

Bayley-III Cognitive and Language Scales in Preterm Children

Megan M. Spencer-Smith, PhD^{a,b}, Alicia J. Spittle, PhD^{b,c,d}, Katherine J. Lee, PhD^{e,f}, Lex W. Doyle, MD^{b,g,h}, Peter J. Anderson, PhD^{b,f}

abstract

BACKGROUND: This study aimed to assess the sensitivity and specificity of the *Bayley Scales of Infant and Toddler Development, Third Edition* (Bayley-III), Cognitive and Language scales at 24 months for predicting cognitive impairments in preterm children at 4 years.

METHODS: Children born <30 weeks' gestation completed the Bayley-III at 24 months and the *Differential Ability Scale, Second Edition* (DAS-II), at 4 years to assess cognitive functioning. Test norms and local term-born reference data were used to classify delay on the Bayley-III Cognitive and Language scales. Impairment on the DAS-II Global Conceptual Ability, Verbal, and Nonverbal Reasoning indices was classified relative to test norms. Scores < -1 SD relative to the mean were classified as mild/moderate delay or impairment, and scores < -2 SDs were classified as moderate delay or impairment.

RESULTS: A total of 105 children completed the Bayley-III and DAS-II. The sensitivity of mild/moderate cognitive delay on the Bayley-III for predicting impairment on DAS-II indices ranged from 29.4% to 38.5% and specificity ranged from 92.3% to 95.5%. The sensitivity of mild/moderate language delay on the Bayley-III for predicting impairment on DAS-II indices ranged from 40% to 46.7% and specificity ranged from 81.1% to 85.7%. The use of local reference data at 24 months to classify delay increased sensitivity but reduced specificity. Receiver operating curve analysis identified optimum cut-point scores for the Bayley-III that were more consistent with using local reference data than Bayley-III normative data.

CONCLUSIONS: In our cohort of very preterm children, delay on the Bayley-III Cognitive and Language scales was not strongly predictive of future impairments. More children destined for later cognitive impairment were identified by using cut-points based on local reference data than Bayley-III norms.

WHAT'S KNOWN ON THIS SUBJECT: There is concern that the Bayley-III overestimates developmental functioning in preterm populations. The ability of the Bayley-III Cognitive and Language scales to predict later functioning in very preterm children has not been examined.

WHAT THIS STUDY ADDS: The norms on the Bayley-III Cognitive and Language scales at 24 months had low sensitivity for impairment across general cognitive, verbal and nonverbal reasoning domains at 4 years, which was better detected using cut-points based on local term-born reference data.

^aSchool of Psychological Sciences, Monash University, Melbourne, Australia; ^bClinical Sciences, ^cClinical Epidemiology and Biostatistics Unit, Murdoch Childrens Research Institute, Melbourne, Australia; ^dDepartment of Physiotherapy, ^eDepartment of Paediatrics, ^fDepartment of Obstetrics and Gynaecology, The University of Melbourne, Melbourne, Australia; ^gNewborn Research, and ^hResearch Office, The Royal Women's Hospital, Melbourne, Australia

Dr Spencer-Smith conceptualized and designed the study, collected data, and drafted the initial manuscript; Dr Spittle assisted in the study design and conceptualizing of the analysis plan, collected data, and reviewed and revised the manuscript; Dr Lee assisted in the study design and conceptualizing of the analysis plan, carried out the statistical analyses, and reviewed and revised the manuscript; Drs Doyle and Anderson assisted in the study design and conceptualizing of the analysis plan and reviewed and revised the manuscript; and all authors approved the final manuscript as submitted.

This trial has been registered with the Australian New Zealand Clinical Trials Registry (identifier ACTRN12606000252516).

www.pediatrics.org/cgi/doi/10.1542/peds.2014-3039

DOI: 10.1542/peds.2014-3039

Accepted for publication Feb 26, 2015

For decades the Bayley Scales of Infant Development (BSID) have been used for the early identification and quantification of developmental delay and to determine eligibility for early intervention services in infants. One limitation of the first and second editions of the Bayley scales (BSID and BSID-II)^{1,2} is that they provided only 2 broad developmental indices: the Mental Development Index (MDI), which evaluated early cognitive and language development, and the Psychomotor Development Index, which evaluated early fine and gross motor development. These broad indices lacked the capacity to differentiate specific delays in cognitive and language development or in fine and gross motor development, which are important for determining appropriate intervention services. A systematic review reported that the ability of the MDI to predict later functioning in preterm populations was variable, indicating a moderate correlation with general cognitive function in preschool-/ school-aged children (meta-analysis of 14 studies: $r = 0.61$; range: 0.39–0.72) and inconsistent associations with language function in preschool-aged children (3 studies; range: 0.23–0.68).³

The most recent edition, the *Bayley Scales of Infant and Toddler Development, Third Edition* (Bayley-III),⁴ has been restructured to include index scores for cognitive, language, and motor domains. The standardization of the 3 editions of the Bayley scales was conducted in a US population and for the Bayley-III included 10% with developmental problems reflecting the general population.⁴ It was hoped that changes in the Bayley-III test structure and restandardization would improve its capacity to identify specific developmental problems in high-risk infants, such as those born preterm. Unfortunately, initial reports suggest that the Bayley-III underestimates rates of developmental delay in preterm and

term-born infants across cognitive, language, and motor domains.^{5–10} In light of these reports, it is important to assess the Bayley-III's capacity to predict later cognitive, language, and motor impairments.

