Use of the Bayley Infant Neurodevelopmental Screener With Low Birth Weight Infants

Carol H. Leonard,1 PhD, Robert E. Piecuch,1 MD, and Bruce A. Cooper,2 PhD
1University of California, San Francisco, and 2California School of Professional Psychology

Objective: To examine the utility of the Bayley Infant Neurodevelopmental Screener (BINS) as a screening technique for premature, low birth weight infants.

Methods: One hundred thirty-three preterm infants <1,500 grams received a BINS assessment at mean adjusted age 6.8 months and a Bayley Scales of Infant Development, Second Edition (BSID-II) assessment at mean adjusted age 12.9 months. Infants’ BINS scores were compared to their BSID-II Mental Development Index (MDI) and Psychomotor Development Index (PDI) scores.

Results: The BINS score showed significant association with the MDI ($r = .40, p = .001$) and with the PDI ($r = .35, p = .001$). The BINS showed moderate predictive validity (67%–76%) for identifying lower functioning infants.

Conclusions: The BINS is a satisfactory screening tool for low birth weight infants when used in conjunction with other known biologic and social risk factors.

Key words: high risk infant; screening; low birth weight; Bayley Scales.

Low birth weight infants (≤1,500 grams) are at increased risk for neurologic abnormalities and developmental delays (Hack, Friedman, & Fanaroff, 1996; Veen et al., 1991). Although some disabilities, such as specific learning disabilities, may be subtle and not apparent until school age, other deficits such as cerebral palsy and significant developmental delay appear earlier. Early diagnosis can lead to interventions, such as physical therapy, occupational therapy, and infant stimulation programs, designed to facilitate optimal development. Screening and assessments in the first year of life are complicated because some infants may have transient tone differences that resolve by one year, and they may be going through periods of reorganization following their hospitalization as a newborn. Additionally, within the category of low birth weight, some infants will have additional medical risk factors such as asphyxia and maternal substance abuse and social risk factors, such as low socioeconomic status (SES), low maternal education, or teenage parenting (Bendersky & Lewis, 1994; Brazy, Goldstein, Oehler, Gustafson, & Thompson, 1993; Degraw et al., 1988; Kirby, Swanson, Kelleher, Bradley, & Casey, 1993; Korner et al., 1993; Msall et al., 1991; Thompson et al., 1997). The developmental course of the low birth weight infant is affected by these many factors.

Several techniques can determine if the infant is making normal developmental progress: neurologic examinations (Wildin et al., 1995; Wildin et al., 1997), parent questionnaires (Heiser, Grimmer, Metze, & Obladen, 1995; Ireton & Glascoe, 1995;
Squires, Bricker, & Potter, 1997), developmental screening techniques such as the Denver Developmental Screening Test II (Frankenburg, Dodds, Archer, Shapiro, & Bresnick, 1992), the Battelle (Glascoe & Byrne, 1993), the CAT/CLAMS (Rossman et al., 1994), and developmental assessments, such as the Bayley Scales of Infant Development, Second Edition (BSID-II) (Bayley, 1992), the Gesell (Knoblock & Pasamanick, 1974), and the Mullen Scales of Early Learning (Mullen, 1985).

The utility of developmental screening measures is affected by many factors: professional staff qualifications to administer, cost of the instrument, time to administer, and adequacy of the results (Dworkin, 1989; Frankenburg, Chen, & Thornton, 1988; Kopp & Kaler, 1989; Meisels, 1988). Full assessments such as the BSID-II Scales require 45 minutes to an hour to administer and require an advanced level of training and expertise in administration. In high-risk clinics set up to follow the growth and development of low birth weight infants, infants may receive either a neurological examination or a developmental assessment; or both examinations which have overlap in items that might be administered, particularly in assessing gross and fine motor skills.

The Bayley Infant Neurodevelopmental Screener (BINS) is a recently developed instrument designed specifically for a high-risk infant population based on an earlier screening instrument, the ENORS (Aylward, 1995; Aylward, Verhulst, & Bell, 1988). It consists of items from the BSID-II Scales that assess cognitive, social, language, gross, and fine motor skills. Typical items in the age ranges under study include reaching for and transferring blocks, looking for fallen items, types of vocalizations, and prewalking progression (scooting, stepping movements). The BINS also includes items that measure neurologically intactness, such as ratings of active and passive tone in the upper and lower extremities, and scoring of quality of movement of the upper and lower extremities.

