

Use of the Bayley Infant Neurodevelopmental Screener with an Environmental Risk Group

Christine Reiner Hess,¹ PhD, Mia A. Papas,² MS, and Maureen M. Black,² PhD

¹Department of Psychology, University of Maryland, Baltimore County, and ²Department of Pediatrics, University of Maryland School of Medicine

Objective To determine predictive validity of the Bayley Infant Neurodevelopmental Screener (BINS) during the first 2 years of life with a group of children at risk for developmental delay due to environmental risk factors. **Method** The setting consisted of home visits to participants. The BINS was administered to 106 children, ages 6 and 13 months, of low-income, African American, adolescent mothers. Three risk groups were identified: low, moderate, and high. The Bayley Scales of Infant Development, second edition (BSID-II), were administered at 24 months and served as the criterion standard. A cut score of 85 (1.00 SD below mean) represented a clinically meaningful indicator of delayed development on the mental and psychomotor developmental indices, as well as a composite of these indices. Two other cut scores on the BSID-II were also included for comparison: 90 (0.75 SD below mean) and 77 (1.50 SD below mean). **Results** Using BSID-II scores at 24 months as the criterion measure, 6- and 13-month BINS scores yielded low sensitivity values but high specificity values, regardless of how BINS risk groups were defined and which cut points on the BSID-II were used. Positive predictive value was higher when the cut score was set below 90 than when it was set below 85. **Conclusions** Low predictive validity of the BINS with an environmental risk group highlights the difficulties inherent in developmental screening among infants who have environmental, but not biological, risk factors. Because infants at environmental risk tend to experience developmental declines after infancy, it may be beneficial for primary care providers to use psychosocial screening tools to identify which children need closer monitoring and referral to enrichment programs to prevent developmental declines during toddlerhood.

Key words environmental risk; predictive validity; Bayley Infant Neurodevelopmental Screener; child development; African American; low income; infant of adolescent parent.

Recently, to ensure receipt of early intervention services, there has been increased emphasis on the identification of infants and toddlers (0–3 years) with developmental delays and disabilities (Individuals with Disabilities Education Act [IDEA] Amendments of 1997). To have the best chance of preventing developmental delays, intervention should begin during the first 3 years of life when children are experiencing rapid growth and development (Committee on Children with Disabilities, 2002). Children with biological risk factors (e.g.,

prematurity, low birth weight) and environmental risk factors (e.g., poverty, substance abuse, adolescent parent) have a higher likelihood of developmental delays than children without these risk factors (Hooper, Burchinal, Roberts, Zeisel, & Neebe, 1998; Sameroff, Seifer, Baldwin, & Baldwin, 1993; Schendel et al., 1997; Wood, Marlow, Costeloe, Gibson, & Wilkinson, 2000).

The federal Early Periodic Screening, Diagnosis, and Treatment (EPSDT) program recommends routine developmental screening during well-child visits following

All correspondence should be sent to Maureen Black, Department of Pediatrics, University of Maryland School of Medicine, 655 W. Lombard Street, Suite 311, Baltimore, Maryland 21201. E-mail: mblack@umaryland.edu.

Journal of Pediatric Psychology 29(5) pp. 321–330, 2004

DOI: 10.1093/jpepsy/jsh036

Journal of Pediatric Psychology vol. 29 no. 5 © Society of Pediatric Psychology 2004; all rights reserved.

Table I. Sensitivity, Specificity, and Positive Predictive Value of a Hypothetical Screening Test

Results of Screening Test	True Characteristics in the Population		
	Delayed	Not Delayed	Total
Delayed	8	10	18
Not delayed	2	80	82
Total	10	90	100

Sensitivity = $8/10 = 80\%$.

Specificity = $80/90 = 89\%$.

Positive predictive value = $8/18 = 44\%$.

Prevalence of developmental delay = 10% .

a schedule determined by each state (Rosenbach & Gavin, 1998). In Maryland, for example, EPSDT guidelines specify that children receive routine developmental screening at every well-child visit from 6 months to 4 years of age (Maryland Medicaid Program, 2003). The EPSDT program is based on a model of preventive care for early detection of illnesses and developmental problems so that services can be put in place to treat existing problems and prevent further delays.

The most widely used screening tool for young children is the Denver Developmental Screening Test II (Frankenburg, Dodds, Archer, Shapiro, & Bresnick, 1992; Frankenburg et al., 1996; Frankenburg, 2002). Although the Denver II has been helpful in identifying children at risk for developmental delays, it may over-identify children as delayed when they are actually typically developing, causing a high overreferral rate (Glascoe et al., 1992). There is a need to examine alternative screening tools that may have better predictive utility and are brief and cost-effective for routine screening.

The Bayley Infant Neurodevelopmental Screener (BINS) (Alyward, 1995) has been developed for children ages 3–24 months and assesses basic neurological functions/intactness, receptive functions, expressive functions, and cognitive processes. Cut scores are used to classify children as high, moderate, or low risk for developmental delays. It is administered by a trained professional in about 10 minutes, making it cost-effective for routine screening. The BINS is based on the Bayley Scales of Infant Development, second edition (BSID-II) (Bayley, 1993), a widely used assessment tool for identifying developmental delays in children ages 1–42 months. The BSID-II requires a highly trained examiner and takes 30–60 minutes, depending on the age of the child, making it an expensive and lengthy exam. Studies have examined the use of the BINS with children who are at high risk for developmental delays due to biological

risk factors such as prematurity and low birth weight (Aylward, Verhulst, & Bell, 1996; Aylward & Verhulst, 2000; Leonard, Piecuch, & Cooper, 2001; Macias et al., 1998), but not with children who have environmental risk factors.

