ABSTRACT. Objective. Studies of developmental outcomes in children with congenital heart disease (CHD) frequently use assessments conducted in infancy as primary endpoints. Whether test scores of CHD patients in infancy are predictive of status at school age has not been evaluated, however.

Methods. In the Boston Circulatory Arrest Study, 135 children with D-transposition of the great arteries repaired by arterial switch operation were administered the Bayley Scales of Infant Development and the Fagan Test of Infant Intelligence at 1 year of age and the Wechsler Intelligence Scale for Children, Third Edition and the Wechsler Individual Achievement Test at 8 years.

Results. Although most 1-year test scores were significantly associated with 8-year test scores, the amounts of shared variance were modest (<10%). All 1-year test scores had poor sensitivity (16%–32%) and poor positive predictive value (35%–42%) but good specificity (80%–93%) and negative predictive value (78%–79%). More than half of the children with low scores at 8 years (≤85) had had scores >84 at 1 year.

Conclusion. This pattern suggests that although test scores at 1 year are modestly associated with test scores at 8 years, many children who are at risk for poor late outcomes will not be identified on the basis of 1-year test scores. Long-term follow-up of children with CHD is necessary to draw inferences about the developmental sequelae of preoperative, intraoperative, and postoperative factors. Pediatrics 2004;114:e572-e576. URL: www.pediatrics.org/cgi/doi/10.1542/peds.2003-0983-L; infant heart surgery, prediction, development, congenital heart defects.

ABBREVIATIONS. BSID, Bayley Scales of Infant Development; DTGA, D-transposition of the great arteries; MDI, Mental Development Index; PDI, Psychomotor Development Index; CI, confidence interval; PPV, positive predictive value; NPV, negative predictive value; OR, odds ratio.

From the *Department of Cardiology, Children's Hospital Boston, Boston, Massachusetts; †Clinical Research Program, Children's Hospital Boston, Boston, Massachusetts; ‡Department of Pediatrics, Harvard Medical School, Boston, Massachusetts; §Department of Biostatistics, Harvard School of Public Health, Boston, Massachusetts; ¶Department of Medicine, Children's Hospital Boston, Boston, Massachusetts; ‡Department of Neurology, Children's Hospital Boston, Boston, Massachusetts; and **Department of Neurology, Harvard Medical School, Boston, Massachusetts.

Accepted for publication Jun 2, 2004. doi:10.1542/peds.2003-0983-L

Reprint requests to (D.C.B.) Department of Neurology, Farley Basement, Box 127, Children's Hospital Boston, 300 Longwood Ave, Boston, MA 02115. E-mail: david.bellinger@childrens.harvard.edu

PEDIATRICS (ISSN 0031 4005). Copyright © 2004 by the American Academy of Pediatrics.
or additional open-heart procedures. Informed consent was obtained from the parents of all infants enrolled in the trial.

**Data Collection**

**One-Year Assessment**

Two neurodevelopmental tests were administered by 1 of 2 examiners.  
1. **BSID.** This is a widely used test of infant development. It yields 2 scores, the Mental Development Index (MDI) and the Psychomotor Development Index (PDI), both of which have a normative population mean of 100 and an SD of 16. Thus, a score \( \leq 84 \) SD below the mean, is considered low. In the restandardization of the BSID, however, the means of contemporaries on this version of the test were found to be considerably higher: 112 for MDI and 110 for PDI. Thus, an MDI or PDI score of 84 on the 1969 version of the BSID is, in effect, ably higher: 112 for MDI and 110 for PDI.

2. **Fagan Test of Infant Intelligence.** This test assesses visual recognition memory using a 10-trial habituation format. In each trial, familiarization to a photograph of a face is followed by paired presentation of the face with either a photograph of a similar but distinct face or with the same face in a different orientation. By observing the infant’s corneal reflections of the 2 stimuli, a “percent Novelty Preference” score is obtained and summed across trials. A score \(<53\%\) identifies children who are at risk for later mental retardation.

