Predictive Value of the BSID-II and the Bayley-III for Early School Age Cognitive Function in Very Preterm Infants

Rachel S. Flynn, MD[1,2] iD **, Matthew D. Huber, MS[1],**
and Sara B. DeMauro, MD, MSCE[1,2]

Abstract

Objective: To compare the predictive validity of the Bayley Scales of Infant Development, Second Edition (BSID-II) and the Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III) for cognitive function at early school age in very preterm infants.
Methods: Seventy-seven former preterm infants (born <32 weeks gestation and ≤2000 g) completed both the BSID-II and the Bayley-III at 2 years corrected age. Children enrolled at hospitals that perform follow-up beyond 2 years had cognitive assessments with the Wechsler Preschool and Primary Scale of Intelligence Fourth Edition (WPPSI-IV). Associations between Bayley and WPPSI scores were assessed using correlation coefficients, linear regression, and Bland-Altman plots.
Results: Thirty-one of 45 eligible children were tested with the WPPSI-IV at 47 ± 11 months. Average BSID-II Mental Development Index (MDI) was 86 ± 19, Bayley-III Cognitive composite score was 101 ± 12 and WPPSI Full Scale IQ (FSIQ) was 96 ± 12. Correlation between MDI and FSIQ was 0.54 ($P<.001$); correlation between Bayley-III cognitive composite score and FSIQ was 0.31 ($P=.03$). Bayley-III language composite had a modestly stronger correlation with FSIQ than cognitive composite (correlation coefficient 0.39; $P=.005$). Linear regression models also demonstrated that BSID-II was more closely correlated with FSIQ than Bayley-III. This bias was consistent across the full range of scores.
Conclusion: The BSID-II underestimated FSIQ and the Bayley-III overestimated FSIQ. Children at risk for impairment might be missed with the Bayley-III. As the Bayley-4 is introduced, clinicians and researchers should be cautious about interpretation of scores until performance of this new measure is fully understood.

Keywords
developmental assessment, prematurity, Bayley scales of infant development, Wechsler preschool and primary scale of intelligence

Received April 22, 2020. Received revised September 16, 2020. Accepted for publication October 12, 2020.

Introduction

Improved survival of very preterm infants (≤32 weeks gestation) due to advances in perinatal care has been associated with increased interest in neurodevelopmental outcomes.[1] Neurodevelopmental follow up during the first few years of life serves to identify children with developmental delay who may benefit from specialized interventions, provides important data for counseling parents, and informs neonatal care. Continued follow up of high risk infants at preschool and school age is critical because some adverse sequelae of prematurity only

become fully apparent when the child begins to develop more complex functioning. Subtle impairments can become more obvious as a child approaches school age and cognitive and social demands increase.[2]

[1]Children's Hospital of Philadelphia, Philadelphia, PA, USA
[2]Hospital of the University of Pennsylvania, Philadelphia, PA, USA

Corresponding Author:
Sara B. DeMauro, MD, MSCE, Department of Neonatology, Children's Hospital of Philadelphia, 3401 Civic Center Blvd, Philadelphia, PA 19104, USA.
Email: demauro@email.chop.edu

The Bayley Scales of Infant Development (BSID) is the most widely reported tool for developmental assessment in the first 2 years of life.[3,4] The BSID was initially published in 1969, updated to the BSID-II in 1993, and again to the Bayley-III in 2006.[5,6] The Bayley-III was expected to provide more refined information than prior versions by separating the original mental developmental index (MDI) into cognitive and language scales. The standardization sample of the Bayley-III included a heterogenous group of children: 10% had high risk for developmental delay.[6] Prior studies have shown that at 18 to 22 months corrected age, Bayley-III scores are significantly higher than BSID-II scores in all domains.[7-9] It is not known whether the Bayley-III underestimates developmental delay, the BSID-II norms are too conservative, or both.[7] Ideally, early developmental assessments such as the Bayley would be helpful for predicting later cognitive outcomes, however studies about the prognostic value of the Bayley are conflicting.[1,4,10]

The aim of the current study is to assess and compare the predictive validity of the BSID-II and the Bayley-III for cognitive function at early school age as measured with the Wechsler Preschool and Primary Scale of Intelligence, Fourth Edition (WPPSI-IV) in very preterm infants. Prior studies comparing the BSID-II and Bayley-III had important methodological limitations such as estimating BSID-II performance based on item overlap.[8,11,12] This study is the first to administer the BSID-II, Bayley-III, and WPPSI-IV in the same cohort of preterm-born children.[13]

Methods

We previously reported a multicenter randomized crossover study comparing the BSID-II and Bayley-III in very preterm infants.[7] The current study is a follow-up of children enrolled in the primary study.