In children born <30 weeks' gestational age (GA) the Bayley-III Motor scale at 24 months underestimates motor impairment at 4 years,⁸ and the Expressive and Receptive Language subscales at 3 years underestimate the prevalence of language impairment at 5 years.¹¹ In the current study we aimed to evaluate the sensitivity and specificity of the Bayley-III Cognitive and Language scales at 24 months for predicting impairments in general cognitive, verbal, and nonverbal functioning at 4 years in children born very preterm (VP).

METHODS

Participants

The VP cohort was recruited for a randomized controlled trial (RCT) of a home-based preventive care program.^{12–14} Infants born <30 weeks' GA were recruited from the Royal Women's and the Royal Children's hospitals in Melbourne, Australia, from January 2005 to January 2007. Exclusion criteria included having a congenital brain anomaly, family not living within a 100-kilometer radius of the hospital, or non-English speaking family. Participants were assessed at 24 months' corrected age (CA) when they completed the Bayley-III and at 4 years' CA when they completed the *Differential Ability Scale, Second Edition* (DAS-II).¹⁵ There was little evidence of difference in cognitive performance between intervention and control groups at the 24-month or 4-year follow-ups^{12,13}; therefore, the data for the 2 groups were combined in the current study.

A local reference group of term-born children (≥ 37 weeks' GA) was used to define developmental delay in the

VP children for the Bayley-III, which acknowledges that Australian children perform above the current test norms.⁵ We also used the test norms to define developmental delay, which is how the Bayley-III was intended to be used. The local reference group for the Bayley-III comprised 220 term-born infants recruited at birth in 2005 as part of a prospective longitudinal cohort study, 202 (92%) of whom participated in the 24-month follow-up.⁵

The Human Research Ethics Committee of the Royal Children's Hospital approved the 24-month and 4-year follow-up studies of the VP children in the RCT and the local reference group at 24 months. Parents provided written informed consent before participation.

Assessment of Cognitive Function

Bayley-III

At 24 months, children were assessed using the Bayley-III⁴ by a psychologist or occupational therapist with Bayley-III certification blinded to the child's perinatal history. The Bayley-III generates scores for 3 composite indices (Cognitive, Language, Motor) and 5 subtests (Cognitive, Expressive Communication, Receptive Communication, Fine Motor, Gross Motor). In this study we focus on the Cognitive scale, which estimates general cognitive functioning on the basis of nonverbal activities involving memory, problem solving, and manipulation, and the Language scale, which estimates receptive communication, including verbal understanding and concept development, as well as expressive communication, including the ability to communicate through words and gestures. Age-standardized scores for each scale were calculated by using test norms (mean = 100; SD = 15). By using test norms and local term-born reference data (Cognitive: mean = 108.9; SD = 14.3; Language: mean = 108.2; SD = 14.8), "mild/

moderate developmental delay” was classified as scores < -1 SD relative to the mean, whereas “moderate developmental delay” referred to scores < -2 SDs relative to the mean. The Cognitive scale scores increase in 5-point steps (eg, 90, 95, 100, etc), whereas the Language scale scores increase in steps of 2 to 3 points. With the use of the Bayley-III norms, the cut-points are <85 and <70 to identify mild/moderate and moderate delay, respectively. By using the local reference data, the cut-points are <95 and <85 , respectively, for the Cognitive scale and <94 and <79 , respectively, for the Language scale, to allow for the discrete nature of the Bayley-III scores.

DAS-II

At 4 years, children were assessed using the DAS-II¹⁵ by a psychologist blinded to the child’s history. The DAS-II comprises a collection of subtests that assess general reasoning and conceptual abilities, which are used to generate a summary measure that is similar to IQ (general conceptual ability [GCA]). Additional summary indices include the Verbal index, which estimates acquired verbal concepts and knowledge, and the Nonverbal Reasoning index, which estimates complex nonverbal, inductive reasoning requiring mental processing. Age-standardized scores were calculated for each outcome by using test norms (mean = 100; SD = 15). “Mild/moderate impairment” was classified as scores <85 , and “moderate impairment” as a score <70 .

Statistical Analyses

Statistical analyses were performed by using Stata 13 (Stata Corp, College Station, TX) and SPSS 22 (IBM SPSS Statistics, IBM Corporation, Armonk, NY). The association between continuous scores on the Bayley-III Cognitive and Language scales at 24 months and the DAS-II GCA, Verbal, and Nonverbal Reasoning indices at

4 years was assessed by using linear regression models. Regression models were fitted with the use of generalized estimating equations to allow for the nonindependence of observations from twins/triplets, by using an exchangeable correlation structure, which assumes within a cluster any 2 observations are equally correlated. Results are presented as regression estimates and 95% confidence intervals (CIs) from separate regression models for each predictor/outcome relationship based on robust SEs, which are valid even if the assumption regarding the correlation structure is not correct. Sensitivity, specificity, and positive and negative predictive values along with 95% CIs were used to assess the mild/moderate and moderate developmental delay classifications on the Bayley-III scales for predicting mild/moderate and moderate impairment on the DAS-II indices. The association between the Bayley-III Language scale and the DAS-II Nonverbal Reasoning index was not examined because these measures estimate different domains. Receiver operating characteristic (ROC) curves were used to identify the optimal cut-points (based on the best combination of sensitivity and specificity) on the Bayley-III scales for predicting mild/moderate impairment and moderate impairment on the DAS-II indices. The

area under the curve (AUC) and its 95% CI were calculated for each ROC curve and are presented with *P* values testing the null hypothesis that AUC equals 0.50, which represents no relationship between the 2 scores.