Validity of the BINS as a screening tool has been evaluated by examining sensitivity, specificity, and positive predictive value (PPV). Sensitivity is defined as the proportion of children who are delayed and are correctly identified as delayed by the screening test, the “true positive” group (Aylward, 1994). Specificity is defined as the proportion of children who are not delayed and are correctly identified as not delayed by the screening test, the “true negative” group. PPV is defined as the proportion of children with a positive screening test result who are truly developmentally delayed.

We have used a brief example to illustrate how sensitivity, specificity, and PPV can be applied to developmental delays. If we have a population of 100 children, of whom 10 have a developmental delay and the other 90 do not, we may use a screening test to identify the children who have a developmental delay. Assume that the screening test has identified 18 children as delayed: 8 who are truly delayed and 10 who are not (see Table I). The sensitivity of the test—the proportion of children who are delayed and have been correctly identified by the screening test as delayed—is 80%. The specificity of the test—the proportion of children who are not delayed and who have been correctly identified by the screening test as not delayed—is 89% (Table I). Thus, the sensitivity of 80% means that 20% of the children who are truly delayed have been misclassified. The specificity of 89% means that the test has misclassified 11% of the children who were not delayed as being developmentally delayed.

The PPV is the probability that a child is truly developmentally delayed when results of the screening test are positive. Unlike the sensitivity and specificity of a screening test, which can be thought of as characteristics of the test itself, the PPV is dependent upon the prevalence of the disorder in the population tested (Gordis, 2000). In the example above, the prevalence of developmental delay in the population was 10%. The PPV is computed as the number of children correctly classified as delayed divided by the total number of children classified as delayed ($8/18 = 44\%$). If the prevalence of developmental delay were increased to 20%, the sensitivity and specificity would remain the same, but the PPV would increase to 64%. When the prevalence of developmental delay within the study

population is thought to be low, the PPV will necessarily be low. Therefore, the sensitivity of the test may be a better measure to evaluate a screening tool for developmental delay than its PPV.

When used with low birth weight children, the BINS showed moderate stability in classification of children over the first 2 years of life, although sensitivity, specificity, and PPVs varied according to the cutoffs used to define risk and delay (Aylward et al., 1996; Aylward & Verhulst, 2000; Leonard et al., 2001; Macias et al., 1998). Across studies of the BINS with low birth weight children, there was no clear consensus about which BINS risk grouping yielded the best predictive validity.

This investigation was undertaken to examine the predictive validity of the BINS among a group of socially disadvantaged infants, specifically African American infants from low-income, urban environments whose primary caregivers were adolescent mothers. Aylward and Verhulst (2000) have suggested that the BINS can be used reliably with children from diverse populations, including those from low socioeconomic backgrounds. This is the first study to evaluate how effective the BINS is with an environmental risk group of children who are not at biological risk. We examined the relationship between BINS performance measured at 6 and 13 months and performance on the BSID-II at 24 months. Three cutoff scores were used on the BSID-II criterion. The primary BSID-II criterion was set at one standard deviation below the mean, that is, a score of less than 85. This is considered to be clinically meaningful because it identifies children who are mildly delayed in developmental skills. Use of this cutoff score allows practitioners to identify which children require closer monitoring to help prevent further declines. Two other cutoff scores were used to determine whether greater sensitivity, specificity, and PPV would be achieved by raising or lowering the cutoff score on the BSID-II criterion.

Method

Participants

Participants included children of adolescent mothers who were enrolled in a longitudinal randomized controlled trial of home intervention designed to promote parenting and adolescent development among low-income families (Black, Siegel, Abel, & Bentley, 2001). Because national policies require that eligibility for public services be restricted to adolescent mothers who are in the guardianship of an adult (U.S. House of

Representatives, 1996), many adolescent mothers live with their mother. We limited our sample to adolescent mothers who were living with their mother (grandmother of the baby).

Eligibility for mothers included age less than 18 years at delivery, first-time delivery, African American, low income (defined as eligible for the Special Supplemental Nutrition Program for Women, Infants, and Children, or WIC, with family income under 185% of poverty level), and no chronic illnesses that would interfere with parenting or adolescent development. Eligibility for infants included full term (≥ 37 weeks), birth weight above 2500 grams, and no congenital problems or chronic illnesses. None of the infants experienced complications following delivery that required neonatal intensive care services.

Over 83% of the eligible mothers agreed to participate and 181 completed the baseline evaluation. There were no differences in maternal age or education between those who completed the baseline evaluation and those who did not. Follow-up evaluations were conducted when infants were 6, 13, and 24 months. Complete data were available from 148 (82%), 127 (70%), and 146 (81%) families at the 6-, 13-, and 24-month evaluations, respectively. One hundred six children (59%) completed both a 6-month and a 24-month visit, and 101 children (56%) completed both a 13-month and 24-month visit. Two children were untestable at the 24-month evaluation due to behavioral difficulties. The remaining families were not compliant with the follow-up visits. There were no differences in maternal age, maternal education, infant birth weight, infant gender, or intervention status between families included in the evaluation of the study and families who were not.