Participants

Details about recruitment, study eligibility, and the study procedures were previously published.[7] In brief, study participants were born <32 weeks gestation and/or weighing ≤2000 g at birth if no gestational age was available, and were 18 to 22 months corrected age (CA) between July 2011 and July 2014. Seventy-seven participants completed both a BSID-II study visit and a Bayley-III clinical visit 4-8 weeks apart, in random order, which was determined in advance using a computer-based random number generator and incorporating variable block sizes. Two of the four clinical sites that enrolled children in the original study routinely continue follow-up beyond 2 years, using the WPPSI-IV[6,13] for cognitive assessments. Children enrolled at those sites were eligible for the current follow-up study.

Outcomes

The primary outcome was the correlation between the WPPSI-IV full scale IQ (FSIQ) and cognitive development as measured by the BSID-II and the Bayley-III. All of these assessments were administered by clinical psychologists. The Bayley-III and WPPSI were administered clinically and the BSID-II was administered for research purposes only. If multiple scores were available, the WPPSI administered closest to school age was used. Bayley scores were adjusted for prematurity; WPPSI scores were unadjusted.

The presence of hearing loss was defined as documented hearing loss on the most recent audiogram. Blindness was defined as no vision in both eyes. Severe neurologic abnormality was defined as cerebral palsy (CP), hypertonia, hypotonia or shunt dependent hydrocephalus by 18-22 months CA. Hypertonia and hypotonia were included in the category of major neurologic abnormality, as these conditions are considered by some to represent variants of CP.[14]

Analysis

Baseline characteristics and outcomes of the children who were included in the current analyses because WPPSI data were available were compared to the rest of the study population using χ^2 tests, Fisher's exact tests with small cell sizes, or one-way ANOVA for continuous variables. The associations between Bayley scores and WPPSI scores were assessed using simple linear regression models, Bland-Altman plots, and Kendall rank correlation coefficients. Statistical analyses were conducted using R, Version 3.6.1 (R Core Team, 2019), and 2-sided $P < .05$ was considered significant.

Ethical Approval and Informed Consent

The study was approved by the Institutional Review Board at The Children's Hospital of Philadelphia (IRB 10-007937).

Results

Thirty-one of the 45 (69%) infants enrolled in the primary study at hospitals that routinely perform follow-up beyond 2 years were assessed with the WPPSI-IV at a clinical visit at 47 months (± 11.4 months). The cohort was 51% male and there was no statistically significant difference in scores by gender. The mean GA at birth was 28 ± 2 weeks.

The perinatal and demographic characteristics of the children who did and did not complete the WPPSI-IV are compared in Table 1. Compared with children who

Table 1. Characteristics of Participants.

