heterogeneity of ESs across the 16 subgroups (eight sites and two birth-weight groups) and then estimates and tests either an averaged ES if there is significant heterogeneity or a pooled ES if not. (Details are in appendix 2, which is deposited in NAPS document 04773.)

To protect against false-positive results in testing efficacy with eight primary outcome measures, a Bonferroni correction was applied, and the .006 (.05/8) significance level was used for each outcome.

**Secondary Analysis.**—When a significant effect of the intervention was found in the primary analysis, a secondary analysis using multiple linear regression was employed to test whether certain initial status variables had an effect on outcome. Included among these variables were site, birth weight (as a continuous variable); gender; a Neonatal Health Index based on length of stay in the newborn nursery, adjusted for birth weight and standardized to a mean of 100, with higher scores indicating better health; and the following sociodemographic variables: maternal education level, maternal age, and maternal race. The main effect of each initial status variable (irrespective of study group) and of the intervention, was tested to ascertain whether overall each variable was independently predictive of the outcome. Further, the interaction of each initial status variable with the intervention was tested, again controlling for all other variables, to ascertain whether the effectiveness of the intervention was related to that initial status variable. Because the regression analysis was employed only when a significant outcome was found in the primary analysis, a .05 significance level was used for these tests.

**RESULTS**

**Sample Description and Retention**

Although the baseline characteristics of the study sample varied greatly among the eight sites (Table 2), overall the randomization procedure resulted in comparable intervention and follow-up groups at study entry. Retention was high, with 913 of the 987 children in the primary analysis group participating in the 36-month assessment for at least one of the primary outcome measures (98% of each study group). This small level of attrition did not affect the comparability of the two study groups at 36 months.

**Primary Outcomes**

The results for each of the eight primary outcome measures by study group are presented in Table 3. Where there are significant differences in ESs between the two weight groups, results are presented separately for the heavier and lighter groups.

**Cognitive Development.**—Primary Analysis. Overall, the mean IQ scores on the Stanford-Binet Intelligence Scale were significantly higher for the intervention children than for the follow-up children. Because the effect of the intervention varied significantly between birth-weight groups, separate ESs were estimated for each weight group. The effect in the heavier intervention vs the heavier follow-up groups was 13.2 IQ points (ES = .83, P < .001), and in the lighter intervention vs lighter follow-up groups, 8.6 IQ points (ES = .41, P < .001) (Table 3).

Controlling for site and initial status variables, the adjusted odds for having IQ scores less than 70, i.e., in the mental retardation range, were 2.7 times greater in the follow-up group (95% con-
Table 4.—Cumulative Stanford-Binet Scale IQ Scores, by Birth Weight and Study Group

<table>
<thead>
<tr>
<th>Cumulative Stanford-Binet Scale IQ Scores at 36 mo*</th>
<th>No. (%) of Patients in Birth-Weight Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>≤1500 g</td>
</tr>
<tr>
<td>&lt;70 intervention group</td>
<td>22 (26.8)</td>
</tr>
<tr>
<td>Follow-up group</td>
<td>43 (28.7)</td>
</tr>
<tr>
<td>&lt;85 intervention group</td>
<td>39 (47.6)</td>
</tr>
<tr>
<td>Follow-up group</td>
<td>95 (63.3)</td>
</tr>
<tr>
<td>&lt;90 intervention group</td>
<td>48 (58.5)</td>
</tr>
<tr>
<td>Follow-up group</td>
<td>104 (69.3)</td>
</tr>
<tr>
<td>&lt;100 intervention group</td>
<td>64 (78.0)</td>
</tr>
<tr>
<td>Follow-up group</td>
<td>123 (82.0)</td>
</tr>
<tr>
<td>Total sample intervention group</td>
<td>82 (100.0)</td>
</tr>
<tr>
<td>Follow-up group</td>
<td>150 (100.0)</td>
</tr>
</tbody>
</table>

The cutoff point of 70 is the IQ level below which the scores fall into the mental retardation range, according to the Stanford-Binet manual. The groups with scores less than 85 also include children with scores that are 1 SD below the mean, which we refer to as subaverage. The groups with scores less than 90 include 23.2% of the distribution, which approximates the lowest quartile. The groups with scores less than 100 include those below the mean.

