Concurrent Validity of Ages and Stages Questionnaires in Preterm Infants

WHAT’S KNOWN ON THIS SUBJECT: Preterm children born between 29 and 36 gestational weeks are at higher risk of developmental delay. The Ages and Stages Questionnaires (ASQ) have been recommended as a developmental screening tool.

WHAT THIS STUDY ADDS: At 12 months’ corrected age (CA), the ASQ was insufficient in identifying delays on both mental and psychomotor scales of the Bayley Scales of Infant Development but was accurate in detecting mental delay at CA of 24 months.

abstract

BACKGROUND: Although preterm infants born at 29 to 36 gestational weeks (GW) are at risk for developmental delay, they do not always benefit from systematic follow-up. Primary care physicians are then responsible for their developmental surveillance and need effective screening tests. This study aimed to determine whether the Ages and Stages Questionnaires (ASQ) at 12 and 24 months’ corrected age (CA) identify developmental delay in preterm infants.

METHODS: With a cross-sectional design involving 2 observations at 12 and 24 months’ CA, 124 and 112 preterm infants were assessed. Infants were born between May 2004 and April 2006 at 29 to 36 GW. The ASQ and the Bayley Scales of Infant Development were used. Concurrent validity was calculated by using κ coefficient, sensitivity, and specificity.

RESULTS: At 12 months’ CA, the ASQ did not perform well in identifying infants with mental delay (κ = 0.08–0.19; sensitivity = 0.20–0.60; specificity = 0.68–0.88). Agreement (κ = 0.28–0.44) and specificity (0.90–0.97) were better for the psychomotor scale, but the sensitivity remained insufficient (0.25–0.52). At 24 months, the ASQ had good sensitivity (0.75–0.92) and specificity (0.55–0.78) for detecting mental delays (κ = 0.45). Results remained unsatisfactory for detecting motor delays (sensitivity = 0.31–0.50; specificity = 0.75–0.92).

CONCLUSIONS: Preterm infants with developmental delays at 12 months’ CA are not adequately identified with the ASQ. At 24 months’ CA, the ASQ identifies mental delays but not psychomotor delays. Additional measures should be used to increase yield of detecting at-risk preterm infants. Pediatrics 2012;130:e108–e114

AUTHORS: Marie-Noëlle Simard, PhD,a Thuy Mai Luu, MD, MSc,b and Julie Gosselin, PhDb,c

School of Psychology, Laval University, Quebec, Quebec, Canada; Pediatrics, Sainte-Justine University Health Centre, Montreal, Quebec, Canada; and School of Rehabilitation, University of Montreal, Montreal, Quebec, Canada

KEY WORDS
Ages and Stages Questionnaires, developmental delay, developmental screening, moderate and late preterm

ABBREVIATIONS
ASQ—Ages and Stages Questionnaires
BSID-II—Bayley Scales of Infant Development, second edition
CA—corrected age
DQ—developmental quotient
MDI—Mental Developmental Index
PDI—Psychomotor Development Index

www.pediatrics.org/cgi/doi/10.1542/peds.2011-3532
doi:10.1542/peds.2011-3532
Accepted for publication Mar 9, 2012
Address correspondence to Julie Gosselin, PhD, School of Rehabilitation, Faculty of Medicine, University of Montreal, CP 8128, Succ. Centre Ville, Montréal, Quebec, Canada H3C 3J7.
E-mail: julie.gosselin@umontreal.ca

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275). Copyright © 2012 by the American Academy of Pediatrics

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: Supported by the Canadian Institutes of Health Research grant 77617.
Infants born between 29 and 36 weeks' gestation represent a substantial and growing proportion of preterm births. Increased mortality and neonatal morbidities are observed among infants in this gestational age range when compared with full-term infants. In addition, they are more likely to display neurodevelopmental disorders including lower intellectual abilities, language and visual-motor deficits as well as behavioral, educational, and emotional problems during school-age years. Yet despite being at higher risk for developmental issues, these children rarely benefit from systematic follow-up and rely on their primary care physicians for developmental surveillance and screening. Therefore, valid and effective screening tools are of utmost importance to enhance identification of preterm infants with developmental problems who will benefit from early intervention to improve later educational and functional attainment.