RESULTS

Characteristics of the Sample

At 24 months’ CA, 115 of the 120 VP infants enrolled in the RCT completed the Bayley-III (3 infants died, 2 withdrew), and 105 of these completed the DAS-II at the 4-year follow-up and are included in the current analyses (9 additional children withdrew or could not be contacted, 1 did not complete the full assessment). A small number of children did not complete all Bayley-III and DAS-II subtests, and it was therefore not possible to calculate the scales and indexes for all 105 children. The mean CA of the VP group at the 24-month follow-up was 24.8 months (SD = 1.0) and at the 4-year follow-up was 53.2 months (SD = 3.1). The perinatal and demographic characteristics were similar between the VP children who did and did not complete the DAS-II (Table 1). The local term-born reference group at 24 months (mean CA = 24.6 months; SD 2.0) who had Bayley-III data (*n* = 190) comprised 89 (45%) boys and 60 (31%) with higher social risk. The VP and local

TABLE 1 Perinatal and Demographic Characteristics of the VP Children in the Current Study and in the Original RCT Who Did Not Complete the DAS-II at 4 Years

	DAS-II at 4 Years	No DAS-II at 4 Years
<i>n</i>	105	15
Gestational age, mean (SD), wk	27.3 (1.5)	27.3 (1.3)
Birth weight, mean (SD), g	1022 (268)	929 (249)
Boys, <i>n</i> (%)	56 (53)	5 (33)
Multiple births, <i>n</i> (%)	35 (33)	4 (27)
Higher social risk (score ≥ 2), ^a <i>n</i> (%)	48 (52)	.
Bronchopulmonary dysplasia, <i>n</i> (%)	33 (31)	2 (13)
Postnatal corticosteroids, <i>n</i> (%)	5 (5)	0 (0)
Grade 3 or 4 IVH, <i>n</i> (%)	6 (6)	0 (0)
Cystic PVL, <i>n</i> (%)	3 (3)	0 (0)

Percentages are based on those with available data. IVH, intraventricular hemorrhage; PVL, periventricular leukomalacia. ^a Social Risk Index¹⁶ scores range from 0 to 12 (higher scores reflect greater risk) and are based on family structure, primary caregiver education, primary income earner employment status and occupation, language spoken at home, and maternal age at birth of the child. Results reported are based on 93 children with social risk data collected at 4 years. There were no data available on social risk for those who did not attend the 4-year assessment.

term-born reference groups were similar for age and gender at the 24-month follow-up, but not for social risk status.

At 24 months, the mean scores on the Bayley-III Cognitive and Language scales in the VP group were 96.8 (SD = 12.6; $n = 105$) and 96.5 (SD = 15.3; $n = 101$), respectively. On the Cognitive scale, 12 (11%) had mild/moderate delay and 1 (1%) had moderate delay according to test norms, but 37 (35%) had mild/moderate and 12 (11%) had moderate delay based on cut-points using local term-born reference data. On the Language scale, 21 (21%) had mild/moderate delay and 2 (2%) had moderate delay using test norms, but 38 (38%) had mild/moderate delay and 16 (16%) had moderate delay relative to local term-born reference data.

At 4 years, the mean scores on the DAS-II indices in the VP group were as follows: GCA = 98.5 (SD = 15.4; $n = 103$); Verbal = 97.1 (SD = 15.1; $n = 104$); and Nonverbal Reasoning = 99.9 (SD = 13.1; $n = 101$). The rates of mild/moderate and moderate impairment, respectively, at 4 years were 17% and 7% on the GCA, 13% and 6% on the Verbal index, and 13% and 1% on the Nonverbal Reasoning index.

Predictive Ability of the Bayley-III Cognitive Scale

Higher Bayley-III Cognitive scores at 24 months were associated with higher DAS-II indices at 4 years: GCA regression coefficient = 0.78 (95% CI: 0.55–1.02; $P < .001$; variance explained = 37%) (Fig 1); Verbal regression coefficient = 0.63 (95% CI: 0.38–0.89; $P < .001$; variance explained = 28%); Nonverbal Reasoning regression coefficient = 0.56 (95% CI: 0.36–0.75; $P < .001$; variance explained = 24%).