**Eight-Year Assessment**

A single examiner (D.C.B.) conducted all 8-year evaluations and did not review a child’s 1-year test results before conducting an evaluation. The evaluation included the following tests:

1. **Wechsler Intelligence Scale for Children, Third Edition.**

   This test yields 3 IQ scores: Full Scale, Verbal, and Performance. Each score is normed to have a mean of 100 and an SD of 15. Thus, a score \( \leq 85 \) is 1 or more SDs below the population mean and therefore was considered to be low.

2. **Wechsler Individual Achievement Test.** This test of key domains of academic achievement yields standard scores (means of 100, SD of 15) for Composite Reading and Composite Mathematics. These 2 composite scores were averaged to derive an overall Achievement score.

**Statistical Analyses**

Pearson correlation coefficients and linear regression methods were used to estimate the associations among the 3 neurodevelopmental test scores obtained at 1 year (MDI, PDI, and percent Novelty Preference) and 4 developmental test scores obtained at 8 years (Full-Scale IQ, Verbal IQ, Performance IQ, and Achievement). For each linear regression analysis, a 95% confidence interval (CI) was calculated for the predicted score at 8 years of age given the test score at 1 year of age. The odds (and 95% CI) of having a low Full-Scale IQ score at 8 years of age (\( \leq 85 \)), a low Verbal IQ score at 1 year (\( \leq 84 \) for MDI and PDI, <53% for novelty preference) was computed. In addition, for each 8-year outcome, the sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated for low scores on each of the three 1-year tests.

**RESULTS**

Of the 171 patients enrolled in the trial, 168 were alive at 1 year of age, and 155 (92%) returned for evaluation. We excluded the data of 10 children who were originally enrolled as a pilot study of trial feasibility. A child with autism could not be tested, and another child was too distressed to complete the developmental assessment. One child refused to cooperate with the administration of the items of the test, precluding calculation of a score. Thus, MDI scores were available for 135 children, and PDI scores were available for 134 children. Novelty Preference scores on the Fagan Test were available for 100 children. Children were excluded from analyses of Fagan Test scores when the test was not completed as a result of the child’s lack of cooperation or because the child’s age was not 11 to 13 months (±1 month of the age for which this test was normed).

At age 8, 165 of the 171 enrolled children were alive. Of the 160 children living in the United States, 155 (97%) were evaluated, including all 135 children whose data were included in analyses of developmental data obtained at 1 year of age.

The mean scores (± SD) on the MDI, PDI, and Novelty Preference were 104.9 ± 14.5, 95.2 ± 15.6, and 58.5 ± 7.7, respectively. Twelve (9%) children had MDI scores \( \leq 84 \), 28 (21%) children had PDI scores \( \leq 84 \), and 23 (23%) children had Novelty Preference scores <53%.

**BSID Scores and 8-Year Outcomes**

Both MDI (Fig 1) and PDI (Fig 2) scores were significantly associated with IQ and achievement test scores at age 8, although the correlations were modest in magnitude, ranging from 0.33 (Full-Scale IQ and MDI) to 0.16 (Achievement and PDI). One-year MDI or PDI scores thus accounted for, at most, 10% of the variance in these 8-year outcomes. For all pairs of 1-year and 8-year scores, the 95% CI for the predicted 8-year scores was wide. In other words, among children who achieved a given score at 1 year, considerable variability would be expected in the IQ and achievement scores achieved at 8 years.

Among children with a low MDI score at age 1, the estimated odds of having a low Full-Scale IQ score (\( \leq 85 \)) at age 8 was 2.5 times the risk among children with MDI scores >85 (95% CI: 0.8–8.6; sensitivity: 16%; specificity: 93%; PPV: 42%; NPV: 78%). Among children with a low PDI score at age 1, the estimated odds of having a low Full-Scale IQ score \( \leq 85 \) at 8 years was 2.1 (95% CI: 0.9–5.3; sensitivity: 31%; specificity: 82%; PPV: 36%; NPV: 79%). Among children with low scores on both MDI and PDI at 1 year (\( n = 5 \)), the estimated odds of having a Full-Scale IQ score \( \leq 85 \) at 8 years old was 5.2 (95% CI: 0.8–32.4).