Table 5.—Multiple Regression Analyses: Relationship of Initial Status Variables to Three Major Outcome Measures*

<table>
<thead>
<tr>
<th>Stanford-Binet Scale</th>
<th>Child Behavior Checklist for Ages 2-3 y</th>
<th>Mother’s Report: Morbidity Index</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Standardized Estimate</td>
<td>p</td>
</tr>
<tr>
<td>Initial status variable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site</td>
<td>&lt;.001</td>
<td>.028</td>
</tr>
<tr>
<td>Race</td>
<td>&lt;.001</td>
<td>.044</td>
</tr>
<tr>
<td>Birth weight</td>
<td>&lt;.001</td>
<td>-.025</td>
</tr>
<tr>
<td>Sex</td>
<td>NS</td>
<td>.097</td>
</tr>
<tr>
<td>Maternal age</td>
<td>-.009</td>
<td>-.096</td>
</tr>
<tr>
<td>Maternal education</td>
<td>-.024</td>
<td>-.165</td>
</tr>
<tr>
<td>Neonatal Health Index</td>
<td>-.090</td>
<td>-.007</td>
</tr>
<tr>
<td>Intervention effect</td>
<td>-.025</td>
<td>-.095</td>
</tr>
<tr>
<td>Site × intervention</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Race × intervention</td>
<td>NS</td>
<td>NS</td>
</tr>
<tr>
<td>Birth weight ×</td>
<td>.067</td>
<td>.014</td>
</tr>
<tr>
<td>Sex × intervention</td>
<td>.040</td>
<td>.045</td>
</tr>
<tr>
<td>Maternal age ×</td>
<td>.024</td>
<td>.066</td>
</tr>
<tr>
<td>Maternal education</td>
<td>.024</td>
<td>.102</td>
</tr>
<tr>
<td>Neonatal Health Index</td>
<td>-.005</td>
<td>-.002</td>
</tr>
</tbody>
</table>

*These are the three outcome measures that showed significant effects of the intervention.
†Standardized estimate is the regression coefficient multiplied by the SD of the outcome measure divided by the SD of the independent variables. They are presented only for the quantitative independent variables to indicate the direction and magnitude of their effects on the outcome measures.
‡The main effect of each initial status variable (irrespective of study group) and of the intervention was tested to ascertain whether overall each variable was independently predictive of the outcome. Further, the interaction of each initial status variable with the intervention was tested, again controlling for all other variables, to ascertain whether the effectiveness of the intervention was related to that initial status variable.
§NS indicates not significant.

Table 4 shows the proportion of children in three birth-weight groups with IQ scores below several cut points of clinical relevance. In each of the three birth-weight groups there was a larger proportion of follow-up group children than intervention group children with low IQ scores. In one subgroup only, the infants who weighed less than or equal to 1500 g and had IQ scores in the range less than 70, this difference was negligible.

Secondary Analysis.—As previous research has found, and as seen in Table 5, the multiple regression analysis indicated a significant main effect of several of the initial status variables on the Stanford-Binet Scale IQ scores. Higher scores were associated with some sites and also with being white, with higher birth weight, higher Neonatal Health Index, and higher maternal education. Controlling for these variables, there was a significant effect of the intervention (adjusted ES = .59, P < .001). Finally, there was an interaction between birth weight and the intervention such that the intervention was more effective for infants of higher birth weight (P = .014).

Behavioral Competence.—Primary Analysis.—The average score on the Child Behavior Checklist was significantly lower for the intervention group than the follow-up group, with higher scores indicating more reported behavior problems (ES = -.18, P = .006) (Table 3). Although the difference between study groups was small, the adjusted odds for having a score above 63, the cut point above which scores are correlated with clinically evident behavior problems, were 1.8 times greater in the follow-up group (95% confidence interval, 1.2 to 2.9). The actual percentages were 18.8% for the follow-up group and 13.9% for the intervention group.

Secondary Analysis.—The multiple regression analysis indicated significant main effects of several initial status variables. Higher scores (suggestive of more behavior problems) were associated with some sites, with being black or Hispanic, with being male, and with lower maternal age and education level (Table 5). Controlling for these variables, there was a significant effect of the intervention (adjusted ES = -.20, P = .003). The only variable that had a significant interaction with the intervention was maternal education (P = .009). With college-educated mothers there seemed to be little difference between the follow-up and intervention groups, whereas with mothers with less education, those in the intervention group reported fewer behavior problems.

Health Status.—Primary Analysis.—Among the six health status measures, the only measure with a significant treatment effect was the Mother’s Report: Morbidity Index. Higher morbidity scores were reported for the lighter-born children in the intervention group than for the lighter-born children in the follow-up group (ES = .29, P < .001); no significant difference was found in the heavier groups (Table 3).

