Original Article

Annals of Rehabilitation Medicine

The Validity of the Bayley-III and DDST-II in Preterm Infants With Neurodevelopmental Impairment: A Pilot Study

Seong Uk Jeong, MD[1], Ghi Chan Kim, MD, PhD[1], Ho Joong Jeong, MD, PhD[1], Dong Kyu Kim, MD[1], Yoo Rha Hong, MD[2], Hui Dong Kim, MD[1], Seok Gyo Park, MD[1], Young-Joo Sim, MD, PhD[1]

Departments of [1]Physical Medicine and Rehabilitation, [2]Pediatrics, Kosin University College of Medicine, Busan, Korea

Objective To identify the usefulness of both the Bayley Scales of Infant and Toddler Development, 3rd edition (Bayley-III) and Denver Developmental Screening Test II (DDST-II) in preterm babies with neurodevelopmental impairment, considering the detection rate as regulation of criteria.

Methods Retrospective medical chart reviews which included the Bayley-III and DDST-II, were conducted for 69 preterm babies. Detection rate of neurodevelopmental impairment in preterm babies were investigated by modulating scaled score of the Bayley-III. The detection rate of DDST-II was identified by regarding more than 1 caution as an abnormality. Then detection rates of each corrected age group were verified using conventional criteria.

Results When applying conventional criteria, 22 infants and 35 infants were detected as preterm babies with neurodevelopmental impairment, as per the Bayley-III and DDST-II evaluation, respectively. Detection rates increased by applying abnormal criteria that specified as less than 11 points in the Bayley-III scaled score. In DDST-II, detection rates rose from 50% to 68.6% using modified criteria. The detection rates were highest when performed after 12 months corrected age, being 100% in DDST II. The detection rate also increased when applying the modified criteria in both the Bayley-III and DDST-II.

Conclusion Accurate neurologic examination is more important for detection of preterm babies with neurodevelopmental impairment. We suggest further studies for the accurate modification of the detection criteria in DDST-II and the Bayley-III for preterm babies.

Keywords Premature birth, infant, Developmental disabilities

INTRODUCTION

Recent developments in perinatal management and neonatology have contributed to increased survival rates of preterm infants, low birth weight infants, and newborn infants with impairments [1]. The number of infants with impairments and subject to special education is increasing along with congenital factors, which in turn increases the importance of pediatric rehabilitation. However, few studies have been conducted on the long-term morbidity of infants with neurodevelopmental impairment.

Compared with normal birth weight babies, preterm and low birth weight infants do not have effective immune systems, and their brains and nervous systems (e.g., spinal cord) are vulnerable to infection or trauma due to structural problems (e.g., germinal matrix), which are likely to lead to neurological impairment [2]. Factors for impairment in infants can be detected using imaging techniques such as ultrasound and MRI, which helps predict their development [3]. However, there may be some developmental impairments that cannot be detected using these techniques, and even if issues are detected, the clinical severities are difficult to determine using imaging. Hence, the clinical practice for neurologic examination includes developmental screenings, in addition to imaging techniques for preterm or low birth weight infants showing abnormalities. The developmental screening tests currently used include the Denver Developmental Screening Test II (DDST-II), Bayley Scales of Infant and Toddler Development, 2nd edition (BSID-II), 3rd edition (Bayley-III), Korean Developmental Screening Test for Infants & Children (K-DST), and Peabody Development Motor Scales and Korean-Ages & Stage questionnaires (K-ASQ) test. These screening tests allow the early detection of developmental impairment and serve as tools for follow-up studies, resulting in therapeutic intervention and improved prognosis [4-6].

The DDST was first published in 1967, with an improved version, DDST-II, published in 1992. The DDST-II targets infants and children 0–6 years old. It is composed of 4 sections (personal-social development, fine motor-adaptive development, language development, and gross motor development), and all items are established based on age.

The Bayley-III published in 2006 targets infants and children 1–42 months old, and consists of 5 sections (motor development, cognition development, language development, social emotion development, and adaptive behavior development). In contrast to the BSID-II, which is based on mental, motor, and behavioral development, the Bayley-III separates mental development into cognition and communication, and motor development into gross motor and fine motor. Therefore, the Bayley-III is more precise in determining the developmental impairment. In addition, it is able to convert composite scores into scaled scores based on age, thereby being more effective in detecting developmental impairment in clinical practice.

In this study, the Bayley-III and DDST-II tests were conducted for preterm infants with neurologic impairment. No studies have evaluated the detection rate of these two tests for infants with neurodevelopmental impairment. Furthermore, a number of cases have been reported where infant patients detected with neurodevelopmental impairment in screening tests have developed normally.

This study therefore aimed to establish appropriate tests and test criteria for infants with developmental impairment, by determining the detection rates of the Bayley-III and the DDST-II, thereby increasing the accuracy of early screening tests by determining the level of agreement between tests and areas requiring mutual supplementation.