The sensitivity of mild/moderate delay (<85) on the Cognitive scale for predicting mild/moderate

impairment on the DAS-II indices was low but specificity was high (Table 2). Results are not reported for the

moderate delay classification on the Bayley-III because this identified only 1 infant who did not have scores for

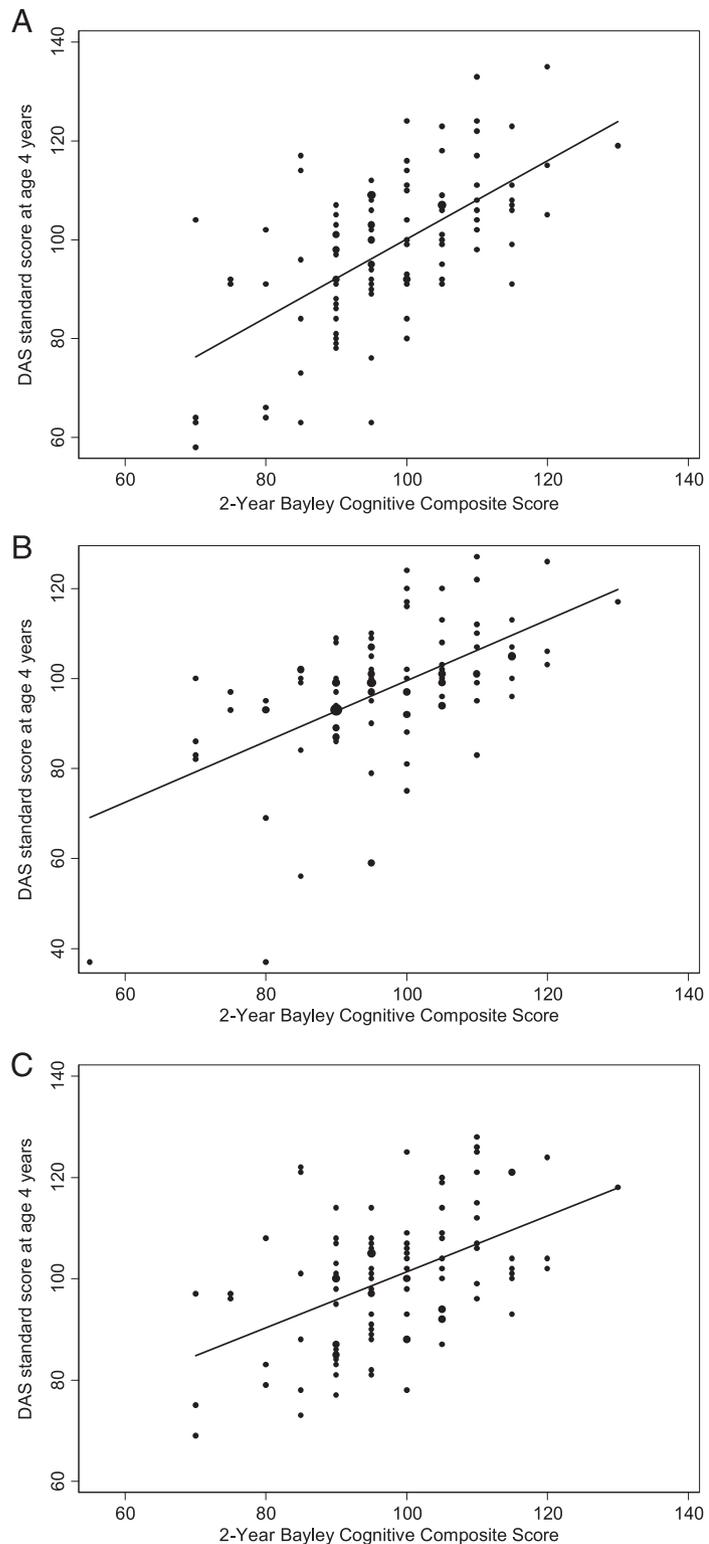


FIGURE 1 Scatterplots of the Bayley-III Cognitive scale score at 24 months against the DAS-II indices at 4 years, with fitted regression line. The size of the circles reflects the number of children with the same score on both measures. A, DAS-II GCA ($n = 103$); B, DAS-II Verbal ($n = 104$); C, DAS-II Nonverbal Reasoning ($n = 101$).

any index on the DAS-II at 4 years. By using local reference data, the sensitivity of mild/moderate delay (<95) on the Cognitive scale for predicting mild/moderate impairment on the DAS-II indices was higher, but specificity was lower than when using the test norms. The sensitivities of moderate delay (<85) on the Cognitive scale for predicting moderate impairment on the DAS-II indices were reasonable, but the CIs were wider than for mild/moderate delay (Table 2).

Predictive Ability of the Bayley-III Language Scale

Higher Bayley-III Language scale scores at 24 months were associated with higher DAS-II GCA and Verbal indices at 4 years: GCA regression coefficient = 0.45 (95% CI: 0.26–0.65; $P < .001$; variance explained = 20%) (Fig 2); Verbal regression coefficient = 0.50 (95% CI: 0.29–0.71; $P < .001$; variance explained = 28%). The sensitivity of mild/moderate delay (<85) on the Language scale for predicting mild/moderate impairment on DAS-II GCA and Verbal indices was low and specificity was high (Table 3). The sensitivity of the moderate delay (<70) on the Language scale for predicting moderate impairment on the DAS-II indices was low, although specificity was 100%. By using local reference data, the sensitivity of mild/moderate

delay (<94) on the Language scale for predicting mild/moderate impairment on the DAS-II indices was improved, but specificity was lower compared with using the test norms. The sensitivity of the moderate delay (<79) on the Language scale for predicting moderate impairment on the DAS-II indices was low, although specificity was high (Table 3).