	All participants (n = 77)	Participants WPPSI performed (n = 31)	Participants WPPSI not performed (n = 46)	P value
Maternal demographic data				
Age, y, mean	29.2 ± 6.4	26.5 ± 6.1	31.0 ± 5.9	.002
Married status*, n (%)				.144
Single	32 (41.6)	17 (54.8)	15 (32.6)	
Married	35 (45.5)	12 (38.7)	23 (50.0)	
Maternal race or ethnicity*, n (%)				.313
Black	41 (53.3)	18 (58.1)	23 (50.0)	
White	30 (39.0)	11 (35.5)	19 (41.3)	
Asian	1 (1.3)	0 (0)	1 (2.2)	
Hispanic	2 (2.6)	2 (6.5)	0 (0)	
Maternal education*, n (%)				.157
< High school	19 (24.7)	11 (35.5)	8 (17.4)	
2 years college	9 (11.7)	2 (6.5)	7 (15.2)	
4 years college	18 (23.4)	8 (25.8)	10 (21.7)	
Graduate level	10 (13.0)	2 (6.5)	8 (17.4)	
Infant birth data				
Gender				.250
Male	39 (50.7)	13 (41.9)	26 (56.5)	
Female	38 (49.4)	18 (58.1)	20 (43.5)	
Gestational age, wk	28.1 ± 2.5	28.4 ± 2.5	28.0 ± 2.4	.464
Birthweight, g	1164 ± 384	1132 ± 357	1170 ± 395	.554
Multiple gestation, n (%)	29 (37.7)	14 (45.2)	15 (32.6)	.339
Small for gestational age, n (%)	2 (2.60)	1 (3.23)	1 (2.17)	1
Infant race or ethnicity, n (%)				.786
Black	41 (53.3)	18 (58.1)	23 (50.0)	
White	35 (45.5)	13 (41.9)	22 (47.8)	
Hispanic	0 (0)	0 (0)	0 (0)	
Asian	1 (1.3)	0 (0)	1 (2.2)	
Neonatal risk factors				
Bronchopulmonary dysplasia*, n (%)	8 (10.4)	6 (19.4)	2 (4.4)	.059
Intraventricular hemorrhage grade 3 or 4*, n (%)	3 (3.9)	1 (3.2)	2 (4.4)	1
Necrotizing enterocolitis*, n (%)	5 (6.5)	2 (6.5)	3 (6.5)	1
Patent ductus arteriosus ligation*, n (%)	10 (13.0)	4 (12.9)	6 (13.0)	1
Retinopathy of prematurity requiring laser treatment,* n (%)	3 (3.9)	2 (6.5)	1 (2.2)	.563
Sepsis*, n (%)	13 (16.9)	6 (19.4)	7 (15.2)	.762
Length of stay*, d	72.1 ± 68.9	65.1 ± 87.0	77.1 ± 53.2	.467
Postnatal steroids*, n (%)	9 (11.7)	4 (12.9)	5 (10.9)	1
Mechanical ventilation*, n (%)	49 (63.6)	22 (71.0)	27 (58.7)	.837
Later risk factors				
Hearing loss*, n (%)	6 (7.8)	2 (6.5)	4 (8.7)	.681
Blindness, n (%)	0 (0)	0 (0)	0 (0)	1
Major neurologic abnormality*, n (%)	7 (9.1)	1 (3.2)	6 (13.0)	.227
Autism*, n (%)	3 (3.9)	2 (6.5)	1 (2.2)	.602
Receiving EI*, n (%)	52 (67.5)	30 (96.8)	22 (47.8)	1
Study characteristics				
Corrected age in months at BSID-II	19.2 ± 2.9	19.6 ± 1.8	18.9 ± 3.5	.300
Corrected age in months at Bayley-III*	18.7 ± 2.4	19.4 ± 1.7	18.3 ± 2.7	.040
Age in months at WPPSI	47.4 ± 11.4	47.4 ± 11.4	N/A	N/A

Data are presented as mean ± SD or No. of %.
*Data analyzed included those for whom the outcome was known. Number missing for married status (10), maternal race (3), maternal education (21), BPD (2), IVH (2), NEC (4), PDA (2), ROP (1), sepsis (2), length of stay (2), postnatal steroids (9), mechanical ventilation (5), hearing loss (9), major neurologic abnormality (4), autism (12), received EI (23), age in months at Bayley-III (2).

Table 2. Summary Statistics.

	BSID-II MDI	Bayley-III cognitive composite	Bayley-III language composite	WPPSI-IV full scale IQ
Mean (SD)	86 (±19)	101 (±12)	94 (±22)	96 (±12)
Median (IQ range)	88 (17)	100 (15)	94 (20)	96 (14)

Table 3. Correlations Between BSID-II and Bayley-III Scores and WPPSI-IV Scales.

	BSID-II MDI	Bayley-III cognitive composite	Bayley-III language composite
Full scale IQ (FSIQ)	0.54, $P < .001$	0.31, $P = .030$	0.39, $P = .005$
Verbal comprehension (VCI)	0.47, $P < .001$	0.22, $P = .129$	0.26, $P = .065$
Visual spatial (VSI)	0.23, $P = .099$	0.15, $P = .291$	0.01, $P = .952$
Working memory (WMI)	0.33, $P = .016$	0, $P = 1$	0.34, $P = .015$

did not complete the WPPSI, children who completed the WPPSI had a younger mean maternal age and were slightly older at the time of Bayley-III testing. When separating out only those children enrolled at hospitals that performed follow up beyond 2 years and therefore would have been eligible for the current study, there was no significant difference in infant birth data, neonatal risk factors, later risk factors, or study characteristics in children who did and did not complete the WPPSI (Supplementary Table).