Secondary Analysis.—The multiple regression analysis indicated significant
Table 6.—Site Variations in Three Primary Outcome Measures*

| Site and Study Group | Outcome Measure | Stanford-Binet Scale | Child Behavior Checklist for Ages 2-3 y | Mother’s Report: Morbidity Index |
|----------------------|-----------------|----------------------|----------------------------------------|
|                      | No. | Mean  | SD | No.  | Mean  | SD | No.  | Mean  | SD |
| Arkansas             |     |       |    |      |       |    |      |       |    |
| Follow-up            | 77  | 85.2  | 16.8 | 76   | 47.8  | 22.7 | 77   | 7.0   | 3.2 |
| Intervention         | 42  | 99.5  | 18.0 | 42   | 39.3  | 16.7 | 41   | 6.9   | 2.7 |
| Einstein             |     |       |    |      |       |    |      |       |    |
| Follow-up            | 78  | 74.2  | 15.7 | 77   | 47.1  | 20.5 | 77   | 6.2   | 2.6 |
| Intervention         | 43  | 84.7  | 18.4 | 43   | 49.0  | 22.5 | 43   | 6.5   | 3.3 |
| Harvard              |     |       |    |      |       |    |      |       |    |
| Follow-up            | 88  | 96.7  | 22.4 | 87   | 44.6  | 20.2 | 88   | 8.8   | 3.5 |
| Intervention         | 43  | 97.1  | 21.5 | 40   | 53.1  | 18.4 | 43   | 9.1   | 4.1 |
| Miami                |     |       |    |      |       |    |      |       |    |
| Follow-up            | 51  | 68.0  | 14.2 | 51   | 53.8  | 21.3 | 49   | 5.5   | 2.4 |
| Intervention         | 40  | 81.0  | 12.0 | 39   | 49.1  | 16.0 | 40   | 6.1   | 2.6 |
| Pennsylvania         |     |       |    |      |       |    |      |       |    |
| Follow-up            | 51  | 82.5  | 16.2 | 51   | 45.6  | 19.9 | 51   | 6.6   | 3.0 |
| Intervention         | 43  | 95.1  | 12.6 | 43   | 43.9  | 24.2 | 43   | 6.9   | 2.4 |
| Texas                |     |       |    |      |       |    |      |       |    |
| Follow-up            | 79  | 80.4  | 12.9 | 75   | 48.1  | 20.6 | 77   | 5.7   | 2.4 |
| Intervention         | 47  | 87.1  | 17.6 | 47   | 42.7  | 17.2 | 47   | 7.8   | 3.5 |
| Washington           |     |       |    |      |       |    |      |       |    |
| Follow-up            | 76  | 92.0  | 21.6 | 73   | 49.1  | 18.7 | 75   | 7.8   | 3.1 |
| Intervention         | 47  | 100.5 | 21.3 | 46   | 45.7  | 17.9 | 47   | 8.8   | 3.6 |
| Yale                 |     |       |    |      |       |    |      |       |    |
| Follow-up            | 61  | 91.1  | 20.0 | 59   | 42.6  | 18.9 | 57   | 7.1   | 3.2 |
| Intervention         | 42  | 102.5 | 17.3 | 38   | 38.7  | 16.6 | 41   | 8.7   | 3.0 |
| Total                |     |       |    |      |       |    |      |       |    |
| Follow-up and        | 561 | 84.5  | 19.9 | 547  | 47.2  | 20.5 | 551  | 6.9   | 3.1 |
| Intervention         | 347 | 93.5  | 19.1 | 338  | 43.7  | 19.1 | 348  | 7.6   | 3.3 |
| Follow-up and        | 908 | 885   |     |      |       |    |      |       |    |
| Intervention         | 896 | 885   |     |      |       |    |      |       |    |

*These are the three outcome measures that showed significant treatment effects.
†For full names of sites, see list of participating universities.

The main effects of several initial status variables. Higher values of the Morbidity Index (suggestive of more morbidity) were associated with some sites, with lower birth weight, with lower Neonatal Health Index, with being male, with being white, and with higher maternal education level. Controlling for these variables, there was a significant effect of the intervention (adjusted ES = .27, P < .001). The only variable that had a significant interaction with the intervention was maternal age, with younger mothers in the intervention group reporting higher scores than younger mothers in the follow-up group (Table 5).