Measures

The BINS (Aylward, 1995) consists of 11–13 items, depending on the child's age. Each item is scored as optimal or nonoptimal, and the optimal responses are totaled to yield a summary score. The summary score reflects the child's level of risk for developmental delays or neurological impairments and is classified as one of three risk groups: low, moderate, or high. Cut scores used to determine risk classification vary according to the child's age. The internal reliability of the BINS has been reported to be high, ranging from .73 to .85 for children ages 3–24 months (Aylward, 1995).

The BSID-II is a standardized assessment tool that measures mental and motor development of children

from 1 to 42 months of age (Bayley, 1993). It is often considered to be the gold standard for identification of developmental delays for children from 0 to 3 years. Items on the mental developmental index (MDI) and the psychomotor developmental index (PDI) are administered individually to the child. The child's response is recorded and scored as *credit*, *no credit*, *refuse*, *report* (mother reports that the child can complete the item), or *omit*. The child's score is determined by the number of items for which credit was received. Raw scores are converted to index scores with a mean of 100 and a standard deviation of 15, based on age-specific norms.

Procedures

Infants and mothers were recruited from three hospitals in Baltimore, Maryland. The study was approved by the institutional review boards at all three hospitals. Mothers were approached shortly after delivery and given a brochure explaining the study. Those who expressed interest in enrolling in the study signed consent forms and were scheduled to receive a baseline home evaluation within 3 weeks. Additional evaluations were conducted in the families' homes when infants were 6, 13, and 24 months. Mothers were paid \$25 for each of the follow-up evaluations. When infants were 6 and 13 months old, the BINS (Aylward, 1995) was administered by research assistants who had been trained by a psychologist in administration and scoring. The research assistants were unaware of intervention-group assignment. When the children were 24 months old, the BSID-II (Bayley, 1993) was administered by a graduate student in psychology who had been trained in administration and scoring by a psychologist and was unaware of group assignment.

Data Analysis

The BINS scores at 6 and 13 months were classified using three different risk cutoffs as established in the scoring manual: low, moderate, and high. At 24 months, MDI and PDI scores from the BSID-II were dichotomized to indicate developmental delay versus nondelay using three criteria. The primary binary variable was created using a cutoff of 1.00 *SD* below the mean (scores ≥ 85 vs. < 85). Children scoring below 85 were classified as mildly delayed in mental skills per the BSID-II scoring manual; therefore, this represents a clinically meaningful criterion cutoff. In addition, two other criterion cutoffs were evaluated: 0.75 *SD* below the mean (scores ≥ 90 vs. < 90) and 1.50 *SD* below the mean (scores ≥ 77 vs. < 77). To determine whether the BINS more accurately

predicted the individual MDI and PDI or a combination of the two, a composite score was computed by averaging the MDI and PDI scores from the BSID-II at 24 months. The same three criteria were used to dichotomize the composite score.

The associations between the infants' BINS risk status and the MDI, PDI, and composite scores were calculated using the Spearman rho, a measure of correlation for nonparametric variables. Sensitivity, specificity, and PPVs were calculated separately for children receiving the intervention versus those who were not.

The receiver operator characteristic (ROC) curve was used to examine the predictive power of the BINS for identifying a child with developmental delay (Harrell, 2001). This procedure plots the true positives (sensitivity) against the false positives ($1.0 - \text{specificity}$). The area under the ROC curve approaches 1.0 when there is a high probability of true positives and a low probability of false positives. In contrast, if individuals are classified by flipping a coin, then the probability of true and false positives would be 50%, and the area under the curve would approach 0.5. Areas under the ROC curve were computed using a cross-validation technique that provides unbiased estimates of prediction error (Harrell, 2001). Since none of the results differed by intervention status, the data were combined for all analyses.

Results

When children were 6 months of age, the mean age of the mothers was 16.8 years ($SD = 1.00$) and they had completed 10.8 years of school ($SD = 1.43$). Mean ages for children at the 6-, 13-, and 24-month evaluations were 6.6 months ($SD = 1.10$), 16.8 months ($SD = 4.80$), and 25.2 months ($SD = 2.80$), respectively.

On the BINS at 6 months, 74%, 21%, and 6% of children were defined as low, moderate, and high risk, respectively. The distribution was similar for children at 13 months, with 79%, 20%, and 1% classified as low, moderate, and high risk, respectively. When we examined the consistency of the BINS from 6 to 13 months, we found that 82% of the children who were classified as low risk on the BINS at 6 months were also classified as low risk at 13 months, and 33% of children classified as moderate risk were also classified as moderate risk at 13 months (see Table II for transitional probabilities from 6 to 13 months for BINS classifications). None of the children classified on the BINS as high risk at 6 months was also classified as high risk at 13 months.