In 2006, the American Academy of Pediatrics proposed an algorithm for developmental surveillance and screening for primary health care professionals. The Ages and Stages Questionnaires (ASQ) have been recommended as a comprehensive and valid screening instrument to identify children who require in-depth developmental assessment to determine their eligibility for rehabilitation services. So far, validity studies among preterm children have either been restricted to the most immature ones, recruited a small sample size, or used a reference test more or less applicable to North American standards.

The objective of the study was to determine the ability of the ASQ at 12 and 24 months' corrected age (CA) to identify preterm children at higher risk of presenting mild developmental delay that would justify additional developmental assessment. More specifically, concurrent validity was calculated by comparing the ASQ against the Bayley Scales of Infant Development, second edition (BSID-II).

### METHODS

#### Participants

Participants were initially recruited as part of a larger longitudinal study on early neurocranial markers in preterm infants that involved a 2-year follow-up. Between May 2004 and April 2006, 142 infants born between 29 and 36 6/7 weeks of gestation were randomly selected and included if they had a birth weight <2500 g and were admitted for at least 24 hours at Sainte-Justine University Health Centre’s NICU. Exclusion criteria were presence of chromosomal anomalies, congenital malformation, consanguinity, congenital infection, documented neonatal stroke, residing outside the metropolitan Montreal area, language spoken at home other than French or English, and significant social problems for 1 or both parents (drug addiction, alcoholism, mental illness, intellectual disability, or history of abuse, neglect, or family violence). During the course of the 2-year follow-up, 18 infants (13%) were not seen at 12 months’ CA. Of the 124 children seen at 12 months’ CA, 15 (12%) were lost to follow-up at 24 months’ CA (Fig 1). However, 3 additional subjects from the original cohort returned for assessment thus yielding 112 participants at 24 months’ CA. The institutional review board approved the protocol, and informed consent was obtained from each participant’s parent.

#### Procedures

At 12 and 24 months’ CA, the child developmental status was assessed at Sainte-Justine University Health Center over a 1-hour period by using the ASQ, second edition, and the BSID-II. The ASQ was first completed on site by the parents who were asked the questions by a research assistant. Then, the BSID-II was administered by 1 of 2 trained assessors who were blind to ASQ scores. In addition, independent assessors completed the assessments at the 12- and 24-month visits.

The ASQ is a parent-completed screening test that serves to identify children who need additional developmental assessment. It includes 19 questionnaires available for children aged 4 months to 60 months. The assessment time frame begins and ends 1 month before and after the questionnaire age, respectively. Each questionnaire contains 30 items covering 5 developmental domains: communication, problem-solving, executive function, motor development, and social and emotional development.

### FIGURE 1
Flow diagram of the study population.
personal-social, gross motor, and fine motor. Each item describes a developmental behavior/ability. Answer options are “yes,” “sometimes,” or “not yet” and are allocated 10, 5, or 0 points, respectively. A score is calculated for each domain. A child has “failed” or is “at risk” of developmental problems in a given domain if the score falls below a cutoff value set at 2 SD below the mean by using reference norms derived from a sample of 1643 US children. For the purpose of this study, we also examined failure cutoffs set at 1.5 SD and 1 SD below the mean still based on US norms. Test-retest and interrater reliabilities are high at 94%. Sensitivity ranges from 51% to 90% and specificity from 81% to 100% depending on age of assessment.

The BSID-II, a standardized measure of developmental performance for children aged 1 to 42 months, was selected as the reference test because of its wide application in large neonatal network studies on developmental outcomes in preterm infants. It allows calculation of 2 developmental quotients (DQ): the Mental Developmental Index (MDI) and the Psychomotor Development Index (PDI; mean = 100, SD = 15). A score <85 reflects mild developmental delay. All children were able to complete the BSID-II, and scores were obtained for each of them.

As part of this study, a distinct and original method was created to facilitate comparison between the screening test and the BSID-II DQ. First, we clustered the 5 ASQ domains into 2 broader categories. The communication, problem-solving, or personal-social domains was then compared against an MDI <85. Similarly, being at risk on any of the fine or gross motor domains was compared against a PDI <85. We used a cutoff of <85 on the BSID-II (1 SD below the mean) to include mild developmental delay, which also represents a level below which we feel infants benefit from developmental intervention.

Statistical Analyses

Statistical analyses were performed with SAS 9.1 (SAS Institute, Inc, Cary, NC). Results on the ASQ and BSID-II were cross-tabulated for each ASQ at-risk category, and Cohen’s was computed to determine chance independent agreement between classifications. Concurrent validity of the ASQ in identifying children with BSID-II scores <85 was further assessed by calculating sensitivity and specificity.