BSID-II, Bayley-III, and WPPSI-IV summary statistics are listed in Table 2. Results from our previous study demonstrated that the order of the Bayley test administration did not influence performance and there was no evidence for confounding by a learning effect.[7] Correlations between BSID-II and Bayley-III subtests and the WPPSI-IV scales are displayed in Table 3. The BSID-II MDI was more strongly correlated with the WPPSI-IV FSIQ (correlation 0.54, $P = <.001$) than the Bayley-III cognitive composite score (correlation 0.31, $P = .030$). Interestingly, the Bayley-III language composite had a modestly stronger correlation with the WPPSI-IV FSIQ (correlation 0.39, $P = .005$) than the Bayley-III cognitive composite score. When compared to the WPPSI-IV primary index scales, the BSID-II MDI had a moderate correlation with verbal comprehension (VCI) (correlation 0.47, $P < .001$) and a weak correlation with working memory (WMI) (correlation 0.33, $P = .016$). The Bayley-III language composite had a weak correlation with working memory (WMI) (0.34, $P = .015$). There was no correlation between BSID-II and Bayley-III scores and WPPSI-IV visual spatial (VSI) scores.

Figure 1 presents correlations between WPPSI-IV FSIQ and BSID-II MDI Scores (1a) and Bayley-III cognitive composite scores (1b). BSID-II MDI scores at 18

to 22 months CA were significantly correlated with WPPSI-IV FSIQ scores (linear regression beta coefficient 0.60, 95% CI 0.31-0.88, $P = .0002$). Bayley-III cognitive composite scores were not as strongly correlated with WPPSI-IV FSIQ (linear regression beta coefficient 0.50, 95% CI 0.05-0.97, $P = .03$).

Bland-Altman plots demonstrate agreement between BSID-II MDI scores (Figure 1c) and Bayley-III cognitive composite scores (Figure 1d) and WPPSI FSIQ scores. The mean difference, or bias, between the WPPSI-IV FSIQ and the BSID-II MDI was 7.57. The mean difference between the WPPSI-IV FSIQ and the Bayley-III cognitive composite score was −2.23. In other words, FSIQ tends to be higher than predicted based on the BSID-II but somewhat lower than predicted based on the Bayley-III.

Discussion

In our cohort of very preterm-born children, BSID-II MDI scores were more closely correlated with cognitive scores at early school age than Bayley-III composite scores. BSID-II underestimated FSIQ and Bayley-III overestimated FSIQ across the full range of scores.

Prior studies evaluating the predictive value of the BSID-II are inconsistent, with some concluding the test was predictive of cognitive functioning at school age,[15-18] and others concluding it was not.[1,14,19-21] Similarly, findings are mixed in the handful of studies evaluating the predictive value of the Bayley-III. Bode et al found that in contrast to prior editions of the BSID, the Bayley-III is a good predictor of cognitive function at 4 years as evaluated by the WPPSI-III.[10] Our study found weaker correlations between Bayley-III cognitive composite scores and WPPSI-IV FSIQ than Bode et al. Similar to the current research, other subsequent
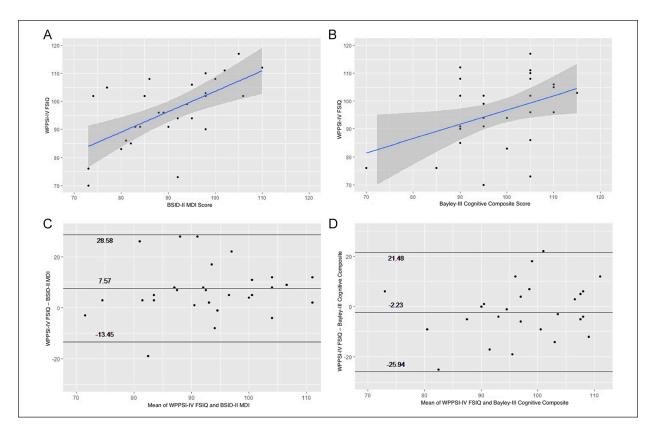
Figure 1. Scatterplots and Bland-Altman plots comparing the BSID-II MDI and Bayley-III cognitive composite with WPPSI-IV FSIQ. (A) Scatterplot of the BSID-II MDI against the WPPSI-IV FSIQ. Regression coefficient is .6008 (95% CI .31-.88, *P* = .0002). (B) Scatterplot of the Bayley-III cognitive composite score against the WPPSI-IV FSIQ. Regression coefficient is 0.5120 (95% CI .05-.97, *P* = .03). (C) Bland-Altman plot of BSID-II MDI score and WPPSI-IV FSIQ. (D) Bland-Altman plot of Bayley-III cognitive composite score and WPPSI-IV FSIQ.

studies found that the Bayley-III was not strongly predictive of future impairments.[4,22] Our work builds on this prior research by demonstrating that the older BSID-II may actually be more predictive of later cognitive function than the Bayley-III. This highlights the importance of rigorous comparisons of different versions of such assessments.