Overall, the positive predictive values of the Bayley-III Cognitive and Language scales at 24 months for predicting impairment on DAS-II indices at 4 years were lower than the negative predictive values. When local term-born reference data were used compared with test norms, the positive predictive values were lower, reflecting the increasing number of false positives identified by using a higher cut-point.

Optimum Cut-points for Bayley-III Cognitive and Language Scales as a Predictor of Impairment on the DAS-II Indices From ROC Curves

The optimum cut-points for the Bayley-III scores and AUCs are shown in Table 4. The cut-points from the ROC curves were always closer to, and sometimes the same as, those from the local term-born reference data than they were to the Bayley-III normative cut-points. In each case, there was strong evidence of an association between delay on the

Bayley-III and impairment on the DAS-II (all $P < .05$).

DISCUSSION

In our cohort of Australian children born <30 weeks' GA, delay on the Bayley-III Cognitive and Language scales at 24 months was not strongly predictive of cognitive impairment at 4 years. The Bayley-III Cognitive and Language scale scores were positively associated with later functioning across general cognitive, verbal, and nonverbal reasoning domains. The strength of these associations was moderate to large for both Bayley-III scales. A recent systematic review showed a similar positive and moderate association for the earlier BSID-II MDI (a single index providing an estimate of cognitive and language development) and later childhood general intellectual but not language functioning in preterm populations.³ The developmental delay classifications of the Bayley-III Cognitive and Language scale scores for predicting future cognitive, verbal, and nonverbal reasoning impairments had concerning low levels of sensitivity, but high specificity, which is consistent with previous research on the Bayley-III scales^{8,11} and other early developmental tools.¹⁷ The use of local term-born reference data to determine developmental delay on

TABLE 2 Sensitivity, Specificity, and Positive and Negative Predictive Values for Delay on the Bayley-III Cognitive Scale at 24 Months Predicting Impairment on DAS-II Indices at 4 Years

Bayley-III Cognitive Scale at 24 Months (Cut-point)	DAS-II Index at 4 Years	<i>n</i>	Sensitivity (95% CI)	Specificity (95% CI)	Positive Predictive Value (95% CI)	Negative Predictive Value (95% CI)
Classified using test norms at 24 months ^a						
Mild/moderate delay (<85)	GCA <85	103	29.4 (10.3–56.0)	94.2 (87.0–98.1)	50.0 (18.7–81.3)	87.1 (78.5–93.2)
	Verbal <85	104	38.5 (13.9–68.4)	92.3 (84.8–96.9)	41.7 (15.2–72.3)	91.3 (83.6–96.2)
	Nonverbal Reasoning <85	101	30.8 (9.1–61.4)	95.5 (88.8–98.7)	50.0 (15.7–84.3)	90.3 (82.4–95.5)
Classified using local reference data at 24 months						
Mild/moderate delay (<95)	GCA <85	103	76.5 (50.1–93.2)	74.4 (63.9–83.2)	37.1 (21.5–51.1)	94.1 (85.6–98.4)
	Verbal <85	104	53.8 (25.1–80.8)	67.0 (56.4–76.5)	18.9 (8.0–35.2)	91.0 (81.5–96.6)
	Nonverbal Reasoning <85	101	76.9 (46.2–95.0)	73.9 (63.4–82.7)	30.3 (15.6–58.7)	95.6 (87.6–99.1)
Moderate delay (<85)	GCA <70	103	71.4 (29.0–96.3)	94.8 (88.3–98.3)	50.0 (18.7–91.3)	97.8 (92.4–99.7)
	Verbal <70	104	50.0 (11.8–88.2)	90.8 (83.3–95.7)	25.0 (5.5–57.2)	96.7 (90.8–99.3)
	Nonverbal Reasoning <70	101	100.0 (2.5–100.0)	93.0 (86.1–97.1)	12.5 (3.2–52.6)	100.0 (96.1–100.0)

Data are presented as percentages unless otherwise indicated.

^a Results are not presented for moderate delay (<70) for the Bayley-III at 24 months using test norms because this only identified 1 participant as impaired who was missing 2 out of the 3 outcomes on the DAS-II.