Site Differences

There was a wide variation in the primary outcomes among the eight sites. This variation is demonstrated in the scores of the follow-up groups and in the differing magnitudes of the treatment effects (Table 6). With the IQ test, for example, the follow-up group scores ranged from a mean of 68.0 at the University of Miami (Fla) School of Medicine to a mean of 96.7 at Harvard Medical School, Boston, Mass. Similarly, the magnitude of the treatment effects varied from an IQ difference of 0.4 at Harvard to 14.3 at the University of Arkansas for Medical Sciences, Little Rock. The multiple regression analyses suggest, however, that the site differences are associated predominantly with variations in the initial status characteristics. Specifically, in Table 5 it can be seen that after controlling for the initial status variables, no significant site-by-intervention interactions are detected, and the effect of the intervention overall is statistically significant for the three measures.

**COMMENT**

This study is the largest reported multisite, randomized clinical trial of an intensive early childhood intervention. It is also the first home- and center-based educational intervention for LBW premature infants from birth to age 3 years. The results of this study indicate the effectiveness of a comprehensive intervention, even for biologically vulnerable infants. Our findings show that the children who received the intervention experienced: (1) significantly higher IQ scores; (2) significantly fewer maternally reported behavior problems; and (3) a small, but significant increase in maternally reported minor morbidity, with no evidence of an increase in reported serious health problems. The magnitude of these effects and also the levels of the three outcome measures in the follow-up groups (Table 6) were influenced by the variation in the initial status characteristics at the sites (Table 2).

**Cognitive Outcome**

The largest treatment effect was the significantly higher cognitive scores achieved by the intervention group compared with the follow-up group at corrected age 36 months. The overall results are consistent with the magnitude of cognitive gains reported previously in single-site intervention studies with normal-birth-weight disadvantaged children. Further, the low scores in the follow-up group parallel those reported in other samples of LBW socially disadvantaged children. Birth weight had a main effect on the level of the IQ scores as well as an influence on the efficacy of the intervention, with a greater effect on heavier-born children. Two important observations regarding birth weight warrant emphasis. First, as shown in Table 4, low and subaverage scores are not limited to the smallest infants. In the follow-up group, more than 18% of those weighing 1501 to 2000 g and 2001 to 2500 g have IQ scores below 70, and almost 50% are below 85. Second, Table 4 also suggests that there is one group in which the effectiveness of the intervention is questionable—those with birth weights less than or equal to 1500 g and with IQ scores less than 70. In this lowest-weight group, there is little difference in the proportion of intervention vs follow-up children who have IQ scores less than 70. In all other weight-by-age groups, the results favored the children in the intervention group.

The improvement in cognitive development in the intervention group was statistically significant at seven of the eight individual sites (P < .05). We speculate that the nonsignificant result at the one site, Harvard, may be related to the sociodemographic characteristics of the site, such as the large proportion of college-educated mothers, as well as the relative abundance of community resources compared with the other sites.

Other issues warrant consideration in the interpretation of the cognitive findings. The first is the potential effect of bilingualism. Bilingual children are reported to perform less well on tests administered in English; here the Stanford-Binet test was administered only in English. Although prerecruitment screening ensured that all mothers...
could communicate in English sufficiently to participate in the program, 25% of the children were regularly exposed to another language. We therefore repeated the primary analysis, excluding all children in bilingual environments. Among the remaining monolingual children, the overall advantage for the intervention group remained large and highly significant (heavier group: ES = .81, P < .001; lighter group: ES = .28, P = .006). Thus, the effect of the intervention is not solely a consequence of English-language facilitation in the subgroup of children in bilingual environments.

Another issue to consider is whether exposure to the test items or to similar materials during the study influenced the magnitude of IQ scores in the intervention group. The intervention curriculum and the cognitive instruments were selected independently, and as a further precaution, the protocol denied specific feedback to the education staff on any child's cognitive test performance.

Finally, although all assessments in this study were based on age corrected for prematurity, repetition of the primary analysis using uncorrected chronological age yielded similar significant results for cognitive outcome and the same mean differences between the two study groups.

**Behavioral Outcome**

Compared with the follow-up group, the intervention group at 36 months experienced a small, yet significant, advantage in behavioral competence, as indicated by lower behavior problem scores on the Child Behavior Checklist. In secondary analysis, the treatment group difference was seen largely in reports from the less-educated mothers. Thus, the intervention may have helped these mothers to become better informed about age-appropriate behaviors and consequently to report fewer behavior problems; it may have taught them more effective techniques for behavior management; or it may have altered the children's behavior. Although the data were collected through maternal report, other studies involving maternally reported behavior problems at this age indicate that such reports correspond to clinically detected problems and may be predictive of longer-term adverse outcomes.