RESULTS

Characteristics of the study population at 12 and 24 months are detailed in Table 1. Children seen at 12 and 24 months’ CA were comparable to the children lost to follow-up in terms of biological and social characteristics (not shown).

Overall, children in our study obtained DQ within the normal ranges on the BSID-II. At 12 months’ CA, 121 infants completed both the ASQ and the MDI and 119, both the ASQ and the PDI. Mean MDI was 92 ± 9 with 18% (n = 22) scoring in the developmentally delayed range (<85), including 1 (0.8%) child scoring <70, whereas mean PDI was 86 ± 12, with 35% (n = 42) scoring 1 SD below the mean including 8% (n = 10) with scores <70. The proportions of at-risk infants failing any 1 of the ASQ communication, problem-solving skills, or personal-social domains were 36% (n = 44), 25% (n = 30), and 13% (n = 16) by using cutoffs of 1, 1.5, and 2 SD, respectively. On the ASQ gross or fine motor domains, 25% (n = 30), 17% (n = 20), and 11% (n = 13) were deemed at-risk with cutoffs of 1, 1.5, and 2 SD, respectively. ASQ at 12 months was not good in accurately detecting infants with MDI <85 on the BSID-II, with coefficients (κ) varying from 0.08 to 0.19 (Table 2). Sensitivity values were low, even with less-restrictive cutoffs (ie, 1 SD). Agreement improved with κ of 0.28 to 0.44 when comparing the ASQ with the PDI (Table 3). However, despite

<table>
<thead>
<tr>
<th>TABLE 1 Characteristics of the Study Population</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Seens at 12 mo CA</strong></td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td>Total, n</td>
</tr>
<tr>
<td>Median gestational age (range), wk</td>
</tr>
<tr>
<td>Gestational age, n (%)</td>
</tr>
<tr>
<td>29-31 (47) wk</td>
</tr>
<tr>
<td>32-33 (57) wk</td>
</tr>
<tr>
<td>34-36 (66) wk</td>
</tr>
<tr>
<td>Median birth weight (range), g</td>
</tr>
<tr>
<td>Birth weight, n (%)</td>
</tr>
<tr>
<td>1000–1499 g</td>
</tr>
<tr>
<td>1500–1999 g</td>
</tr>
<tr>
<td>2000–2499 g</td>
</tr>
<tr>
<td>Male gender, n (%)</td>
</tr>
<tr>
<td>66 (53)</td>
</tr>
<tr>
<td>Mean age at assessment (range)</td>
</tr>
<tr>
<td>Mean maternal age at birth (range)</td>
</tr>
<tr>
<td>Maternal education high school or less</td>
</tr>
<tr>
<td>Maternal ethnicity non-Canadian</td>
</tr>
<tr>
<td>Yearly family income &lt;50 000$ CAN</td>
</tr>
</tbody>
</table>
good specificity (90%–97%), sensitivity remained insufficient with values of 52% with a cutoff of 1 SD to as low as 25% with 2 SD, resulting in high rates of underreferred cases.

At 24 months’ CA, 109 infants had both ASQ and MDI scores available for analysis, and 107 had both the ASQ and the PDI. Mean MDI and PDI were 95 ± 15 and 95 ± 14, respectively, with 24% (n = 26) obtaining a score in the developmentally delayed range on the MDI (including 6% [n = 7] with scores <70) and 25% (n = 27) on the PDI (including 5% [n = 5] with scores <70). Failure on any 1 of the communication, problem-solving skills, and personal-social scales was observed in 55% (n = 60), 41% (n = 45), and 34% (n = 37) of infants by using cutoffs at 1, 1.5, and 2 SD, respectively. On the ASQ gross or fine motor domains, 33% (n = 35; 1 or 1.5 SD) or 14% (n = 15; 2 SD) of infants were considered at risk. Screening accuracy of the ASQ to identify infants with mental developmental delay (<85) was better than at 12 months’ CA (Table 4). The optimal cutoff was set at 1.5 SD (κ = 0.45) with a sensitivity of 88% and a specificity of 72%. On the other hand, concurrent validity at 24 months’ CA between the ASQ and the PDI was poor (Table 5).