Although using items from the BSID-II Scales, the scoring is different and the inclusion of tone and quality of movement items differentiates this screening test from other developmental screening tests and the BSID-II. The BINS was validated on a high-risk infant population as well as normal infants in test construction. One study has reported on its concurrent validity with two other instruments in a mixed high-risk infant (term and preterm) population (Macias et al., 1998). This study examines the relationship of the BINS scores at 6 months to later development as measured by the Bayley Scales at 1 year.

### Method

The study group consisted of 133 preterm infants weighing <1,500 grams born April 1994 through September 1997 and treated in an intensive care nursery at an urban university medical center. The longitudinal study of the outcome of these infants is approved by the university Committee on Human Research. The mean gestational age of the infants was 27.5 weeks (range: 24–34 weeks), and mean birth weight was 976 grams (range: 570–1,465 grams). The racial composition of the group was representative of our geographical catchment area with 57% white and 43% nonwhite. Table I presents characteristics of the study group.

<table>
<thead>
<tr>
<th>Table I. Characteristics of Infants (n = 133)</th>
<th>%</th>
<th>M ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gestational age (wks)</td>
<td>27.5</td>
<td>2.3</td>
</tr>
<tr>
<td>Birth weight (grams)</td>
<td>976</td>
<td>235</td>
</tr>
<tr>
<td>Male</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Appropriate for gestational age</td>
<td>88</td>
<td></td>
</tr>
<tr>
<td>Small for gestational age</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>Days in supplemental oxygen</td>
<td>29.6</td>
<td>33.5</td>
</tr>
<tr>
<td>Intracranial hemorrhage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>Grades I–II</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>Grades III–IV</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>Periventricular leukomalacia</td>
<td>4.5</td>
<td></td>
</tr>
<tr>
<td>Maternal education &gt;12 years</td>
<td>93</td>
<td></td>
</tr>
</tbody>
</table>

Following attendance at a BINS training workshop for providers and practice in administration and scoring supervised by a psychologist, the BINS was routinely administered by the physician or nurse in the program at the 6-month visit, and the BSID-II was administered by the psychologist at 1 year or soon thereafter. Ages were adjusted for prematurity. The mean age of testing on the BINS was 6.8 ± 1.4 months (Mean ± standard deviation) and on the BSID-II was 12.9 ± 1.2 months. The mean time between the BINS assessment and the BSID-II...
assessments was 6.1 ± 1.5 months. Five infants had their PDI scores deferred due to late arrival for appointments or in one case, an infant who was casted for club feet. All five infants were seen by the physician for a neurodevelopmental examination and were considered to have normal neurologic development. The number of subjects for the PDI data is 128 and for the MDI data is 133.

The BINS consists of 11–13 items for different age levels. The total number of items failed places the infant in a category of low, moderate, or high risk for developmental delay. Within the moderate risk range, a cut score is indicated at which optimal sensitivity and specificity are reached. Scores below this cut line are considered high moderate scores, indicating that the infant is approaching a high-risk range. Scores above the cut line are considered low-moderate scores, indicating that the infant is closer to the low-risk range. We could categorize an infant’s BINS score thus as 0 = low risk, 1 = low moderate risk, 2 = high moderate risk and 3 = high risk.

The BSID-II MDI and PDI have mean scores of 100, with a standard deviation of 15. Although traditional cut-offs of one and two standard deviations below the mean are often utilized, we included a third category of 1.5 standard deviations below the mean as this is the criterion for entrance to some infant special education programs. Since the lowest BSID-II MDI or PDI score that can be obtained is a categorical score of <50, the MDI and PDI scores cannot be used directly in mathematical calculations unless scores of <50 are assigned a common data point, such as 50 and the data are treated as continuous. We chose to do categorical analyses that are consistent with the noncontinuous form of the original data set. We therefore categorized the MDI and PDI scores as within or above one standard deviation from the mean, (scores ≥85); from 1 to 1.5 SD below the mean, (scores 77–84); 1.5–2 SD below the mean (scores 70–76); and at 2 or more two standard deviations below the mean (scores ≤69).