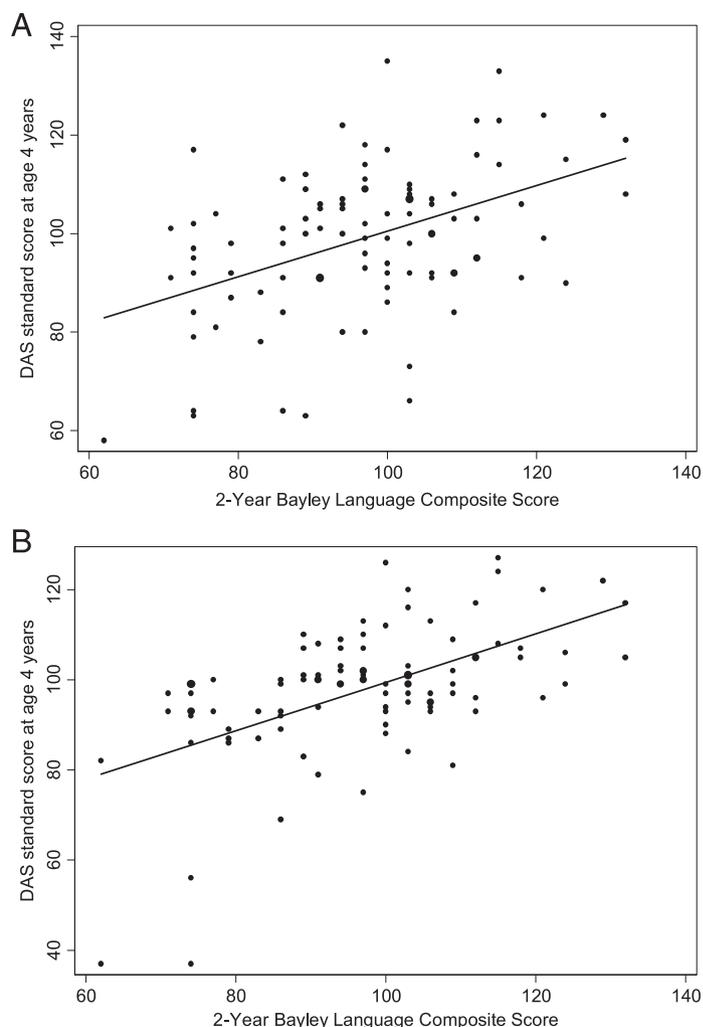


FIGURE 2 Scatterplots of the Bayley-III Language scale score at 24 months against the DAS-II indices at 4 years, with fitted regression line. The size of the circles reflects the number of children with the same score on both measures. A, DAS-II GCA ($n = 99$); B, DAS-II Verbal ($n = 100$).

the Bayley-III improved sensitivity, and cut-points were more consistent with those obtained from ROC curve

analysis than those from the Bayley-III normative data. This pattern of results suggests that the Bayley-III

cut-points for developmental delay might be too low and could be more useful if cut-points were higher, which is consistent with the suggestion of recent investigations by other groups.¹⁸ The Cognitive scale was equally associated with future verbal and nonverbal reasoning. A previous study in very low birth weight (<1500 g) children using the BSID MDI at 24 months reported higher sensitivity compared with our results (62%; 95% CI: 44–77%) and a similar level of specificity (89%; 95% CI: 79–95%) for predicting general intellectual functioning at 3.5 years by using the mild/moderate delay classification and low sensitivity (37%; 95% CI: 18–61%) and high specificity (97%; 95% CI: 91–99%) by using the moderate delay classification.¹⁹

The Bayley-III Cognitive and Language scale average scores were high (96.8 and 96.9, respectively) and rates of developmental delay with the use of test norms were low in our cohort of VP children at 24 months (mild/moderate delay: 11% and 21%, respectively; moderate delay: 1% and 2%, respectively). Higher rates of delay on the Cognitive and Language scales were observed when local term-born reference data were used to determine cut-points (mild/moderate delay: 35% and 38%, respectively; moderate delay: 11% and 16%, respectively). Our group has previously reported underestimation of the rates of delay

TABLE 3 Sensitivity, Specificity, and Positive and Negative Predictive Values for Delay on the Bayley-III Language Scale at 24 Months Predicting Impairment on the DAS-II Indices at 4 Years

Bayley-III Language Scale at 24 Months (Cut-point)	DAS-II Index at 4 Years (Cut-point)	<i>n</i>	Sensitivity (95% CI)	Specificity (95% CI)	Positive Predictive Value (95% CI)	Negative Predicted Value (95% CI)
Classified using test norms						
Mild/moderate delay (<85)	GCA <85	99	46.7 (21.3–73.4)	85.7 (76.4–92.4)	36.8 (16.3–61.6)	90.0 (81.2–95.6)
	Verbal <85	100	40.0 (12.2–73.8)	81.1 (71.5–88.6)	19.0 (5.4–41.9)	92.4 (84.2–97.2)
Moderate delay (<70)	GCA <70	99	16.7 (0.4–64.1)	100.0 (96.1–100.0)	100.0 (2.5–100.0)	94.9 (88.5–98.3)
	Verbal <70	100	25.0 (6.3–80.6)	99.0 (94.3–100.0)	50.0 (1.3–98.7)	96.9 (91.3–99.4)
Classified using local reference data						
Mild/moderate delay (<94)	GCA <85	99	66.7 (38.4–88.2)	69.0 (58.0–78.7)	27.8 (14.2–45.2)	92.1 (82.4–97.4)
	Verbal <85	100	70.0 (34.8–83.3)	65.6 (54.8–75.3)	18.4 (7.7–34.3)	95.2 (86.5–99.0)
Moderate delay (<79)	GCA <70	99	50.0 (11.8–88.2)	88.2 (79.8–94.0)	21.4 (4.7–50.8)	96.5 (99.0–99.3)
	Verbal <70	100	75.0 (19.4–99.4)	86.5 (78.0–92.6)	18.8 (4.0–45.6)	98.8 (93.5–100.0)

Data are presented as percentages unless otherwise indicated.