**DISCUSSION**

This study highlights the limits of parent-completed questionnaires for developmental screening in high-risk preterm infants. Our goal was to determine whether the ASQ were sufficient on their own in detecting which child would benefit from additional evaluation and early intervention. Our results suggest that the ASQ had poor concurrent validity, albeit high specificities, on both MDI and PDI of the BSID-II at 12 months’ CA. In contrast, good agreement between the ASQ and the BSID-II MDI was achieved at 24 months, particularly if using an optimal stricter cutoff of 1.5 SD. However, the ASQ failed to accurately discriminate children with motor delays. On the basis of our results, an unacceptably high number of preterm children at risk for developmental problems as determined by formal evaluation with a developmental assessor would be missed if relying solely on a screening questionnaire.

Standards for good screening tests include sensitivity of at least 70% to 80% and specificity close to 80%.15,29 With these criteria, we found the ASQ to be valid for detecting mental developmental delay on the BSID-II at 24 months’ CA.
but not at 12 months’. Such findings may be at least partially explained by the fact that the target population is at higher risk of mild and moderate disabilities than of severe handicaps. These mild and moderate disabilities are known to be of late-emergent nature. With respect to motor delay, an unacceptably high number of infants were missed with the ASQ at both ages regardless of the cutoff threshold used. To our knowledge, 3 studies have assessed clinimetric characteristics of the ASQ in preterm populations aged between 12 and 24 months’ CA. Skellern et al calculated a sensitivity of 50% and a specificity of 91% for detecting a BSID-II MDI <85 at 18 months in infants born <31 weeks by using a cutoff of <2 SD on the ASQ.16 At 24 months, sensitivity and specificity were 75% and 62%, respectively, when compared against the Griffith Mental Developmental Scale. Another study of 169 infants with birth weight <1000 g found a sensitivity of 63% and a specificity of 75% for identification of developmental delay at 18 to 22 months defined as either a BSID-II MDI or PDI <85 also by using a cutoff of <2 SD on the ASQ.17 Finally, with the standard of 1 failed domain considered as a failed screen, Flamant et al obtained a sensitivity of 0.88 and a specificity of 0.57 for the identification of a global DQ <85 on the Brunet-Lezine Test with a cohort of children aged 2 years’ CA and born <35 weeks’ gestational age.18 Varying results may be partly explained by use of different definitions for either failed screening test or diagnosis of developmental delay, different reference tests, as well as different population characteristics.

The purpose of developmental screening is to correctly identify children who need referral for additional assessment or intervention, which may be determined by a DQ <85 on the BSID-II. By lowering a cutoff threshold on a screening test, we diminish the risk of missing a child who may have required more developmental services. This must be balanced against unnecessary and extensive referrals to overwhelmed developmental specialists. However, because the early years lay the foundation to school readiness and later success, the cost of over-referring may outweigh the financial and societal burden of long-term drop-outs.19,20 By using an optimal cutoff of 1.5 SD on the ASQ at 2 years’ CA, as determined by the κ coefficient, we found that only 3% of infants would be under-referred and 22% over-referred for mental delay, whereas 11% and 21% would be under- and over-referred for motor delay, respectively. Although the percentages of over-referrals appear high, it has been shown that children with a false-positive screening test may actually benefit from developmental intervention services because their performance on language, intelligence, and academic achievement measures was significantly lower than those who passed screening and did not have developmental problems after diagnostic assessment.21 Nevertheless, the high under-referral rate for detection of motor delays argues against the ASQ for this domain.

The ASQ has previously demonstrated good sensitivity and specificity in a representative US population.16 However, our study reveals that using developmental screening questionnaires in high-risk preterm children may not be as valid. One issue relates to administration of the questionnaires. Although the manual specifies that questionnaires should be preferably mailed to parents so they can actually attempt activities described in each item with their child, it also allows for completion of questionnaires on site. This method was also used by Rydz et al and is representative of primary care practice.11 Nevertheless, it is possible that parents in our study overestimated their child’s abilities because they did not try each item activity directly. A second concern that pertains to questionnaires is parents’ misinterpretation of some items. For example, a statement such as, “does your child run fairly well?” can be answered “yes” by parents when the child is, in fact, only walking fast. In addition, the nuance between “yes” and “sometimes” may be unclear. To a statement that asks whether the child kicks a ball by swinging his leg, a parent may choose “yes” when the child actually walks into the ball most of the time.