**Analyses**

The association (Spearman rho) between the infant’s BINS risk status and MDI score, and BINS risk status and PDI score, was calculated.

Positive predictive value (PPV) is a measure of the ability of the screening test to accurately identify low scoring infants, as measured by a later criteria. We used the Bayley Scales of Infant Development at 1 year as our criteria for development. Positive predictive values were calculated for the BINS and the MDI, and BINS and the PDI. We looked at positive predictive value comparing and contrasting the different categories of performance on both tests.

**Results**

The BINS showed a significant association with the BSID-II MDI score (Spearman rho = .40, p = <.001) and with the BSID-II PDI score (Spearman rho = .35, p = <.001), when four risk ranges on the BINS were used (low, low-moderate, high-moderate, and high) and the MDI and PDI scores were categorized into four ranges.

Ten of 15 infants obtaining high risk BINS scores at 6 months of age showed MDI scores of <85 at 1 year, giving a positive predictive value of 67% on the BINS screening (Table II). Similarly for the PDI, positive predictive value was 67%, with 10 of 15 infants designated high risk on the BINS actually having scores <85 on the BSID-II PDI.

In Table II, the highest PPVs are seen when the low and moderate BINS scores are considered versus high risk BINS scores. This is true for the MDI scores, regardless of whether the MDI scores grouped −1, −1.5, or −2 standard deviations from the mean. For the PDI, the same trend is seen with the exception that for infants with PDIs of −1 standard deviation (>85 versus ≥84), PPV increased when the BINS scores were grouped as low and low-moderate versus high-moderate and high. The PPV increased from 67% to 72% in this grouping.

Since the BINS is not specific for cognitive versus motor deficits, an analysis was done looking at infants who had either an MDI or a PDI or both scores <85. The predictive validity of the BINS (grouping low and low-moderate versus high-moderate and high risk BINS scores) was 76%.

Among the false negatives were eight infants who scored in the low risk range on the BINS, but who obtained both BSID-II MDI and PDI scores of <85. Four additional infants scored in the low risk range on the BINS and had MDI scores of <85, with PDI scores ≥85. There were no infants in the remaining category of low risk on BINS with MDI ≥85 and PDI <85. (Table III).

As Table III shows, of the eight infants scoring <85 on the both MDI and PDI, seven infants were in the gestational age range 24–26 weeks. In the
total study group of 133, 39% of the sample were in the gestational age range of 24–26 weeks. The eighth infant in this group of infants on both the MDI and PDI was born at 29 weeks gestation and had a diagnosis of cerebral palsy at 1 year that was not identified at 6 months.

Since seven of eight false negatives, or infants who were low risk on the BINS but who scored <85 on the Bayley MDI and PDI, were of extremely low gestation, we separated the study group by gestational age (27 weeks versus 26 weeks) and reanalyzed the data. With the BINS categorized into low versus moderate and high risk scores, and Bayley MDI scores within or 1.0 SD below the mean (<85 versus ≥84), the PPV was 33% for the ≥27 weeks’ gestation infants and 59% for the infants ≤26 weeks’ gestation. For the PDI, the PPV was 65% for the ≥27 weeks’ gestation infants and 66% for the ≤26 weeks’ gestation infants.

Reasons for poor performance on the BSID-II PDI were those generally seen in this low birth weight population: infants of 12–13 months adjusted age who were showing a normal progression of skills (crawling, pulling to stand, cruising) but who may not have been standing alone, or taking steps alone. One infant had a diagnosis of cerebral palsy. Based on review of test reports commenting on performance for each infant in the areas of gross motor skills, fine motor skills, language skills, problem-solving skills and behavior, the reasons for poor scores on the MDI fell into three patterns: (1) poor attention with increased activity level interfer-
Discussion

By definition a screening test casts a wide net to select those children who may need closer monitoring. Any screening test will make classification errors; both in identifying infants as abnormal who were later found to be normal and in underidentifying infants who presented later with deficits. This
study examined the relationship of the 6-month BINS score to the 12-month Bayley scores. The Bayley Scales of Infant Development, as with other infant tests, does not predict long-term outcome. Infant tests are measures of present developmental status and are widely utilized as formal indicators of need for timely intervention services.