**Health Status Outcomes**

Given the positive effects of this program in cognitive development and behavioral competence, it is important to ask whether group participation in the child development centers led to an increase in serious health problems within the intervention group. The only health measure with a significant intervention effect was Mother's Report: Morbidity Index, where the lighter intervention group had higher reported morbidity scores than the lighter follow-up group. Further analysis of the components of the Morbidity Index indicated that the difference was accounted for primarily by an increase in the number of reported brief illnesses and conditions. No significant differences were reported by the mothers in the two groups in the Serious Morbidity Index.

At face value these findings suggest that the intervention led to a slight, but significant, increase in morbidity. However, other factors may well have contributed to these findings. First is the issue of reporting bias. Other studies indicate that maternal report may underestimate the child's health conditions, particularly in disadvantaged groups.

A second issue is the more intense health surveillance and health education in the intervention group compared with the follow-up group.

It is unclear whether the higher Morbidity Index score in the intervention group reflects more complete and accurate reporting or a real increase in acute conditions (that may well occur with children in group care). Nevertheless, during a 2-year period at eight different rigorously controlled child development centers there was not a single serious infectious epidemic or a single major accident. It is also possible that the pediatric follow-up program extended to all the children, regardless of group assignment, improved the health of both study groups relative to LBW children in most communities.

**Retention and Compliance**

In this randomized clinical trial we largely avoided the concern of differential attrition in the two groups by achieving a 93% retention rate in the intervention and the follow-up groups, with good comparability in the initial characteristics of the subjects at entry and of those retained at 36 months. The level of participation in the intervention was generally high and similar across sites; further detailed analyses are forthcoming in a separate article.

The issue of "crossover" between the intervention and follow-up groups also merits a mention of concern. In the follow-up group, 30% had entered a community day-care center by 36 months corrected age; conversely, 14% of the intervention group never attended the child development center. Because subjects were retained and analyzed in the group to which they were randomly assigned, the effect of such crossovers would be to reduce the apparent magnitude of the effect of the intervention, i.e., the actual treatment effects might be somewhat larger than those we have documented.

**Conclusions**

We conclude that this comprehensive and intensive early intervention program shows substantive promise of decreasing the number of LBW premature infants at risk for later developmental disability. We also document its apparent safety in a biologically vulnerable population. Additional exploratory studies are under way to examine the variations in ESs across different subgroups of children.

The long-term significance of these findings is being addressed in the continued follow-up of the study cohort. From the low-SES studies of normal-birth-weight disadvantaged children in preschool programs, it seems that IQ effects attenuate when such children enter the usual inner-city schools but that subsequent advantages of the interventions may accrue in the primary grades. Such advantages include higher academic achievement, fewer placements in remedial classes, and lowered risk of grade retention and school dropout; later, there may be reduced juvenile delinquency and increased employment.

Improving the status of young children is the subject of policy debate currently and in the foreseeable future. The results of this study are especially timely and relevant to the concerns at the federal and state levels for providing appropriate interventions for children at risk for developmental delay.

The Infant Health and Development Program was funded by grants from the Robert Wood Johnson Foundation to the Department of Pediatrics, Stanford University, Stanford, Calif; the Frank Porter Graham Child Development Center, University of North Carolina at Chapel Hill; and the eight participating universities. Additional support for the National Study Office was provided to the Department of Pediatrics, Stanford University, from the Pew Charitable Trusts; the Bureau of Maternal and Child Health and Resources Development and the National Institute of Child Health and Human Development, Health Resources and Services Administration, Public Health Service, US Department of Health and Human Services (grant MCH 006515); and the Stanford Center for the Study of Families, Children, and Youth.

We express our deep appreciation to Ruby P. Hearn, PhD, vice president of the Robert Wood Johnson Foundation, for her guidance and support throughout the study.

See NAPS document #04773 for four pages of supplementary material from NAPS's Microfiche Publications, PO Box 3513, Grand Central Station, New York, NY 10016-3513. Remit with your order: not under separate cover, in US funds only $7.75 for photocopies or $4 for microfiche. Outside the United States or Canada, add postage of $4.50 for the first 20 pages and $1 for each 10.
1. references