We also calculated the odds of having a low Full-Scale IQ at 8 years using an MDI score \( \leq 96 \) and a PDI score \( \leq 94 \) as cutoffs. As noted, because of the upward drift in infants’ scores on the 1969 version of the BSID, these scores are ~1 SD below contemporary means. For MDI, the estimated odds ratio (OR) was unchanged (2.5), although the 95% CI was narrower (1.1–5.8) as a result of the larger number of infants classified as having a low score. For PDI, the OR was only 1.1, approximately half the size of the OR obtained using 84 as the cutoff. As with MDI, the 95% CI was narrower (0.5–2.5).

**Fagan Test of Infant Intelligence Novelty Preference Score and 8-Year Outcomes**

Novelty Preference scores were significantly correlated with Full-Scale (\( P = .04 \)) and Performance IQ (\( P = .04 \)) scores at age 8 and marginally correlated with Verbal IQ (\( P = .055 \)) and Achievement (\( P = .06 \)) scores. The correlations were modest (0.19–0.21), indicating that the amounts of variance shared between Novelty Preference score and all 8-year out-
comes were <5%. Among the 23 children who achieved "at risk" scores at age 1 year, the OR for achieving a Full-Scale IQ score of 85 or lower was 1.9 (95% CI: 0.7–5.2; sensitivity: 32%; specificity: 80%; PPV: 35%; NPV: 78%). The odds of having a full-scale IQ ≤85 at 8 years given low scores on all 3 infant test scores (MDI, PDI, and Novelty Preference; n = 3) was 6.4, but the 95% CI was very wide (0.6–74.4).

DISCUSSION

These secondary analyses indicate that among children who had DTGA and underwent infant heart surgery, those with poor neurodevelopmental scores at age 1 year were at significantly increased risk for poor developmental scores at 8 years of age. For instance, children who scored 1 or more SDs below the population mean on the MDI of the 1969 version of the BSID at 1 year (ie, ≤84) were between 2 and 3 times as likely as children with scores of 85 or greater to achieve a Full-Scale IQ score at 8 years that was 1 or more SDs below the population mean. In general, a low MDI score was more predictive of developmental difficulties at 8 years of age than were either a low PDI score or a low Novelty Preference score on the Fagan Test. The greater prediction of later IQ afforded by the MDI score than the PDI score is not surprising, given that, at 1 year of age, the PDI score reflects largely a child’s gross motor function. The fact that the MDI score was more predictive than the Novelty Preference score is somewhat more surprising in light of the claim that scores in infancy on information processing tests are better predictors of later IQ than are scores on sensory-motor tests. Although these 3 scores varied somewhat in their characteristics as screening tests, all had poor sensitivity (16%–32%) and poor PPV (35%–42%). However, all had reasonably good specificity (80%–93%) and NPV (78%–79%). This pattern suggests that a relatively small subset of children who scored 1 or more SDs below the population norm at age 1 year continued to score poorly at 8 years of age. Furthermore, a large majority of children who scored in the normal range at 1 year also scored in the normal range at 8 years. Perhaps most important, however, the relatively low PPVs of 1-year test scores indicates that the majority of children with reduced performance at 8 years had test scores that were in the normal range at 1 year. Therefore, although children with low scores at 1 year should be placed under the closest developmental surveillance, children with higher scores should

Fig 1. Score on 8-year IQ and achievement as a function of 1-year PDI. Scatterplots, regression lines, and 95% CIs for an individual predicted score are presented for the effects of PDI on Full-Scale IQ (A), Verbal IQ (B), Performance IQ (C), and Achievement (D). Pearson correlation coefficients and P values are also reported.
also be monitored for indications of emerging problems.