Previous research suggests that the Bayley-III underestimates impairment, leading to decreased identification of children at risk who may benefit from intervention and therapies.[7,22] We found that the FSIQ tends to be lower than predicted based on Bayley-III cognitive composite and higher than predicted based on the BSID-II MDI. These findings suggest that based on Bayley-III scores, children at risk for future impairment might be missed.

Among Bayley-III subtests, the Bayley-III language composites correlated more strongly with WPPSI-IV FSIQ scores than the cognitive composite. Language scores are often not included in reports of developmental outcomes, and are inconsistently included as a component of composite endpoints in research studies.[23] Our research suggests that Bayley-III language composite scores are correlated with later outcomes and therefore should be analyzed and reported.

The BSID-II MDI was also more strongly correlated with WPPSI-IV primary index scales than either the Bayley-III cognitive composite or the language composite. While the BSID-II MDI was moderately correlated with WPPSI-IV verbal comprehension scores (VCI), the Bayley-III language composite was not predictive of WPPSI-IV VCI. Both the BSID-II MDI and the Bayley-III language composite scores were weakly correlated with WMI. Bayley scores were not predictive of VSI, perhaps because this index has a motor component, which may not be captured with cognitive scores. This suggests that the older BSID-II may be more predictive of both global intellectual functioning, as well as more specific domains of cognitive functioning.

There are several limitations to our study. Our sample size is small and not all children included in the

primary research study had later clinical follow-up with the WPPSI. Since this component of the study was added on later, we did not contact families whose child did not complete WPPSI testing, which may have contributed to the attrition rate. Nevertheless, participants served as their own controls, which provided sufficient power to draw important conclusions despite the sample size. Additionally, the WPPSI-IV was administered at a wide range of ages. The Bayley is a developmental assessment and not an intelligence test; as such, it is not meant to predict future cognitive function. However, it continues to be used as the key measure of development in high risk populations, with an expectation that it will correlate with later cognitive measures. If new versions of the Bayley correlate less strongly with IQ and even overestimate children's skills, we may lose an opportunity to identify and help high risk children and underestimate the true incidence of delay.

With the introduction of the new Bayley-4 in Fall 2019, we encourage clinicians and researchers to be cautious about interpretation of the scores until the performance of this new measure is fully understood. While the Bayley-4 technical manual reports that Bayley-4 and WPPSI FSIQ scores are similar when both are administered at a mean age of 37 months, it remains unknown if Bayley-4 scores at 2 years correlate with later early school age IQ.[24] Future prospective studies are needed to investigate the associations between the new Bayley version and longer-term outcomes.

Author Contributions

All authors made a substantial contribution to the concept or design of the work, acquisition, analysis or interpretation of data, drafted the article or revised it critically for important intellectual content, and approved the version to be published.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Rachel S. Flynn   https://orcid.org/0000-0002-3320-6991

Supplemental Material

Supplemental material for this article is available online.

References

1. Santos dos ESL, de Kieviet JF, Königs M, van Elburg RM, Oosterlaan J. Predictive value of the Bayley scales of infant development on development of very preterm/very low birth weight children: a meta-analysis. *Early Hum Dev*. 2013;89:487-496. doi:10.1016/j.earlhumdev.2013.03.008

2. Aylward GP. Cognitive and neuropsychological outcomes: more than IQ scores. *Ment Retard Dev Disabil Res Rev*. 2002;8:234-240. doi:10.1002/mrdd.10043

3. Anderson PJ, de Luca CR, Hutchinson E, Roberts G, Doyle LW. Underestimation of developmental delay by the new Bayley-III Scale. *Arch Pediatr Adolesc Med*. 2010;164:352-356. doi:10.1001/archpediatrics.2010.20

4. Månsson J, Stjernqvist K, Serenius F, Ådén U, Källén K. Agreement between Bayley-III measurements and WISC-IV measurements in typically developing children. *J Psychoeduc Assess*. 2018;309:073428291878143. doi:10.1177/0734282918781431

5. Bayley N. *Bayley Scales of Infant Development*. 2nd ed. Psychological Corporation; 1993.

6. Bayley N. *Bayley Scales of Infant and Toddler Development*. 3rd ed. Harcort Assessment, Inc; 2006.

7. Sharp M, DeMauro SB. Counterbalanced comparison of the BSID-II and Bayley-III at eighteen to twenty-two months corrected age. *J Dev Behav Pediatr*. 2017;38:322-329. doi:10.1097/DBP.0000000000000441

