Journal of Pediatric Psychology, 41(5), 2016, 573–581
doi: 10.1093/jpepsy/jsv101
Advance Access Publication Date: 4 November 2015
Original Research Article

OXFORD

Developmental Screening Among Children Born Preterm in a High-Risk Follow-Up Clinic

Allison G. Dempsey,[1] PhD, Catherine W. Abrahamson,[2] EDM, and Milena A. Keller-Margulis,[2] PhD

[1]University of Texas Health Science Center at Houston and [2]University of Houston

All correspondence concerning this article should be addressed to Allison G. Dempsey, PhD, Department of Pediatrics and the Center for Clinical Research and Evidence-Based Medicine, University of Texas Health Science Center at Houston, Houston, TX 77030, USA. E-mail: Allison.dempsey@uth.tmc.edu

Received May 6, 2015; revisions received September 28, 2015; accepted October 2, 2015

Abstract

Objective The psychometric properties of two formats of developmental screening tools that may be used in follow-up clinics providing primary care to children born preterm are presented. **Methods** 28 children born extremely preterm (<27 weeks) attending a high-risk clinic at the time of their 18–24 month visit were administered the Child Development Review, Brigance Early Head Start Screen II, and Bayley Scales of Infant and Toddler Development—Third Edition. **Results** Both screeners identified the majority of the sample as at-risk. The Brigance Screen II more accurately identified children at-risk compared with the Child Developmental Review (sensitivity: 1.00 and 0.44; specificity: 0.60 and 0.80; positive predictive value: 79% and 80%; negative predictive value: 100% and 44%, respectively). **Conclusions** Screening assessments using direct skills assessment may be an efficient and effective method of identifying children with developmental delays, particularly high-frequency but lower severity difficulties, in high-risk follow-up care settings.

Key words: developmental disabilities; low birthweight; neonatology; prematurity.

Pediatric psychologists have the assessment training to identify, adopt, administer, and interpret appropriate screening tools for developmental delays. This allows them to oversee screening of children born preterm in high-risk follow-up clinics. According to the 2006 American Academy of Pediatrics (AAP) guidelines (Council on Children with Disabilities, 2006), developmental screening with reliable and valid measures should occur at 9, 18, and 30 months of age during well-child visits.

The screening process in high-risk clinics is particularly essential for children born preterm, who are already at increased risk for developmental delays, cognitive impairments, academic challenges, and behavior difficulties (Doyle, 2001; Hintz, 2005). It is well-established that as the degree of prematurity increases, the likelihood of developmental delays increases (Aarnoudse-Moens, Weisglas-Kuperus, van Goudoever, & Oosterlaan, 2009; Bhutta, Cleves, Casey, Cradock, & Anand, 2002). However, despite guidelines for screening, many children do not receive available support services. Roughly half of children born preterm with moderate to severe disabilities and three-fourths of children with mild disabilities do not receive early intervention services by age 2 years (Roberts et al., 2008). High-risk clinics may have higher referral rates and use of early intervention services (Keller-Margulis, Abrahamson, Llorens, & Dempsey, 2013). The purpose of this study was to examine the technical adequacy of two screening

measures using different assessment methodologies for identifying developmental delays and need for intervention services among children born preterm.

Identification of Developmental Delays

In a review of considerations in selecting screening measures, Jenkins, Hudson, and Johnson (2007) discuss two indicators of a screening tool's validity: criterion validity and classification accuracy (diagnostic accuracy). In criterion validity, the correlation of the screening measure with an established, reference-standard assessment tool is examined using continuous scores on both measures. In contrast, when describing classification accuracy, both scores are treated as dichotomous (at-risk or not for the screening measure and delayed or not for the reference-standard tool). The classification accuracy is described by the instrument's sensitivity and specificity. Sensitivity is the probability that a screening measure will accurately identify those children identified with a developmental delay using the reference-standard assessment. That is, sensitivity refers to the measure's ability to identify true positives. In contrast, specificity is the probability that a screener will not identify children who do not have delay as indicated by performance on the reference-standard assessment. In other words, specificity indicates the screening measure's ability to identify true negatives. A perfect screener has 100% sensitivity and specificity, thus resulting in no false negatives and no false positives. The AAP's 2006 guidelines state that screening tools should be technically adequate, with sensitivity and specificity ranging from 70% to 80% (Council on Children with Disabilities, 2006). However, given the high-risk nature of children born preterm, clinicians may wish to select a tool with higher sensitivity (thus possibly sacrificing specificity).