TABLE 4 Optimal Cut-points on the Bayley-III Scales and AUCs from ROC Analysis Assessing the Bayley-III Scales at 24 Months as a Diagnostic Tool for Predicting Impairment on the DAS-II Indices at 4 Years

	Impaired on DAS-II Index, <i>n</i>	Optimal Bayley-III Cut-point	AUC	95% CI	<i>P</i>
Bayley-III Cognitive scale					
Mild/moderate impairment on DAS-II index					
GCA	17	<95	0.824	0.728–0.919	<.001
Verbal	13	<90	0.724	0.569–0.878	.009
Nonverbal Reasoning	13	<95	0.831	0.730–0.932	<.001
Moderate impairment on DAS-II index ^a					
GCA	7	<90	0.917	0.813–1.0	<.001
Verbal	6	<90	0.812	0.662–0.963	.011
Bayley-III Language scale					
Mild/moderate impairment on DAS-II index					
GCA	17	<91	0.748	0.615–0.882	.002
Verbal	13	<94	0.718	0.548–0.888	.024
Moderate impairment on DAS-II index					
GCA	7	<91	0.798	0.618–0.979	.015
Verbal	6	<77	0.909	0.820–0.997	.006

^a AUC not calculated for moderate cognitive impairment because only 1 child had moderate impairment on the Nonverbal Reasoning index.

using the Bayley-III Motor scale with the use of test norms in this same cohort of VP children.⁸ Another group has reported a similar pattern of underestimation of the rates of delay for the Bayley-III Expressive and Receptive subscales with the use of test norms in a cohort of VP children.¹¹ An increasing number of published studies are in agreement that the Bayley-III scales overestimate developmental status in preterm populations, evidenced by higher average scores^{6,7,9} and lower rates of delay^{6,9,20} compared with earlier versions on the BSID-II MDI scores and higher rates of delay when using local term-born reference data to calculate cut-points for classifying the level of delay.^{5,6} Importantly, the current study contributes to this body of research by revealing discordance between developmental delay on the Bayley-III and later cognitive impairment. We observed higher rates of delay on the Bayley-III Language scale than with the Bayley-III Cognitive scale, almost twofold, which is consistent with some studies in preterm populations,^{5,9} and in contrast to other studies.^{20,21}

Our study findings should be considered in the context of some limitations. Although the Bayley-III was not designed to predict future cognitive functioning, understanding its predictive ability is crucial because it is often used across research and clinical settings in this way. Developmental assessments such as the Bayley-III are designed to assess developmental delay, and there is an expectation that some children who are delayed will show catch-up to their peers; therefore, modest agreement between Bayley-III and later IQ measures is to be expected. We assessed VP children at 24 months on the Bayley-III; however, its capacity to predict future functioning could vary with age at assessment, limiting the generalizability of our findings. A potentially stronger association between the Bayley-III Language scale and later language functioning might be revealed if a more comprehensive assessment of language was performed at the 4-year follow-up. Our findings suggest that DAS-II indices also underestimate impairment when using test norms, as evidenced by the mean scores and

rates of impairment of the VP cohort and the local term-born reference group. Another possible limitation is that the study sample was part of a clinical trial and may not represent all infants born <30 weeks' GA.

CONCLUSIONS

Although the Bayley-III Cognitive and Language scales at 24 months are associated with cognitive functioning at 4 years, developmental delay on these scales has low sensitivity in predicting later impairment. The use of local term-born reference data improved the identification of children destined to have later intellectual impairment. ROC analysis identified optimum cut-point scores for the Bayley-III that were more consistent with using local term-born reference data than with using the normative data for the Bayley-III. The clinical implications of these findings are that some at-risk children seen at 24 months are not being classified as delayed on the Bayley-III and accordingly might not receive the level of monitoring or early intervention that is warranted.

Address correspondence to Megan M. Spencer-Smith, PhD, School of Psychological Sciences, Monash University, Building 17, Wellington Rd, Clayton, Victoria 3800, Australia. E-mail: megan.spencer-smith@monash.edu

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

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

FUNDING: This study was funded by grants from the National Health and Medical Council (project grant 284512; Early Career Fellowship 1053767 [Dr Spittle], Senior Research Fellowship 1081288 [Dr Anderson]; Centre of Research Excellence 1060733), the Cerebral Palsy Alliance, Murdoch Children's Research Institute, Myer Foundation, Allens Arthur Robinson, Thyne Reid Foundation, and the Victorian Government's Operational Infrastructure Support Program.

POTENTIAL CONFLICT OF INTEREST: The authors have indicated they have no potential conflicts of interest to disclose.