8. Vohr BR, Stephens BE, Higgins RD, et al. Are outcomes of extremely preterm infants improving? Impact of Bayley assessment on outcomes. *J Pediatr*. 2012;161:222-228. doi:10.1016/j.jpeds.2012.01.057

9. Gasparini C, Caravale B, Rea M, et al. Neurodevelopmental outcome of Italian preterm children at 1year of corrected age by Bayley-III scales: an assessment using local norms. *Early Hum Dev*. 2017;113:1-6. doi:10.1016/j.earlhumdev.2017.06.007

10. Bode MM, D'Eugenio DB, Mettelman BB, Gross SJ. Predictive validity of the Bayley, third edition at 2 years for intelligence quotient at 4 years in preterm infants. *J Dev Behav Pediatr*. 2014;35:570-575. doi:10.1097/DBP.0000000000000110

11. Moore T, Johnson S, Haider S, Hennessy E, Marlow N. Relationship between test scores using the second and third editions of the Bayley Scales in extremely preterm children. *J Pediatr*. 2012;160:553-558. doi:10.1016/j.jpeds.2011.09.047

12. Reuner G, Fields AC, Wittke A, Löpprich M, Pietz J. Comparison of the developmental tests Bayley-III and Bayley-II in 7-month-old infants born preterm. *Eur J Pediatr*. 2013;172:393-400. doi:10.1007/s00431-012-1902-6

13. Wechsler D. *Wechsler Preschool and Primary Scale of Intelligence*. 4th ed. The Psychological Corporation; 2012.

14. Hack M, Taylor HG, Drotar D, et al. Poor predictive validity of the Bayley Scales of Infant Development for cognitive function of extremely low birth weight children

at school age. *Pediatrics*. 2005;116:333-341. doi:10.1542/peds.2005-0173

15. Dezoete JA, MacArthur BA, Tuck B. Prediction of Bayley and Stanford-Binet scores with a group of very low birthweight children. *Child Care Health Dev*. 2003;29:367-372.

16. Munck P, Niemi P, Lapinleimu H, Lehtonen L, Haataja L, PIPARI Study Group. Stability of cognitive outcome from 2 to 5 years of age in very low birth weight children. *Pediatrics*. 2012;129:503-508. doi:10.1542/peds.2011-1566

17. Sajaniemi N, Hakamies-Blomqvist L, Katainen S, Wendt von L. Early cognitive and behavioral predictors of later performance: a follow-up study of ELBW children from ages 2 to 4. *Early Child Res Q*. 2001;16:343-361. doi:10.1016/s0885-2006(01)00107-7

18. van Vliet EOG, de Kieviet JF, van der Voorn JP, Been JV, Oosterlaan J, van Elburg RM. Placental pathology and long-term neurodevelopment of very preterm infants. *Am J Obstet Gynecol*. 2012;206:489.e1-e7. doi:10.1016/j.ajog.2012.03.024

19. O'Shea TM, Joseph RM, Allred EN, et al. Accuracy of the Bayley-II mental development index at 2 years as a predictor of cognitive impairment at school age among children born extremely preterm. *J Perinatol*. 2018;38:908-916. doi:10.1038/s41372-017-0020-8

20. Potharst ES, Houtzager BA, van Sonderen L, et al. Prediction of cognitive abilities at the age of 5 years using developmental follow-up assessments at the age of 2 and 3 years in very preterm children. *Dev Med Child Neurol*. 2012;54:240-246. doi:10.1111/j.1469-8749.2011.04181.x

21. Roberts G, Anderson PJ, Doyle LW, Group TVICS. The stability of the diagnosis of developmental disability between ages 2 and 8 in a geographic cohort of very preterm children born in 1997. *Arch Dis Child*. 2010;95:786-790. doi:10.1136/adc.2009.160283

22. Spencer-Smith MM, Spittle AJ, Lee KJ, Doyle LW, Anderson PJ. Bayley-III cognitive and language scales in preterm children. *Pediatrics*. 2015;135:e1258-e1265. doi:10.1542/peds.2014-3039

23. Haslam MD, Lisonkova S, Creighton D, et al. severe neurodevelopmental impairment in neonates born preterm: impact of varying definitions in a Canadian cohort. *J Pediatr*. 2018;197:75–81.e4. doi:10.1016/j.jpeds.2017.12.020

24. Bayley N, Aylward GP. *Bayley Scales of Infant and Toddler Development*. 4th ed. NCS Pearson; 2019.