Types of Screening Measures

Developmental screeners that involve parent report of skills, such as the Ages and Stages Questionnaire, Third Edition (Squires, Bricker, & Potter, 2009) and Child Developmental Review (CDR; Ireton & Vader, 2004), which is part of the Child Developmental Inventories (Ireton, 1992), are widely adopted owing to their ease of use and minimal clinician administration time (Gollenberg, Lynch, Jackson, McGuinness, & Msall, 2010; Rydz et al., 2006). Parent report of language skills in early development has demonstrated high correlations with direct assessment of language delay (Sachse & Suchodoletz, 2008). In contrast, parent report tools used in general developmental screening of multiple functional domains in community samples yield sensitivity levels that are generally below AAP guidelines, with sensitivity ranging from 0.50 to 0.67 and specificity ranging from 0.39 to 0.86 (Rydz et al., 2006). Additionally, parent and clinician agreement regarding a child's developmental level may vary across skills (Bortolus et al., 2002; Simard, Luu, & Gosselin, 2012).

In contrast, clinician-administered, direct assessment screening measures have the benefit of direct assessment of skills. Examples of direct assessment screening measures include the Brigance Early Head Start Screen II (Brigance Screen II; Glascoe, 2010) and Bayley Infant Neurodevelopmental Screener (Aylward, 1995). They can typically be administered in less than 15 min. Drawbacks to their implementation include administration time, concerns that child behaviors and fatigue during the assessment may affect validity of results, low compensation for clinician time, and lack of guidance from professional associations regarding instrument selctions (for a review, see Rydz et al., 2005).

Current Study

Although there are advantages and disadvantages to both screening approaches, there has been little work comparing screening methods among populations of children born preterm. This study explored child performance on the Brigance Screen II and parental responses/unstructured clinician observations on the CDR. Children's scores on both were compared with a reference-standard assessment (Bayley Scales of Infant and Toddler Development—Third Edition [Bayley III]; Bayley, 2006) among children in a high-risk follow-up clinic. Specifically, this study will answer the following research questions related to the use of two screening measures among children born preterm at 18–24 months adjusted age: (1) What is the rate of positive screens on the Brigance Screen II and CDR? and (2) How do scores on the two screening measures relate to identification of developmental delays on a reference-standard measure, the Bayley III?

Method

Participants

Participants were 28 children recruited from a high-risk infant follow-up clinic over a 3-year period starting in 2011. They are drawn from a larger sample ($n = 56$) of children participating in a longitudinal study assessing developmental screening of children born preterm in early childhood. The clinic serves children born preterm (<29 weeks gestational age) and other children with significant medical needs. Children were recruited at their 18–24 month well-child clinic visit or research follow-up visit. Children and their families were excluded if they were non-English speaking and/or if they were unable to be tested owing to deafness, blindness, or severe

neurological impairment. However, none of the children approached to participate presented with deafness, blindness, or severe neurological impairment. Only 28 of the 56 children in the larger study had scores for all three measures reported in this study, as one of the measures (Bayley III) was administered as part of a different study and signed release of information was obtained to use the scores from that measure for research. The mean chronological age of the current sample was 22.25 ($SD = 3.47$) months and, after adjusting for prematurity, the mean age was 18.82 months ($SD = 3.54$). To calculate adjusted age, the number of weeks born before 40 weeks is subtracted from the child's chronological age (Committee on Fetus and Newborn, 2004). For example, if the child was born at 25 weeks gestational age, 15 weeks ($40 - 25 = 15$) is subtracted from the child's chronological age to create an adjusted age. Although controversy exists regarding in what situations and with which groups of infants to use adjusted age (for a review, see Wilson & Cradock, 2004), we adopted the guidelines to correct for all scores in this study. It is common practice to adjust for prematurity in very and extremely premature infants until age 2 years (Aylward, 2002) and is consistent with the practices of the Neonatal Research Network practices for developmental follow-up in early childhood (Vohr et al., 2012). Demographic characteristics of the sample, life-saving interventions at birth, and current medical conditions are reported in Table I.

Measures

Child Development Review

The CDR includes open and closed-ended clinician-administered interview questions and a chart of developmental milestones across five domains that are assessed via parent response and direct observations (Ireton & Vader, 2004). The CDR is a two-page form that requires completion of open-ended questions regarding general behaviors and a checklist of potential problems experienced by children. The Child Development Chart is the second page of the CDR and is completed using both parent report and direct, unstructured observation of the child's spontaneous behaviors and skills across five domains (social, self-help, gross motor, fine motor, and language). The clinician completes the chart using parent report and clinician observations of the child's skills, based on whether the child (a) displays the behavior/skill regularly or performs the behavior "pretty well" (indicated with a check mark), (b) is just beginning to display the behavior/skill or only displays the behavior/skill "sometimes" (indicated with a B), or is not yet displaying the behavior/skill. A positive screen (indicating need for further assessment) is determined by drawing a horizontal line across the chart at the child's current age and another line at 70% of the child's age. Because the sample in this study was premature, lines were drawn at adjusted age and 70% of adjusted age. The CDR has been demonstrated to identify children with developmental delays, health issues, and behavior problems and has adequate sensitivity (0.68) and specificity (0.88) in a study of 3- and 4-year-olds screened in a community sample for developmental problems (Ireton & Vader, 2004).