The predictive utility of the BSID, a test of sensory-motor skills, did not differ appreciably from the predictive utility of the Fagan Test, an information-processing test. The OR associated with a low MDI score at 1 year was ~30% higher than the OR associated with a low Fagan Test score (2.5 vs 1.9). Although a low score on the Fagan test had twice the sensitivity of a low MDI score, it had somewhat lower specificity and PPV. The screening test statistics associated with a low PDI score were very similar to those associated with a low score on the Fagan Test. It seemed that accuracy of prediction was greater when classification was based on MDI and PDI scores together or on MDI, PDI, and Fagan Test scores together. For both classification strategies, the ORs exceeded those obtained when classification was based on a single test score. Because of the relatively small number of children who achieved low scores on >1 test, however, the upper bounds of the CIs for these ORs were very wide (up to 74). In a larger sample than ours, joint classifications might have proved to yield significantly more accurate predictions of 8-year scores than classifications based on single test scores.

We used the expected mean and SD published in the manual for the 1969 version of the BSID to identify children with low scores at 1 year. As noted earlier, the upward drift over time in children's scores on this test has resulted in a current expected mean of ~112 for the MDI and 110 for the PDI. Although the secular trend in test scores does not affect the internal validity of our findings, it does mean that our definition of a “low” score at age 1 year (<84) corresponds to >1.5 SDs below the expected score rather than 1 SD. Defining a low MDI score as one that was ≤96 (16 points lower than 112) afforded prediction that was the same as that obtained using ≤84 as the cutoff (ie, ORs of 2.5). This was not the case when a low PDI score was defined as one that was ≤94 (16 points lower than 110), as the OR was only 1.1 (vs 2.1 using the lower cutoff).

Our findings are similar to those of other studies of the predictive utility of scores achieved on neurodevelopmental tests by children at increased medical risk. In a study of 196 low birth weight, premature infants, low (<84) scores on the MDI of the BSID at
4 months of age had the following screening characteristics for predicting cognitive delay between 3 and 8 years of age: sensitivity, 39%; specificity, 92%; PPV, 52%; and NPV, 87%. In a study of 196 infants for whom Novelty Preference scores on the Fagan Test at 7 to 10 months of age and IQ scores at 5 years of age were available, the correlations between test scores were 0.20 and 0.16 at the 2 clinics that contributed study subjects. The sensitivity of novelty preference scores for identifying children with IQ scores below the 10th percentile was 0 at 1 clinic and 22% at the other.

In our study cohort, fewer than half of the children who had low test scores at 1 year had low test scores at 8 years (true positives), and nearly all of those who had IQ scores within normal limits at 8 years had scores within the normal range at 1 year (true negatives). More important, however, more than half of the children who had low IQ scores at 8 years did not have low scores at 1 year (false negatives). Targeting children with low scores at 1 year for intervention services will capture a substantial subgroup of the children at increased risk of cognitive problems at school age. At the same time, however, it is clear that conclusions drawn about the long-term developmental outcomes of children with congenital heart disease based only on evaluations conducted at 1 year would, for many children, be incorrect. Long-term follow-up remains necessary to discern the contributions of preoperative, intraoperative, and postoperative factors to children’s developmental risk.

ACKNOWLEDGMENT

This study was supported by National Institutes of Health grants HL 41786, RR 02172, and P30-HD18655.

REFERENCES

3. McCall RB, Carriger MS. A meta-analysis of infant habituation and recognition memory performance as predictors of later IQ. Child Dev. 1993;64:57–79
Prediction of IQ and Achievement at Age 8 Years From Neurodevelopmental Status at Age 1 Year in Children With D-Transposition of the Great Arteries

Ellen McGrath, David Wypij, Leonard A. Rappaport, Jane W. Newburger and David C. Bellinger

*Pediatrics* 2004;114;e572; originally published online October 18, 2004; DOI: 10.1542/peds.2003-0983-L
Prediction of IQ and Achievement at Age 8 Years From Neurodevelopmental Status at Age 1 Year in Children With D-Transposition of the Great Arteries
Ellen McGrath, David Wypij, Leonard A. Rappaport, Jane W. Newburger and David C. Bellinger

Pediatrics 2004;114:e572; originally published online October 18, 2004;
DOI: 10.1542/peds.2003-0983-L

The online version of this article, along with updated information and services, is located on the World Wide Web at:
/content/114/5/e572.full.html