Table II. Transitional Probabilities from 6 to 13 Months, no. (%), *N* = 84

	BINS Risk-Group Category at 13 Months		
	Low (<i>n</i> = 65)	Moderate (<i>n</i> = 18)	High (<i>n</i> = 1)
BINS Risk-Group Category at 6 Months			
Low (<i>n</i> = 65)	53 (82)	11 (17)	1 (2)
Moderate (<i>n</i> = 15)	10 (67)	5 (33)	0 (0)
High (<i>n</i> = 4)	2 (50)	2 (50)	0 (0)

BINS = Bayley Infant Neurodevelopmental Screener.

At 24 months, average BSID-II scores were 85.3 (*SD* = 7.61) for the MDI, 94.6 (*SD* = 6.94) for the PDI, and 90.1 (*SD* = 5.77) for the composite. Forty-three percent of children scored at least 1.0 *SD* below the mean on the MDI, 8% scored at least 1.0 *SD* below the mean on the PDI, and 14% scored at least 1.0 *SD* below the mean on the composite score (see Table III). Few children scored less than 1.5 *SD* below the mean (cutoff of < 77) on the MDI, PDI, or composite at 24 months. Therefore, this criterion cutoff was eliminated from analyses.

Sensitivity of the Bayley Infant Neurodevelopmental Screener

Sensitivity of the BINS was low at both 6 and 13 months when the BSID-II criterion cutoff was set below 85 (see Table IV). At both 6 and 13 months, higher sensitivity was achieved when the low-risk BINS category was compared with the moderate- and high-risk BINS categories to predict 24-month MDI, PDI, and composite criterion measures than when the low- and moderate-risk BINS categories were compared with the high-risk BINS category. When a BSID-II cutoff of 85 was used at 6 months, sensitivity was highest when predicting the MDI–PDI composite compared with the individual MDI or PDI criterion measures (see Table IV, second column). In this case, 29% of children who had a developmental delay on the MDI–PDI composite at 24 months were identified by the 6-month BINS. Using the same cutoff at 13 months, sensitivity was highest when predicting the MDI compared with the PDI or composite (see Table IV, fourth column). That is, 27% of children who had a developmental delay on the MDI at 24 months were identified by the BINS at 13 months. The pattern of results was similar when the BSID-II criterion cut point was set below 90, with one exception. The highest sensitivity at 6 months occurred when the PDI was the criterion (see Table IV, second column). That is, 40% of children who were identified as mildly de-

Table III. Descriptive Statistics for 24-Month BSID-II Scores, *N* = 144

BSID-II	Mean (<i>SD</i>)	Cut Score (%)		
		< 90	< 85	< 77
MDI	85.3 (7.61)	70	45	17
PDI	94.6 (6.94)	24	9	0
Composite	90.1 (5.77)	37	12	0

BSID-II = Bayley Scales of Infant Development, second edition; MDI = motor developmental index; PDI = psychomotor developmental index.

velopmentally delayed on the PDI at 24 months were identified on the 6-month BINS.

Specificity of the Bayley Infant Neurodevelopmental Screener

The specificity of the screening test was very high, reaching 100% in some cases (see Table V). High specificity levels were achieved when the low- and moderate-risk BINS categories were compared with the high-risk BINS category. This pattern was consistent for the MDI, PDI, and composite measures regardless of whether the BSID-II cutoff was set below 85 or below 90. Specificity was only marginally higher for the below-90 cutoff versus the below-85 cutoff. The pattern of results was similar when using the 6-month BINS to predict 24-month BSID-II measures and when using the 13-month BINS to predict 24-month BSID-II measures. In this case, the BINS screening test at both 6 and 13 months did a good job of classifying children as not delayed who did not experience a developmental delay on the BSID-II at 24 months.

Table IV. Sensitivity Using Different Categories of Scores on 6- and 13-Month BINS and 24-month BSID-II

	BINS Score at 6 Months ^a		BINS Score at 13 Months ^b	
	Low and Moderate vs. High (%)	Low vs. Moderate and High (%)	Low and Moderate vs. High (%)	Low vs. Moderate and High (%)
BSID-II at 24 Months				
90 Cutoff (0.75 <i>SD</i>)				
MDI	4	23	2	24
PDI	4	40	0	13
Composite	2	25	0	22
85 Cutoff (1.00 <i>SD</i>)				
MDI	2	21	0	27
PDI	0	22	0	14
Composite	6	29	0	13

BINS = Bayley Infant Neurodevelopmental Screener; BSID-II = Bayley Scales of Infant Development, second edition; MDI = motor developmental index; PDI = psychomotor developmental index.

^a 106 children completed both the 6- and 24-month evaluations.

^b 101 children completed both the 13- and 24-month evaluations.

Table V. Specificity Using Different Categories of Scores on 6- and 13-Month BINS and 24-month BSID-II

	BINS Score at 6 Months ^a		BINS Score at 13 Months ^b	
	Low and Moderate vs. High (%)	Low vs. Moderate and High (%)	Low and Moderate vs. High (%)	Low vs. Moderate and High (%)
BSID-II at 24 months				
90 Cutoff (0.75 SD)				
MDI	91	66	100	86
PDI	94	78	99	77
Composite	91	72	98	81
85 Cutoff (1.00 SD)				
MDI	91	69	98	83
PDI	94	73	99	78
Composite	94	74	99	78

BINS = Bayley Infant Neurodevelopmental Screener; BSID-II = Bayley Scales of Infant Development, second edition; MDI = motor developmental index; PDI = psychomotor developmental index.