Brigance Early Head Start Screen II

The Brigance Early Head Start Screen II (Brigance Screen II) is a norm-referenced, standardized measure of development for children ages 0–35 months used to assess language, physical health and development, and social-emotional development through direct assessment or observation (Glascoe, 2010). The screening measure can be administered in approximately 15 min. The Toddler Screen was administered to participants between the ages of 18 and 23 months and the Two-Year-Old Screen was administered to participants between the ages of 24 and 29 months.

Table I. Demographic and Medical Characteristics of Study Sample

Characteristic	All ($N = 28$)
	n (%)[a]
Male	13 (46.4%)
Race/ethnicity	
Black	19 (67.9)
White	3 (10.7)
Hispanic/Latino	6 (21.4)
Maternal education	
Did not finish high school	3 (10.8)
High school or equivalent	5 (17.9)
Some college	10 (35.7)
4-year college or more	7 (25.0)
Not reported	3 (10.7)
Gestational age at birth (weeks), *M* (*SD*)	24.75 (1.01)
23 weeks	4 (14.3)
24 weeks	6 (21.4)
25 weeks	11 (39.3)
26 weeks	7 (25.0)
Birthweight (grams), *M* (*SD*)	734.75 (114.42)
Length of stay in NICU (days), *M* (*SD*)	121.86 (59.98)
Received early childhood intervention services	12 (42.9)
History of phototherapy treatment	21 (75.0)
History of mechanical ventilation treatment	26 (92.9)
History of CPAP treatment	28 (100)
Current retinopathy of prematurity	10 (35.7)
Current neurological disorder	2 (7.1)
Current feeding problems	6 (21.4)
Current cerebral palsy	1 (3.6)
Current chronic lung disease	5 (17.9)
Current other medical condition	13 (46.4)

Note. NICU = Neonatal Intensive Care Unit; CPAP = Continuous Positive Airway Pressure. [a]Results presented as *n* (% of column *N*) unless otherwise specified as mean (*SD*).

None of the children had current cardiac conditions or chronic kidney disease.

Adjustment to the participant's chronological age was used for all children included in the study to determine the appropriate screen. Both versions of the Brigance Screen II yield raw scores that are totaled and weighted to produce a total score. There are cutoff scores for children at various age ranges that indicate whether the child is likely to have a developmental disability and/or academic delay, indicating a need for a referral for full testing. The screens also provide standardized cluster scores in communication/language and nonverbal/motor, though only the overall score was used for the purpose of this study.

Bayley Scales of Infant and Toddler Development—Third Edition

The Bayley Scales of Infant and Toddler Development—Third Edition (Bayley III) is a standardized, norm-referenced developmental battery that provides information regarding children's developmental skills across cognitive, language, and motor domains (Bayley, 2006). The measure is used for children 0–48 months of age and has a five-factor solution, producing three total domain scores (Cognition, Language, Motor) and five subdomain scores (cognitive, receptive language, expressive language, fine motor, and gross motor).

Procedures

The study was approved by the institutional review boards at the two institutions of the authors. Families were invited for participation in the study during well-child or research visits in the follow-up clinic. Parents signed a release to allow study personnel to collect specific medical history information from clinic staff (information included in Table I), as well as scores on the Bayley III, which was administered as part of a different series of research projects by research-approved examiners. Research examiners for this study were blinded to gestational age and weight at birth and medical history. Trained doctoral students administered study measures (the Brigance Screen II, CDR, and Bayley III) and parents completed a demographic form. Examiners were not blinded to performance across the three measures, but measures were not scored until visit completion to reduce scoring and administration bias. Adjusted age was used to calculate scores on all measures.

Data analytic procedures included descriptive statistics for performance on the CDR and Brigance Screen II. Next, scores on the Brigance Screen II were examined using correlations with standard scores on the Cognitive, Language, and Motor Bayley III scores, treating all scores as continuous (criterion validity). The CDR does not produce raw or standard scores so its criterion validity was not examined. Next, rate of positive screens on CDR and Brigance Screen II was compared with identification of delay on the Bayley III using a series of 2×2 contingency tables (classification accuracy).