REFERENCES

1. Bayley N. *Bayley Scales of Infant Development*. New York, NY: The Psychological Corporation; 1969
2. Bayley N. *Bayley Scales of Infant Development*. 2nd ed. San Antonio, TX: The Psychological Corporation; 1993
3. Luttikhuis dos Santos ES, 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(7):487–496
4. Bayley N. *Bayley Scales of Infant and Toddler Development*. San Antonio, TX: Harcourt Assessment; 2006
5. Anderson PJ, De Luca CR, Hutchinson E, Roberts G, Doyle LW; Victorian Infant Collaborative Group. Underestimation of developmental delay by the new Bayley-III Scale. *Arch Pediatr Adolesc Med*. 2010; 164(4):352–356
6. Lowe JR, Erickson SJ, Schrader R, Duncan AF. Comparison of the Bayley II Mental Developmental Index and the Bayley III Cognitive Scale: are we measuring the same thing? *Acta Paediatr*. 2012;101(2):e55–e58
7. Moore T, Hennessy EM, Myles J, et al. Neurological and developmental outcome in extremely preterm children born in England in 1995 and 2006: the EPICure studies. *BMJ*. 2012;345:e7961
8. Spittle AJ, Spencer-Smith MM, Eeles AL, et al. Does the Bayley-III Motor Scale at 2 years predict motor outcome at 4 years in very preterm children? *Dev Med Child Neurol*. 2013;55(5):448–452
9. Vohr BR, Stephens BE, Higgins RD, et al; Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network. Are outcomes of extremely preterm infants improving? Impact of Bayley assessment on outcomes. *J Pediatr*. 2012;162(2): 222–228.e3
10. Serenius F, Källén K, Blennow M, et al; EXPRESS Group. Neurodevelopmental outcome in extremely preterm infants at 2.5 years after active perinatal care in Sweden. *JAMA*. 2013;309(17): 1810–1820
11. Woods PL, Rieger I, Wocadlo C, Gordon A. Predicting the outcome of specific language impairment at five years of age through early developmental assessment in preterm infants. *Early Hum Dev*. 2014;90(10):613–619
12. Spencer-Smith MM, Spittle AJ, Doyle LW, et al. Long-term benefits of home-based preventive care for preterm infants: a randomized trial. *Pediatrics*. 2012; 130(6):1094–1101
13. Spittle AJ, Anderson PJ, Lee KJ, et al. Preventive care at home for very preterm infants improves infant and caregiver outcomes at 2 years. *Pediatrics*. 2010;126(1). Available at: www.pediatrics.org/cgi/content/full/126/1/e171
14. Spittle AJ, Ferretti C, Anderson PJ, et al. Improving the outcome of infants born at <30 weeks' gestation—a randomized controlled trial of preventative care at home. *BMC Pediatr*. 2009;9(1):73
15. Elliott C. *Differential Ability Scales-II (DAS-II)*. San Antonio, TX: Harcourt Assessment; 2007
16. Roberts G, Howard K, Spittle AJ, Brown NC, Anderson PJ, Doyle LW. Rates of early intervention services in very preterm children with developmental disabilities at age 2 years. *J Paediatr Child Health*. 2008;44(5):276–280
17. Brown N, Spittle A. Neurobehavioral evaluation in the preterm and term infant. *Curr Pediatr Rev*. 2014;10(1): 65–72
18. Johnson S, Moore T, Marlow N. Using the Bayley-III to assess neurodevelopmental delay: which cut-off should be used? *Pediatr Res*. 2014;75(5):670–674
19. Ross G, Lipper EG, Auld PA. Consistency and change in the development of premature infants weighing less than 1,501 grams at birth. *Pediatrics*. 1985; 76(6):885–891
20. Silveira R, Filipouski G, Goldstein D, O'Shea T, Procianny R. Agreement between Bayley Scales second and third edition assessments of very low-birth weight infants. *Arch Pediatr Adolesc Med*. 2012;166(11):1075–1076
21. Greene MM, Patra K, Silvestri JM, Nelson MN. Re-evaluating preterm infants with the Bayley-III: patterns and predictors of change. *Res Dev Disabil*. 2013;34(7): 2107–2117

Bayley-III Cognitive and Language Scales in Preterm Children

Megan M. Spencer-Smith, Alicia J. Spittle, Katherine J. Lee, Lex W. Doyle and Peter J. Anderson

Pediatrics 2015;135:e1258

DOI: 10.1542/peds.2014-3039 originally published online April 20, 2015;

Updated Information & Services	including high resolution figures, can be found at: http://pediatrics.aappublications.org/content/135/5/e1258
References	This article cites 16 articles, 4 of which you can access for free at: http://pediatrics.aappublications.org/content/135/5/e1258#BIBL
Subspecialty Collections	This article, along with others on similar topics, appears in the following collection(s): Developmental/Behavioral Pediatrics http://www.aappublications.org/cgi/collection/development:behavioral_issues_sub Cognition/Language/Learning Disorders http://www.aappublications.org/cgi/collection/cognition:language:learning_disorders_sub Fetus/Newborn Infant http://www.aappublications.org/cgi/collection/fetus:newborn_infant_sub Neonatology http://www.aappublications.org/cgi/collection/neonatology_sub
Permissions & Licensing	Information about reproducing this article in parts (figures, tables) or in its entirety can be found online at: http://www.aappublications.org/site/misc/Permissions.xhtml
Reprints	Information about ordering reprints can be found online: http://www.aappublications.org/site/misc/reprints.xhtml

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™



PEDIATRICS®

OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

Bayley-III Cognitive and Language Scales in Preterm Children

Megan M. Spencer-Smith, Alicia J. Spittle, Katherine J. Lee, Lex W. Doyle and Peter J. Anderson

Pediatrics 2015;135:e1258

DOI: 10.1542/peds.2014-3039 originally published online April 20, 2015;

The online version of this article, along with updated information and services, is located on the World Wide Web at:

<http://pediatrics.aappublications.org/content/135/5/e1258>

Pediatrics is the official journal of the American Academy of Pediatrics. A monthly publication, it has been published continuously since 1948. Pediatrics is owned, published, and trademarked by the American Academy of Pediatrics, 141 Northwest Point Boulevard, Elk Grove Village, Illinois, 60007. Copyright © 2015 by the American Academy of Pediatrics. All rights reserved. Print ISSN: 1073-0397.

American Academy of Pediatrics

DEDICATED TO THE HEALTH OF ALL CHILDREN™