^a 106 children completed both the 6- and 24-month evaluations.

^b 101 children completed both the 13- and 24-month evaluations.

Positive Predictive Value of the Bayley Infant Neurodevelopmental Screener

Table VI shows the PPV of the BINS at both 6 and 13 months. Consistent with the findings for sensitivity, PPV was higher when using low versus moderate and high BINS grouping than when using the low and moderate versus the high BINS grouping. When using the below-85 BSID-II cutoff, the PPV ranged from 0 to 52%, with the highest PPV achieved when using the 13-month BINS to predict 24-month MDI. When using the below-90 BSID-II cutoff, the PPV improved. Seventy-six percent of children who were identified as delayed on the 13-month BINS were actually delayed on the MDI at 24 months.

The BINS was not significantly correlated with the MDI, PDI, or composite score. This finding held for both the 6-month and the 13-month BINS, regardless of the classification used.

Receiver Operator Characteristic Curve

The BINS score at 6 months did no better than chance for predicting developmental delay at 24 months using the classification of MDI = 85 (area under the ROC curve = .45). Similar results were found at 13 months (area under the ROC curve = .54). These results were consistent for the classification of MDI = 90, PDI = 85, PDI = 90, and composite = 85 or 90 (area under the ROC curve ranged from .39 to .63).

Table VI. Positive Predictive Value Using Different Categories of Scores on 6- and 13-Month BINS and 24-Month BSID-II

	BINS Score at 6 Months ^a		BINS Score at 13 Months ^b	
	Low and Moderate vs. High (%)	Low vs. Moderate and High (%)	Low and Moderate vs. High (%)	Low vs. Moderate and High (%)
BSID-II at 24 months				
90 Cutoff (0.75 SD)				
MDI	50	61	100	76
PDI	17	36	0	14
Composite	17	46	0	52
85 Cutoff (1.00 SD)				
MDI	17	36	0	52
PDI	0	7	0	5
Composite	17	18	0	10

BINS = Bayley Infant Neurodevelopmental Screener; BSID-II = Bayley Scales of Infant Development, second edition; MDI = motor developmental index; PDI = psychomotor developmental index.

^a 106 children completed both the 6- and 24-month evaluations.

^b 101 children completed both the 13- and 24-month evaluations.

Discussion

Using 6- and 13-month BINS scores to predict BSID-II scores at 24 months yielded low sensitivity values but high specificity values. Results for sensitivity and specificity were similar regardless of whether delayed development was defined as BSID-II scores below 85 (1.00 SD below the mean) or below 90 (0.75 SD below the mean). Positive predictive value was higher for the BSID-II criterion cutoff below 90 than below 85. Sensitivity and PPV were better when BINS risk groups were defined as low (no delay) versus moderate and high (delay) than when BINS risk groups were defined as low and moderate (no delay) versus high (delay). Specificity was better when BINS risk groups were defined as low and moderate (no delay) versus high (delay).

The children's scores on the BSID-II at 24 months indicated higher than expected rates of developmental delays. Although PDI scores were within normal limits, the mean MDI scores were approximately one standard deviation below the mean score expected within a normal sample. Thus, consistent with findings reported from other samples of low-income toddlers (Black, Hess, & Berenson-Howard, 2001; Burchinal, Campbell, Bryant, Wasik, & Ramey, 1997; Luster & McAdoo, 1996), the detrimental effect of environmental risk factors on children's development was evidenced by 24 months.

Although it has been argued that the BINS can be used reliably with children from low socioeconomic backgrounds (Aylward & Verhulst, 2000), we have

found that it has low predictive validity with an environmental risk group. The high specificity values in our study indicated that most children who were not delayed on the BSID-II at 24 months were accurately identified as not delayed on the screening tool at 6 and 13 months. However, the low sensitivity values indicated that infants who were delayed on the BSID-II at 24 months were not identified as delayed on the screening tool at 6 or 13 months. Despite the relatively high base rate of delayed cognitive development in our sample (i.e., 46% of children had an MDI < 85), the PPV of the BINS was only slightly higher than chance. The PPV was affected more by which BSID-II cutoff was used than was sensitivity or specificity, as it increased when the BSID-II cutoff was changed from below 85 to below 90. Prediction of children's development at 24 months from 6- and 13-month BINS scores was no better than chance when using ROC curve analyses. In this investigation, the low predictive validity of the BINS when using a clinically meaningful cutoff below 85 to identify those children with mild developmental delays suggests that the BINS alone is not an ideal measure for use with an environmental risk group of infants who have no biological risks. The BINS may be a better measure of risk for biological risk samples than for infants with environmental risks, because it includes items that capture skills and competencies in infancy that continue to be problematic for that group over time (Aylward et al., 1996; Aylward & Verhulst, 2000; Leonard et al., 2001; Macias et al., 1998). Preterm and low birth weight infants who score low on neurological and motor items in infancy may continue to score low on these items in toddlerhood.