Presence of delay was defined as a standard score <85 on the Cognitive, Language, or Motor domain of the Bayley III. Research involving both preterm and other children with chronic conditions has indicated that the Bayley III produces potentially inflated scores compared with the previous edition (Moore, Johnson, Haider, Hennessy, & Marlow, 2012; Vohr et al., 2012) and a substantial decrease in the proportion of children born preterm identified with neurodevelopmental delays, based on cognitive scores <70. Therefore, for the current study, delay was determined by a Cognitive, Language, or Motor score >1 standard deviation below the mean (i.e., standard score <85).

Any domain rating $<70\%$ of the child's adjusted age on the CDR was considered a positive screen. Finally, a positive screen on the Brigance Screen II was calculated using both the raw score cutoff from the manual and a standard score of <85. Sensitivity was calculated as the probability that a positive screening was consistent with a Bayley III score <85. Specificity was calculated as the probability that a negative screening was consistent with a Bayley III score >85. Positive predictive validity (PPV) was the probability that a child with a score <85 on the Bayley III had a positive screen on the screening measure, and negative predictive validity (NPV) was the probability that a child with a score >85 on the Bayley III had a negative screen. False-positive and false-negative rates were also examined.

Results

A total of 28 children were included in this study. Data entry was checked by randomly sampling 20% of the data and comparing the responses on the measures to the data entry database. For the Brigance Screen II, after adjusting for prematurity, 25 participants fell in the 12–23 month window and 3 fell in the 24–29 month window. Three of the 28 children were administered the wrong version for their adjusted age and so were excluded from analyses. This resulted in 25 of the 28 children included in analyses involving the Brigance Screen II scores.

Performance on Screening Measures

Descriptive statistics for the variables of interest are included in Table II. The mean Brigance Screen II score was below the average range. Based on raw score cutoffs provided in the Brigance Screen II manual (which differed depending on the age of the child), 19 of 25 children (67.9%) had a positive screen. Additionally, 17 (60.7%) children had a standard score <85 on the Brigance Screen II.

Table II. Performances on the Screening and Reference-Standard Measures

Measure	All		Subset included in Brigance Screen II analyses ($n = 25$)
	N	*n* (%)[a]	*n* (%)[a]
Brigance Screen standard score, *M* (*SD*)	25		77.76 (12.83)
Brigance Screen raw score below cutoff score	25		19 (76.0%)
Brigance Screen standard score < 85	25		17 (68.0%)
Brigance Screen standard score < 70	25		7 (28.0%)
Child Developmental Review score < 70%	25		
Any domain	25		9 (36.0%)
Social	25		6 (24.0%)
Self-help	25		1 (4.0%)
Fine motor	25		3 (12.0%)
Gross motor	25		2 (8.0%)
Language	25		4 (16.0%)
Bayley III standard score, *M* (*SD*)	28		
Cognitive composite	28	85 (13.95)	87.00 (13.31)
Receptive communication	28	7.00 (1.90)	7.16 (1.86)
Expressive communication	28	6.71 (2.86)	6.88 (2.79)
Language composite	28	82.03 (12.36)	83.00 (11.81)
Fine motor	28	8.25 (2.78)	8.52 (2.81)
Gross motor	28	7.86 (2.69)	8.08 (2.72)
Motor composite	28	88.18 (15.70)	89.64 (15.85)
Bayley III standard < 85			
Cognitive Composite standard score < 85	28	9 (16.1%)	8 (32.0%)
Language Composite standard score < 85	28	16 (57.1%)	14 (56.0%)
Motor Composite standard score < 85	28	11 (39.3%)	8 (32.0%)
Bayley III standard < 70			
Cognitive Composite standard score < 70	28	3 (10.7%)	2 (8.0%)
Language Composite standard score < 70	28	4 (14.3%)	3 (12.0%)
Motor Composite standard score < 70	28	4 (14.3%)	3 (12.0%)

Note.[a]Results presented as *n* (% of column *N*) unless otherwise specified as mean (*SD*).

Of the 28 children with completed CDR forms, 10 (35.7%) had a positive screen on the CDR based on a score <70% on any domain. Furthermore, seven (25.0%) had a positive screen on the social domain, one (3.6%) on the self-help domain, two (7.1%) on the gross motor domain, three (10.7%) on the fine motor domain, and five (17.9%) on the language domain. According to parent response to questions whether there were concerns about the child's health or development, 11 (39.3%) parents indicated a concern regarding their child's health and/or development.

Finally, scores on the Bayley III Cognitive, Language, and Motor composite scores were in the low average range. Eighteen (64.3%) of the children had one or more composite score <85; six (21.4%) of the children had one or more composite score <70.