Few screening tests have been validated with a cohort of preterm infants aged 12 or 24 months’ CA. Pritchard et al used the Parents’ Evaluation of Developmental Status against the BSID-II with a cohort of 110 children aged 2 years’ CA and born with a birth weight <1250 g.31 By using a DQ <70 for their classification, they calculated a κ coefficient of 0.21, suggesting fair agreement between the Parents’ Evaluation of Developmental Status and the BSID-II. Sensitivity was 39%, whereas specificity was 84%. Two other screening tests have also been studied with cohorts of preterm infants between 12 and 24 months’ CA: the Parent Report of Children’s Abilities-Revised32 and the Cognitive Adaptive Test/Clinical Linguistic and Auditory Milestone Scale.33 However, both screening tests focused on detection of severe cognitive delay as determined by an MDI <70. Agreement between the Parent Report of Children’s Abilities—Revised and BDII MDI score was deemed to be poor (κ = 0.38) when tested on a cohort of 146 very preterm children aged 24 months’ CA. Sensitivity and specificity values were high at 83% and 72%, respectively. Finally, Vincer et al used the Cognitive Adaptive Test/Clinical Linguistic and Auditory Milestone Scale to predict severe cognitive delay at 18 months’ CA in 147 children born <31
weeks’ gestation and found optimal sensitivity of 88% and specificity of 97% by using different cutoff scores. Direct comparisons with our results should be performed with caution because those screening tests were used to detect an MDI < 70 as opposed to < 85. With only 6 items by section in the ASQ, mild mental delay is more difficult to identify accurately.

This study draws its strength from its appropriate sample size for calculation of acceptably precise sensitivity and specificity values. Camp advocated for a “minimum sample of 20 children who fall below the cutoff on the reference test” to minimize uncertainty (ie, 95% confidence interval) around calculated values. In addition, all infants received both screening and reference tests to prevent verification bias. Finally, grouping similar domains to create 2 distinct at-risk categories to increase precision of the screening test with the equivalent reference domain (ie, mental versus psychomotor development) can be useful in a clinical practice to determine more specifically what type of assessment is necessary for a given child. However, since this study was conducted, the third version of both the ASQ and the BSID has been released, thus limiting applicability of our results to current tests. The ASQ-3 has revised cutoff threshold based on new standardization, whereas the BSID-III has been renormed and modified to now include 3 scales (namely, cognitive, language, and motor) as well as new socioemotional and adaptive questionnaires.

**CONCLUSIONS**

Compared with children born at term age, children born between 29 and 36 weeks’ gestation are more at risk to suffer from neurodevelopmental difficulties such as intellectual disabilities, language, and visual-motor deficits, as well as behavioral, educational, and emotional problems during their childhood years. Valid screening tools are essential to primary health care providers to allow identification of children who could potentially present neurodevelopmental problems. According to the results of the current study, the ASQ failed in the identification of preterm children born at 29 to 36 6/7 weeks’ gestation with developmental delays at 12 months’ CA. The questionnaire was better at identifying children with an MDI < 85 at 24 months’ CA, but it performed poorly in discriminating children with a PDI < 85. We have shown in previous studies that a neurologic assessment completed at term age or concomitantly to the developmental assessment is a valid predictor of developmental performances at preschool age. Combining the use of ASQ and of neurologic assessment may lead to a better allocation of health care resources to the children who need it the most.

**REFERENCES**

17. Woodward BJ, Papile LA, Lowe JR, et al. Use of the Ages and Stages Questionnaire and...


Concurrent Validity of Ages and Stages Questionnaires in Preterm Infants
Marie-Noëlle Simard, Thuy Mai Luu and Julie Gosselin

Pediatrics 2012;130:e108; originally published online June 11, 2012;
DOI: 10.1542/peds.2011-3532

Updated Information & Services

including high resolution figures, can be found at:
/content/130/1/e108.full.html

References

This article cites 34 articles, 8 of which can be accessed free at:
/content/130/1/e108.full.html#ref-list-1

Citations

This article has been cited by 5 HighWire-hosted articles:
/content/130/1/e108.full.html#related-urls

Post-Publication Peer Reviews (P³Rs)

2 P³Rs have been posted to this article
/cgi/eletters/130/1/e108

Subspecialty Collections

This article, along with others on similar topics, appears in the following collection(s):
Developmental/Behavioral Pediatrics
/cgi/collection/development:behavioral_issues_sub

Permissions & Licensing

Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at:
/site/misc/Permissions.xhtml

Reprints

Information about ordering reprints can be found online:
/site/misc/reprints.xhtml