Traditionally, measures of sensitivity and specificity have been used as one means of assessing the validity of screening tools. Recently, however, it has been argued that measures of sensitivity and specificity are not applicable to children's general development (Frankenburg, 2002) and should not be used to assess screening tools. Frankenburg (2002) suggests that validity assessments of screening tools should focus on the way in which developmental norms were established, that is, the method of standardization for the screener. Although standardization is indeed an important factor in assessing a tool's validity, it alone is not enough. In order to have clinical utility, it is essential to know how accurately the screener can predict children's performance on a full-scale assessment, and what the error rate is in making classification decisions. Only by assessing a screening tool's sensitivity and specificity is it possible to know whether a cost-effective screening tool can be

reliably and accurately substituted for a more lengthy and expensive developmental assessment tool during routine pediatric exams.

One possible explanation for the poor predictive validity of the BINS found for this low-income, urban risk group is that items on the BINS administered at the 6- or 13-month evaluations may not predict later developmental delays for environmental risk groups. Compared with low birth weight and premature infants, infants from low-income families are less likely to have neurological and motor difficulties, and thus are unlikely to score low on these items at early ages. Children from environmental risk groups tend to have more difficulty with language and problem-solving tasks than with neurological or motor tasks (Aylward, 1992). It is likely that other screening tools that assess sensorimotor and neurological skills would also lack predictive utility from infancy to toddlerhood for environmental risk groups. When routine screening is conducted during the first year of life with infants in environmental risk groups, screening tools should include more items to assess verbal and cognitive abilities so that they are more sensitive to detecting delays that children in environmental risk groups are likely to experience. In addition, it may be useful for screening tools to have domain subscales. We cannot rule out the possibility that gains in motor skills may have masked delays in language and cognitive areas, yielding lower overall sensitivity of the BINS for this environmental risk group. Use of domain scores for screening tools could help identify children's specific areas of weakness and alleviate the problem of children's gains in one area masking delays in other areas.

Another possible explanation for low predictive validity from the first to the second year of life for environmental risk groups is that risk for delayed development is low during the first year of life but increases over time. The canalization theory of development helps support this explanation. According to canalization theory, there is a species-specific process that helps ensure that development unfolds similarly for all children during early life (McCall, 1981). Strong self-righting processes protect young infants from all but the most devastating environmental influences (Bretherton, Bates, Benigni, Camaioni, & Volterra, 1979; Rutter, 1985). However, after infancy, development is less canalized, self-righting processes are less effective, and individual differences stabilize (Kopp & McCall, 1982). Variability in development becomes more apparent after infancy, as environmental influences exert themselves (Aylward, 1992). Low predictive validity of the BINS

may therefore be a result of increased risk for delayed development during the second year of life as protection from self-righting processes decrease.

Empirical evidence to support this explanation comes from several cross-sectional and longitudinal studies that have indicated that children from low-income, urban households are more vulnerable to delayed development during toddlerhood than during infancy (Black et al., 2001; Burchinal et al., 1997; Luster & McAdoo, 1996; Ramey & Campbell, 1991; Schweinhart & Weikart, 1989; Werner, Bierman, & French, 1971; Werner & Smith, 1982). Thus, children from low-income families who appear to be developing normally during infancy may be at heightened risk for developmental delays after infancy. If canalization theory explains the poor predictive validity of the BINS from infancy to toddlerhood, most other screening tests would be unlikely to detect developmental problems for this group during the first year, since problems do not tend to manifest themselves until toddlerhood. As children age, environmental factors have more time to influence development, and identification of delayed development becomes more likely. Unfortunately, many children with environmentally caused psychosocial problems are not identified in primary care (Regier, Goldberg, & Taube, 1978). Screening during the preschool years can be beneficial because when children with developmental delays are not identified and do not receive early intervention, they are at increased risk to enter school with substantial delays and to become high users of the health care system (Janicke & Finney, 2000).

Limitations

One limitation of the present study was the restricted range of scores on the MDI and PDI. In our sample, the standard deviation was about half of what is expected for the general population, probably because our sample comprised a relatively homogeneous group of children at environmental risk. Thus, predictive validity may have been affected by the restricted range on the 24-month outcomes. We also experienced attrition over the 2-year evaluation period. Although there were no differences on several demographic variables between those who did and did not complete follow-up evaluations, it is unclear whether there may have been differences between the families on the child outcome measures studied here. Another limitation was that the BINS was not measured during the second year of life (e.g., 18 months) to predict 24-month BSID-II performance, nor were the tests measured at concurrent age periods. Thus it is not

possible to determine whether the BINS could be effectively substituted for the BSID-II during the second year to predict 24-month performance, or whether it could adequately predict current, rather than future, performance. A fourth limitation was that the present study did not include a comparison group of low-risk children, so it is unclear whether the low predictive validity of the BINS is specific to children with environmental risk or is specific to the BINS as a screening tool. Future studies should compare this environmental risk group with other risk groups and with a normative pediatric group using the BINS, as well as other screening tools, to help determine whether it is the population or the tool that is the source of the low predictive validity.