Brigance Screen II and Bayley III

Continuous Scores

The results of the correlations used to examine relationships among Brigance Screen II and Bayley III continuous scores (criterion validity) are displayed in Table III. The results indicate that the Brigance Screen II standard score demonstrated moderate, positive, and significant correlations with the Bayley III Cognitive, Language, and Motor composite scores. The Brigance Screen II had the lowest correlation with the Bayley III Cognitive composite ($r = .52$, $p < .01$). The Brigance Screen II had the highest correlation with the Bayley III Language composite ($r = .78$, $p < .01$).

Dichotomous Scores: Raw Score Cutoffs

The results of the analyses comparing dichotomous scores on Brigance Screen II and Bayley III (classification accuracy) are displayed in Table IV ($n = 25$). All children with a Bayley III score <85 had a positive screen on the Brigance Screen II, based on raw score cutoffs (sensitivity = 1.0). Of the 10 children with all Bayley III composite scores >85, 6 had a negative screen on the Brigance Screen II (specificity = 0.60). Results indicate that in this sample the Brigance Screen II produced no false negatives (negative screen when a delay exists) when using raw score cutoffs. However, 40% of those with a positive screen did not have a developmental delay on any domain on the Bayley III.

Table III. Correlation Between Brigance Early Head Start Screen II and Bayley Scales of Infant Development Scores

		1	2	3	4	5	6	7
1.	Brigance standard score	–						
2.	Bayley Cognitive Composite standard score	.52*	–					
3.	Bayley Receptive Language Scale score	.69**	.57*	–				
4.	Bayley Expressive Language Scale score	.65**	.58*	.50*	–			
5.	Bayley Language Composite standard score	.78**	.66**	.80**	.92**	–		
6.	Bayley Fine Motor Scale score	.56*	.79**	.61**	.56*	.67**	–	
7.	Bayley Gross Motor Scale score	.67**	.72**	.50*	.53*	.61**	.74**	–
8.	Bayley Motor Composite standard score	.66**	.82**	.59**	.57*	.67**	.93**	.92**

Note. Includes Brigance 12–23 months scores and Brigance > 24 months score (only one case); $n = 28$ for Bayley–Bayley correlations, $n = 25$ for Bayley–Brigance correlations.
*$p \leq .01$; **$p \leq .001$.

Table IV. Psychometric Properties of Brigance Early Head Start Screen II and Child Development Review at Predicting Deficits on Reference-Standard Measure, the Bayley III

Classification accuracy statistics	Bayley Any SS < 85	Bayley Cognitive SS < 85	Bayley Language SS < 85	Bayley Motor SS < 85
Brigance cutoff scores				
Sensitivity	1	1	1	1
Specificity	0.60	0.35	0.55	0.35
PPV	0.79	0.42	0.74	0.42
NPV	1	1	1	1
False positive rate	0.40	0.65	0.46	0.65
False negative rate	0	0	0	0
Brigance SS <85				
Sensitivity	0.87	0.75	0.93	0.88
Specificity	0.60	0.35	0.64	0.41
PPV	0.77	0.35	0.77	0.41
NPV	0.75	0.75	0.88	0.88
False positive rate	0.40	0.65	0.36	0.59
False negative rate	0.13	0.25	0.07	0.13
CDR any score <70%				
Sensitivity	0.44	0.56	0.43	0.55
Specificity	0.80	0.74	0.75	0.77
PPV	0.80	0.50	0.70	0.60
NPV	0.44	0.78	0.50	0.72
False positive rate	0.20	0.26	0.25	0.24
False negative rate	0.56	0.44	0.44	0.46

Note. SS = standard score; PPV = positive predictive value; NPV = negative predictive value.

Dichotomous Scores: Standard Scores <85

Of the 15 children with a Bayley III score <85, 13 of them had a positive screen on the Brigance Screen II based on standard score <85 (sensitivity = 0.87). Of the 10 children with all Bayley III composite scores >85, 6 had a negative screen on the Brigance Screen II (specificity = 0.60). Results indicate that in this sample, 13% with a negative screen on the Brigance Screen II had at least 1 Bayley III composite score <85. Additionally, 40% of those with a positive screen did not have a developmental delay on any domain on the Bayley III.

CDR and Bayley III

The results of the analyses comparing dichotomous scores on CDR and Bayley III (classification accuracy) are displayed in Table IV ($n = 28$). Of the 18 children with a Bayley III score <85, 8 of them had a positive screen on the CDR (sensitivity = 0.44). Of the 10 children with all Bayley III composite scores >85, 8 had a negative screen on the CDR (specificity = 0.80). Results indicate that in this sample, 56% with a negative screen on the CDR had at least 1 Bayley III composite score <85. Additionally, 20% of those with a positive screen did not have a developmental delay on any domain on the Bayley III.