This study examined the predictive value of using the BINS at a typical 6-month adjusted age examination of low birth weight infants to screen for emerging disabilities, as a lower cost measure than administering a full Bayley assessment at 6 months, and as able to be incorporated in the physician’s examination of the infant. This represents cost efficiency in staff time and length of visit for the patient.

The BINS was moderately effective in correctly predicting as low risk those infants at 6 months who later had a normal developmental assessment on the BSID-II and in identifying abnormal infants who remained abnormal at 12 months. Consideration of the efficacy of any screening measure depends on the base rate in the population being screened. A base rate of 50% normal and 50% abnormal is likely to be the most efficacious for using a screening measure (Murphy and Davidshofer, 1988). In the premature, low birth weight infant population, more than 50% of infants are likely to have normal development at one year (Hack et al. 1996; Piecuch, Leonard, & Cooper, 1998; Piecuch, Leonard, Cooper, & Sehring, 1997). The BINS was able to predict 67%–72% of the low functioning infants correctly, given a likely base rate of abnormality of 20%–40% in this population.

The BINS manual is clear in acknowledging the moderate risk range as one requiring clinical judgment in deciding whether to further assess or refer infants obtaining a moderate risk score. The moderate risk range has been found to be a less consistent classification than initial classification in the low or high risk range in a study of BINS performance over the first 2 years of life for high-risk infants (Aylward, Verhulst, & Bell, 1996). In the age ranges in this study, infants had to miss only two or three items to fall in the moderate risk range. If an infant is at the lower end of the age range, failure of items may be more likely than at the upper end of the age range. In this study, however, 13/14 infants at the lower end of the age range obtained scores in a low risk range. The best clinical practice may be to report scores in the moderate risk range as low-moderate risk or high-moderate risk, utilizing the cut score on the record form, and to monitor those infants obtaining high-moderate scores more frequently than those obtaining low-moderate scores.

In the total group of 133 infants, some infants in each BINS risk range were receiving services from infant programs by 1 year of age. Eleven percent (9/84) of infants who had BINS scores in the low risk range were enrolled in an infant program as were 25% (4/16) of infants who had a low-moderate risk score, 22% (4/18) who had a high-moderate risk score, and 27% (4/15) who had a high risk BINS score. Reasons for referral to the programs varied with “prematurity” as the usual intake reason.

The BINS at 6 months would not be expected to accurately identify subtle aspects of development emerging at 1 year of age: refinements in fine motor skills, progress in receptive and expressive language, efficiency in imitation, and problem solving skills. The BINS at 6 months may also not accurately identify developmental delay that is not neurologically based but may reflect subtle aspects of behavioral organization and attention span that are expected to emerge by 1 year of age.

Many aspects of development are affected by factors in the social environment, such as adequacy of parenting, levels of parental education, and cultural differences in child rearing practices, among others. Medical risk factors such as significant intracranial hemorrhage, periventricular leukomalacia, and maternal substance abuse in the high risk infant's medical history may place him or her at risk for development. Extremely low gestational age may be a biological risk for later cognitive skills. Our finding of the preponderance of infants at 24–26 weeks gestational age who had low risk BINS scores but who gave a poor performance on the BSID-II at 1 year may indicate a medical risk factor of very early gestational age, which has been reported elsewhere (Piecuch et al., 1997). Serial evaluations may reveal a pattern of change related to biological risk, similar to those observed by Liaw and Brooks-Gunn (1993) in low birth weight infants.

Prediction of developmental outcome at 1 year from a 6-month screening for a high-risk infant is a complicated issue. In addition to social factors and medical risk factors, there is the physical recovery from the neonatal hospitalization and the resilience of the infant, which affects the trajectory of developmental outcome. No screening tool at 6 months
is likely to accommodate to a perfect degree the multitudinous effects on an infant's development at 1 year. Social or biologic risk factors, and perhaps their additive effects, as discussed by Aylward (1992), continue to be an important aspect of the assessment of developmental risk to be considered in conjunction with test scores in initiating referral for services. Our clinical experience with the BINS and this study suggest that it is a useful tool.

Received May 28, 1999; revisions received October 14, 1999; accepted February 29, 2000

References