Clinical Implications

Socially disadvantaged children remain a group at risk for developmental problems after infancy. In our sample, declines in development were evidenced by age 2 as measured by the BSID-II. Although few children in our sample experienced delays that were substantial enough to meet the criteria for early intervention under the Individuals with Disabilities Education Act (IDEA Amendments of 1997) it is likely that declines will continue unless developmental intervention is initiated. Close monitoring of this group is warranted. Use of screening tools that measure psychosocial risk is recommended for use in primary care settings. As the number of environmental risk factors increases, children's performance in cognitive and language domains decreases (Sameroff, Seifer, Barocus, Zax, & Greenspan, 1987; Hooper et al., 1998). If at-risk children are identified early, they can be referred to Early Head Start, Head Start, and preschool programs to receive early intervention services designed to help prevent developmental delays. Additional work is necessary to develop and validate psychosocial screeners. The Family Psychosocial Screener is an example of a promising tool that can help identify children exposed to environmental factors that may negatively affect their development (Kemper & Kelleher, 1996). It can be used in primary care settings as part of routine clinical intake to aid in the decision about which children need close monitoring and referral to early intervention or developmental enrichment programs.

In addition to professionally administered screening tests, tools that use parent report can be useful and cost-effective in documenting children's developmental progress during the first 2 years. There is evidence that parental concerns regarding specific aspects of their

child's development are highly correlated with results from developmental screening tests (Glascoe, 1997), suggesting the importance of relying on parental report for screening. During the first 2 years, emphasis should also include efforts to promote infants' health and well-being by ensuring that infants are up-to-date on their immunizations, by helping families develop healthy communication patterns with their infants (Black & Teti, 1997), and by providing anticipatory guidance to families (Slaby & Stringham, 1994; Zuckerman & Parker, 1995). A combination of surveillance and anticipatory guidance may be an effective strategy in working with environmentally at-risk children.

Acknowledgement

This research was supported by grant MCJ-240301 from the Maternal and Child Health Research Program, U.S. Department of Health and Human Services.

References

- Aylward, G. P. (1992). The relationship between environmental risk and developmental outcome. *Journal of Developmental and Behavioral Pediatrics*, *13*, 222–229.
- Aylward, G. P. (1994). *Practitioner's guide to developmental and psychological testing*. New York: Plenum Medical Book Company.
- Aylward, G. P. (1995). *The Bayley Infant Neurodevelopmental Screener*. San Antonio, TX: The Psychological Corporation.
- Aylward, G. P., & Verhulst, S. J. (2000). Predictive validity of the Bayley Neurodevelopmental Screener (BINS) risk status classifications: Clinical interpretation and application. *Developmental Medicine and Child Neurology*, *42*, 25–31.
- Aylward, G. P., Verhulst, S. J., & Bell, S. (1996). Predictive validity of the Bayley Neurodevelopmental Screener (BINS) risk status classifications. *Developmental Medicine and Child Neurology*, *138*(Suppl 74), 26.
- Bayley, N. (1993). *Bayley Scales of Infant Development, Second Edition*. San Antonio, TX: The Psychological Corporation.
- Black, M. M., Hess, C. R., & Berenson-Howard, J. (2001). Toddlers from low-income families have below normal mental, motor, and behavior scores on the revised Bayley scales. *Journal of Applied Developmental Psychology*, *21*, 655–666.
- Black, M. M., Siegel, E. S., Abel, Y., & Bentley, M. E. (2001). Home and videotape intervention delays early complementary feeding among adolescent mothers. *Pediatrics*, *107*, e67. Retrieved April 4, 2004, from <http://www.pediatrics.org/cgi/content/full/107/5/e67>
- Black, M. M., & Teti, L. (1997). Videotape: A culturally sensitive strategy to promote communication and healthy nutrition between adolescent mothers and infants. *Pediatrics*, *99*, 432–437.
- Bretherton, I., Bates, E., Benigni, L., Camaioni, L., & Volterra, V. (1979). Relationships between cognition, communication, and quality of attachment. In E. Bates (Ed.), *The emergence of symbols: Cognition and communication in infancy* (pp. 223–269). New York: Academic.
- Burchinal, M. F., Campbell, F. A., Bryant, D. M., Wasik, B. H., & Ramey, C. T. (1997). Early intervention and mediating processes in cognitive performance of children of low-income African American families. *Child Development*, *68*, 935–954.
- Committee on Children with Disabilities. (2002). Developmental surveillance and screening of infants and young children. *Pediatrics*, *108*, 192–202.
- Frankenburg, W. K. (2002). Developmental surveillance and screening of infants and young children. *Pediatrics*, *109*, 144–145.
- Frankenburg, W. K., Dodds, J., Archer, P., Bresnick, B., Maschka, P., Edelman, N., et al. (1996). *The DENVER II technical manual*. Denver, CO: Denver Developmental Materials, Inc.
- Frankenburg, W. K., Dodds, J., Archer, P., Shapiro, H., & Bresnick, B. (1992). The Denver II: A major revision and restandardization of the Denver Developmental Screening Test. *Pediatrics*, *90*, 477–490.
- Glascoe, F. P. (1997). Parents' concerns about children's development: Prescreening technique or screening test? *Pediatrics*, *99*, 522–528.
- Glascoe, F. P., Byrne, K. E., Ashford, L. G., Johnson, K. L., Chang, B., & Strickland, B. (1992). Accuracy of the Denver-II in developmental screening. *Pediatrics*, *89*, 1221–1225.
- Gordis, L. (2000). *Epidemiology* (2nd ed.). Philadelphia: W. B. Saunders.
- Harrell, F. E., Jr. (2001). *Regression modeling strategies: With applications to linear models, logistic regression, and survival analysis*. New York: Springer-Verlag.