Discussion

The purpose of this study was to describe the use of screening tools using two different approaches in a sample of children born extremely preterm (<27 weeks) and followed in a high-risk follow-up clinic. As children born preterm are at significant risk of developmental delay (Aylward, 2002; Saigal & Doyle, 2008), adoption of screening tools sensitive for detecting delays in an already high-risk population is critical. Screening measures with high sensitivity (and a low false-negative rate) will assist clinicians in quickly and accurately identifying children in need for a more thorough developmental assessment and/or intervention. However, with higher sensitivity, there is increased risk for a higher false-positive rate (i.e., children without delays will be referred for further evaluation). Given the high base rate of developmental delays and high-risk nature of children born preterm, a higher rate of false positives is preferable to not detecting a child in need of further evaluation and treatment.

The results of this describe the use of two screening measures, the CDR and the Brigance Screen II, in a high-risk clinic setting. As expected, given the high base rate of developmental delays among children born <27 weeks gestational age (Vohr, Wright,

Poole, McDonald, & NICHD Neonatal Research Network Follow-up Study, 2005), the majority of children were identified as at-risk using both parent report supplemented with direct, unstructured clinician observation (the CDR) and clinician-administered (Brigance Screen II) approaches. A much greater percentage of children were identified as at-risk based on the Brigance Screen II, which is a direct skills measure, when compared with the CDR, which does not rely on direct, structured skill assessment and is instead based on parent report supplemented with clinician observation of child skills.

The Brigance Screen II exceeded recommended levels of sensitivity, particularly when using raw score cutoffs. No children identified with a developmental delay on the Bayley III were missed (i.e., had a negative screen) when using the Brigance Screen II raw score cutoffs. However, specificity levels were below AAP recommendations (Council on Children with Disabilities, 2006), as 40% of children with a positive screen did not have a developmental delay based on Bayley III scores.

Notably, overall scores on the Bayley III were systematically higher than on the Brigance Screen II. The mean standard scores of the children in the current sample fell in the borderline range on the Brigance Screen II and in the low average range on the Bayley III. This discrepancy is consistent with recent research suggesting possible inflation of scores on the Bayley III among children born preterm (Moore et al., 2012; Vohr et al., 2012). This potential inflation in scores in the reference-standard measure may have resulted in a higher false-positive rate of the Brigance Screen II because children with true delays may have had a positive screen, but an inflated score on the Bayley III that did not indicate presence of delay. This, in turn, would yield lower specificity rates. Additionally, correlations among the Brigance Screen II and the Bayley III were higher for language than cognitive composites, as were the other screening performance statistics (e.g., specificity, PPV, NPV). This suggests that the Brigance Screen II is better at detecting language difficulties in comparison with cognitive and/or motor skill delays.

In contrast to the Brigance Screen II, the use of the CDR in this sample yielded much lower sensitivity, but adequate specificity. This is consistent with a previous study investigating its use in a community sample, in which the instrument had low sensitivity (0.50) and adequate specificity (0.86; Rydz et al., 2006). In the current sample, of the children with negative CDR screens, over half had a developmental delay based on Bayley III scores. This finding has significant implications for the use of the CDR as a screening measure and indicates that it may not adequately identify children who do indeed have a developmental delay.

Children followed in high-risk follow-up clinics are an already at-risk population and priority needs to be given to identifying all those in need of more intensive intervention services. Owing to the strong need to identify those who require additional services, the results of this study indicate that the Brigance Screen II, or other direct screening measures of child development that have high sensitivity may be a more accurate screening approach for use with children born preterm than the CDR and other similar measures that combine parent report and clinician observations. Screening tools involving direct assessment are time efficient and result in a high sensitivity rate. While some clinics may complete full assessments on *all* children, these results suggest that the use of a technically adequate, direct-skills assessment screening tool reduces the amount of children who need full testing and improve the accuracy of those who are referred for additional testing.

Limitations

This study has several limitations, including a small sample size. Unfortunately, owing to the small sample size, there was not sufficient power to complete a head-to-head comparison of the two screening approaches. Another major limitation is that the screening measures were not administered separately from one another or from the Bayley III for the majority of children. Therefore, examiners were not blinded to performance on the screening and reference-standard assessments, though none of the measures were scored until all were completed. Additionally, we did not collect information regarding identified central nervous system complications in the neonatal period, which may affect development, such as hypoxic-ischemic events, periventricular leukomalacia, intraventricular hemorrhage, or prolonged mechanical ventilation. One of the main limitations of the current study is the ability to generalize the findings to other children born preterm owing to the unique setting of a high-risk infant clinic. This sample may only attract a population that has particular characteristics found in follow-up clinics in which patients receive frequent monitoring and support. Regular screening and surveillance may occur more frequently in high-risk clinics, particularly those adopting a comprehensive care model (Nehra, Pici, Visintainer, & Kase, 2009), resulting in a higher proportion of the population being identified and referred for services. Additionally, research has demonstrated the positive associations with the use of these clinics, including improved parent knowledge about their child and increased attendance to follow-up appointments (Nehra et al., 2009).

Practical Implications

Despite the limitations with the present study, the results contribute to the growing knowledge about screening and surveillance, particularly among high-

risk populations. Results provide compelling evidence that clinician-administered screening assessments may be an efficient and effective approach to identifying children already at increased risk for delays who may need a more comprehensive developmental assessment. The potential impact of having an accurate screening measure for this population cannot be understated. Given that the population of children born preterm is well known to have difficulties that impact functioning in school and beyond (Chyi, Lee, Hintz, Gould, & Sutcliffe, 2008; Johnson et al., 2009), early and accurate identification of children who are most at-risk is critical to providing intervention that positively alters the trajectory of performance. The results of this study indicate that the time invested in use of a screening measure using direct assessment of skills, like the Brigance Screen II, is invaluable given that all children at-risk based on criterion performance were identified using this screener. An important direction for future research is a continued focus on evaluation of children born preterm in early childhood and beyond. More information regarding the validity of both screening and comprehensive developmental assessment tools is needed.

Acknowledgments

The authors wish to thank Ashlie Llorens and Erika Gonzalez for their contribution to this project. The authors also want to thank Tom Northrup, PhD, for reviewing earlier versions of this manuscript.

Funding

This work was supported by a University of Houston Small Grant Award.
Conflicts of interest: None declared.

References

Aarnoudse-Moens, C., Weisglas-Kuperus, N., van Goudoever, J. B., & Oosterlaan, J. (2009). Meta-analysis of neurobehavioral outcomes in very preterm and/or very low birth weight children. *Pediatrics*, *124*, 717–728. doi: 10.1542/peds.2008-2816

Aylward, G. P. (1995). *Bayley infant neurodevelopmental screener*. San Antonio, TX: Psychological Corporation.

Aylward, G. P. (2002). Cognitive and neuropsychological outcomes: More than IQ scores. *Mental Retardation and Developmental Disabilities Research Reviews*, *8*, 234–240. doi: 10.1002/mrdd.10043

Bayley, N. (2006). *Bayley scales of infant and toddler development* (3rd ed.). San Antonio, TX: The Psychological Corporation.

Bhutta, A. T., Cleves, M. A., Casey, P. H., Cradock, M. M., & Anand, K. J. S. (2002). Cognitive and behavioral outcomes of school-aged children who were born preterm: A meta-analysis. *Journal of the American Medical Association*, *288*, 728–737. doi:10.1001/jama.288.6.728

Bortolus, R., Parazzini, F., Trevisanuto, D., Cipriani, S., Ferrarese, P., & Zanardo, V. (2002). Developmental assessment of preterm and term children at 18 months: Reproducibility and validity of a postal questionnaire to parents. *Acta Paediatrica*, *91*, 1101–1107. doi: 10.1111/j.1651-2227.2002.tb00106.x

Chyi, L. J., Lee, H. C., Hintz, S. R., Gould, J. B., & Sutcliffe, T. L. (2008). School outcomes of late preterm infants: Special needs and challenges for infants born at 32 to 36 weeks gestation. *The Journal of Pediatrics*, *153*, 25–31. doi:10.1016/j.jpeds.2008.01.027

Committee on Fetus and Newborn (2004). Age terminology during the perinatal period. *Pediatrics*, *114*, 1362–1364. doi: 10.1542/peds.2004-1915

Council on Children with Disabilities, Section on Developmental Behavioral Pediatrics, Bright Futures Steering Committee & Medical Home Initiatives for Children with Special Needs Project Advisory Committee (2006). Identifying infants & young children with developmental disorders in the medical home: An algorithm for developmental surveillance & screening. *Pediatrics*, *118*, 405–420. doi: 10.1542/peds.2006-1231

Doyle, L. W. (2001). Outcome at 5 years of age of children 23 to 27 weeks' gestation: Refining the prognosis. *Pediatrics*, *108*, 134–141. doi:10.1542/peds.108.1.134

Glascoe, F. P. (2010). *Brigance Early Head Start Screen II*. North Billerica, MA: Curriculum Associates, Inc.

Gollenberg, A. L., Lynch, C. D., Jackson, L. W., McGuinness, B. M., & Msall, M. E. (2010). Concurrent validity of the parent-completed Ages and Stages Questionnaires, 2nd E., with the Bayley Scales of Infant Development II in a low-risk sample. *Child: Care, Health, and Development*, *36*, 485–490. doi: 10.1111/j.1365-2214.2009.01041.x

Hintz, S. R. (2005). Changes in neurodevelopmental outcomes at 18 to 22 months' corrected age among infants of less than 25 weeks' gestational age born in 1993-1999. *Pediatrics*, *115*, 1645–1651. doi:10.1542/peds.2004-2215

Ireton, H. R. (1992). *Child development inventory*. Minneapolis, MN: Behavior Science Systems, Inc.