- Hooper, S. R., Burchinal, M. R., Roberts, J. E., Zeisel, S., & Neebe, E. C. (1998). Social and family risk factors for infant development at one year: An application of the cumulative risk model. *Journal of Applied Developmental Psychology, 19*, 85–96.
- Individuals with Disabilities Education Act Amendments of 1997, PL 105-17, 20 U.S.C.
- Janicke, D. M., & Finney, J. W. (2000). Determinants of children's primary health care use. *Journal of Clinical Psychology in Medical Settings, 7*, 29–39.
- Kemper, K. J., & Kelleher, K. J. (1996). Family psychosocial screening: Instruments and techniques. *Ambulatory Child Health, 1*, 325–339.
- Kopp, C. B., & McCall, R. B. (1982). *Predicting later mental performance for normal, at-risk, and handicapped infants: Life-span development and behavior* (pp. 33–61). New York: Academic Press.
- Leonard, C. H., Picuch, R. E., & Cooper, B. A. (2001). Use of the Bayley Infant Neurodevelopmental Screener with low birth weight infants. *Journal of Pediatric Psychology, 26*, 33–40.
- Luster, T., & McAdoo, H. (1996). Family and child influences on educational attainment: A secondary analysis of the High/Scope Perry preschool data. *Developmental Psychology, 32*, 23–39.
- Macias, M. M., Saylor, C. F., Greer, M. K., Charles, J. M., Bell, N., & Katikaneni L. D. (1998). Infant screening: The usefulness of the Bayley Infant Neurodevelopmental Screener and the Clinical Adaptive Test/Clinical Linguistic Auditory Milestone Scale. *Journal of Developmental and Behavioral Pediatrics, 19*, 155–161.
- Maryland Medicaid Program (2003, November 5). Retrieved April 2, 2004, from <http://www.dhmh.state.md.us/mma/mmahome.html>
- McCall, R. B. (1981). Nature-nurture and the two realms of development: A proposed integration with respect to mental development. *Child Development, 52*, 1–12.
- Ramey, C. T., & Campbell, F. A. (1991). Poverty, early childhood education, and academic competence: The Abecedarian experiment. In A. C. Huston (Ed.), *Children in poverty: Child development and public policy*. Cambridge, MA: Cambridge University Press.
- Regier, D., Goldberg, I. D., & Taube, C. A. (1978). The de facto U.S. mental health services system: A public health perspective. *Archives of General Psychiatry, 35*, 685–693.
- Rosenbach, M. L., & Gavin, N. I. (1998). Early and Periodic Screening, Diagnosis, and Treatment and managed care. *Annual Reviews of Public Health, 19*, 507–525.
- Rutter, M. (1985). Resilience in the face of adversity: Protective factors and resistance to psychiatric disturbance. *British Journal of Psychiatry, 147*, 598–611.
- Sameroff, A. J., Seifer, R., Baldwin, A., & Baldwin, C. (1993). Stability of intelligence from preschool to adolescence: The influence of social and family risk factors. *Child Development, 64*, 80–97.
- Sameroff, A. J., Seifer, R., Barocus, R., Zax, M., & Greenspan, S. (1987). Intelligence quotient scores of 4-year-old children: Social-environmental risk factors. *Pediatrics, 79*, 343–350.
- Schendel, D. E., Stockbauer, J. W., Hoffman, H. J., Herman, A. A., Berg, C. J., & Schramm, W. F. (1997). Relation between very low birth weight and developmental delay among preschool children without disabilities. *American Journal of Epidemiology, 146*, 740–749.
- Schweinhart, L. J., & Weikart, D. P. (1989). The High/Scope Perry preschool study: Implications for early childhood care and education. *Prevention in Human Services, 7*, 109–132.
- Slaby, R. G., & Stringham, P. (1994). Prevention of peer and community violence: The pediatrician's role. *Pediatrics, 94*, 608–616.
- U.S. House of Representatives. (1996). *Personal responsibility and work opportunity reconciliation act of 1996*. Conference report H.R. 3734. Report No. 104–725. Washington, DC: U.S. Government Printing Office.
- Werner, E. E., Bierman, J. M., & French, F. E. (1971). *The children of Kauai: A longitudinal study from the prenatal period to age 10*. Honolulu: University of Hawaii Press.
- Werner, E. E., & Smith, R. S. (1982). *Vulnerable but invincible: A longitudinal study of resilient children and youth*. New York: McGraw-Hill.
- Wood, N. S., Marlow, N., Costeloe, K., Gibson, A. T., & Wilkinson, A. R., for the EPICure Study Group. (2000). Neurologic and developmental disability after extremely preterm birth. *New England Journal of Medicine, 343*, 378–384.
- Zuckerman, B., & Parker, S. (1995). Preventive pediatrics—new methods of providing needed health services. *Pediatrics, 95*, 758–762.