Ireton, H. R., & Vader, H. (2004). *Child development review*. Minneapolis, MN: Behavior Science Systems, Inc.

Jenkins, J. R., Hudson, R. F., & Johnson, E. S. (2007). Screening for at-risk readers in a response to intervention framework. *School Psychology Review*, *36*, 582–600.

Johnson, S., Hennessy, E., Smith, R., Trikic, R., Wolke, D., & Marlow, N. (2009). Academic attainment and special educational needs in extremely preterm children at 11 years of age: the EPICure study. *Archives of Disease in Childhood Fetal and Neonatal Edition*, *94*, 283–289. doi: 10.1136/adc.2008.152793

Keller-Margulis, M. A., Abrahamson, C. W., Llorens, A. V., & Dempsey, A. G. (2013). Early intervention service utilization in children born preterm. *Clinical Practice in Pediatric Psychology*, *1*, 344–354. doi: 10.1037/cpp0000029

Moore, T., Johnson, S., Haider, S., Hennessy, E., & Marlow, N. (2012). Relationship between test scores using the second and third editions of the Bayley Scales in extremely preterm children. *The Journal of Pediatrics*, *160*, 553–558. doi:10.1016/j.jpeds.2011.09.047

Nehra, V., Pici, M., Visintainer, P., & Kase, J. S. (2009). Indicators of compliance for a developmental follow-up of infants discharged from a regional NICU. *Journal of Perinatal Medicine*, *37*, 677–681. doi: 10.1515/JPM.2009.135

Roberts, G., Howard, K., Spittle, A. J., Brown, N. C., Anderson, P. J., & Doyle, L. W. (2008). Rates of early intervention services in very preterm children with developmental disabilities at age 2 years. *Journal of Paediatrics and Child Health*, *44*, 276–280. doi: 10.1111/j.1440-1754.2007.01251.xk

Rydz, D., Shevell, M. I., Majnemer, A., & Oskoui, M. (2005). Topical review: Developmental screening. *Journal of Child Neurology*, *20*, 4–21. doi: 10.1177/08830738050200010201

Rydz, D., Srour, M., Oskoui, M., Marget, N., Shiller, M., Birnbaum, R., … Shevell, M. I. (2006). Screening for developmental delay in the setting of a community pediatric clinic: A prospective assessment of parent-report questionnaires. *Pediatrics*, *118*, 1178–1186. doi:10.1542/peds.2006-0466

Sachse, S., & Suchodoletz, W. V. (2008). Early identification of language delay by direct language assessment or parent report? *Journal of Developmental and Behavioral Pediatrics*, *29*, 34–41. doi: 10.1097/DBP.0b013e318146902a

Saigal, S., & Doyle, L. W. (2008). An overview of mortality and sequelae of preterm birth from infancy to adulthood. *The Lancet*, *371*, 261–269. doi:10.1016/S0140-6736(08)60136-1

Simard, M. N., Luu, T. M., & Gosselin, J. (2012). Concurrent validity of Ages and Stages Questionnaires in preterm infants. *Pediatrics*, *130*, 108–114. doi:10.1542/peds.2011-3532

Squires, J., Bricker, D., & Potter, L. (2009). *Ages and Stages Questionnaires User's Guide* (3rd ed.). Baltimore: Paul Brookes Publishing.

Vohr, B. R., Stephens, B. E., Higgins, R. D., Bann, C. M., Hintz, S. R., Epi, M., … Fuller, J; Eunice Kennedy Shriver National Institute of Child Health, Human Development Neonatal Research Network. (2012). Are outcomes of extremely preterm infants improving? Impact of Bayley assessment on outcomes. *The Journal of Pediatrics*, *161*, 222–228. doi:10.1016/j.jpeds.2012.01.057

Vohr, B. R., Wright, L. L., Poole, W. K., McDonald, S. A. , & for the NICHD Neonatal Research Network Follow-up Study (2005). Neurodevelopmental outcomes of extremely low birth weight infants <32 weeks' gestation between 1993 and 1998. *Pediatrics*, *116*, 635–643. doi: 10.1542/peds.2004-2247

Wilson, S. L., & Cradock, M. M. (2004). Review: Accounting for prematurity in developmental assessment and the use of age-adjusted scores. *Journal of Pediatric Psychology*, *29*, 641–649. doi: 10.1093/jpepsy/jsh067