MATERIALS AND METHODS

Subjects

Subjects were selected from infant patients who visited or were admitted to Department of Pediatrics and Rehabilitation Medicine, Kosin University Gospel Hospital from June 2010 until August 2013. Infants diagnosed with neurological impairment after neurological assessment by a rehabilitation specialist, and who underwent both the Bayley-III and DDST-II, were included in the study.

The neurological examinations included tests for muscle tone, primitive reflex, Vojta reaction, cranial nerve, muscle power, deep tendon reflex, sensory function, and general movement. The subjects of this study were 74 premature infants less than 37 weeks old, with an average gestational age of 31 weeks and 5 days (min 25 weeks; max 36 weeks and 6 days). The average birth weight was 1,641.3 g (min 610 g; max 2,940 g). There were 28 infants with extremely low birth weight (<1,500 g). Five patients

were excluded because they were only tested for some of the items of the Bayley-III. Therefore, a final of total 69 patients were selected as subjects of this study. The time between administration of the Bayley-III and the DDST-II did not exceed one week, and the two tests were conducted by a well-trained occupational therapist, to ensure an objective and consistent study.

Methods

The Bayley-III, a tool used to determine the degree of development in infants aged 1–42 months, consists of tests for cognition, language (receptive and expressive), motor skills (fine/gross), and social-emotion and adaptive behaviors. The test was conducted in accordance with standard guidelines. The raw scores for respective items were converted into scaled scores for application, and the range of the scaled scores was set from 1 to 19 points. In the Bayley-III evaluation criteria, 10 points is average, 7 points is -1 standard deviation (SD), and 4 points is -2 SD. The raw scores of respective items were converted into scaled scores, and cases where patients scored less than 7 points were identified as developmental impairment. In the DDST-II, if there are two or more cautions and/or one or more failures at baseline, the subject is suspected of having a developmental delay.

In this study, we modified the criteria for the Bayley-III and DDST-II tests, with the aim to increase the detection rate of neurodevelopmental impairment. For the Bayley-III, we changed the base for an abnormal scaled score adjusting the score by 1 point, thereby considering up to below 11 points to indicate gradually increasing abnormality. For modified criteria of the DDST-II, one or more cautions were considered abnormal.

We compared items and corrected age in the Bayley-III and DDST-II to determine the detection rate for respective age groups. Corrected age groups spanned up to 24 months, in intervals of 6 months. Additionally, the level of agreement of the sector between the two tests was determined by the kappa value.

We applied each item of the DDST-II and the Bayley-III to their corresponding sectors, using 'cognitive' for 'social', 'fine motor' for 'fine motor-adaptive', 'language (receptive communication)' and 'language (expressive communication)' for 'language', and 'gross motor' for 'gross motor'.

RESULTS

Of the 69 infants with neurological impairment, the gestational age of 15 infants was 34– weeks (late preterm). There were 40 males and 29 females. Diagnostic implications included 19 patients with hemorrhagic lesion, 25 with cavum septi pellucidum or cavum vergae, 6 with cystic lesion, and 19 with no abnormal findings in brain imaging (Table 1).

The infants were tested for all items for their age group in the Bayley-III and DDST-II. As per the conventional criteria, which considers a scaled score of below 7 points as abnormal for the Bayley-III, the detection rate was 22 infants (31.88%). When the modified criteria with a scaled score from below 8 to below 11 were applied, the detection rate increased from 30 infants (43.48%) to 68 infants (98.55%) gradually, indicating that increase in the reference point for abnormality increases the detection rate (Table 2).

Among infant patients with suspected neurological abnormalities who underwent the DDST-II, 35 infants (50.7%) had neurological abnormalities based on criteria of two or more cautions and or/one or more delays at age baseline. If the criteria, considering one or more cautions to be abnormal are applied, the total detection rate was 48 infants (69.6%) (Table 3).

Table 1. Demographic characteristics

Characteristic	Value
Gestational age (wk)	31.5±3.1
≤34	54
>34 and <37	15
Sex	
Male	40
Female	29
Birth weight (g)	1,653.9±617.5
Brain image findings	
Hemorrhagic lesion[a]	19
Cavum septi pellucidum or cavum vergae	25
Cystic lesion	6
None	19

Values are presented as mean±standard deviation or number of cases.

[a]Hemorrhagic lesion includes all of hemorrhage in brain (e.g., germinal matrix hemorrhage, microbleed, subependymal hemorrhage).

Table 2. Detection rate using modified criteria for the Bayley-III scaled scores

	Bayley-III scaled score (criteria for abnormal scaled score)				
	<7	<8	<9	<10	<11
Cognitive	5	7	12	18	28
Receptive communication	3	8	14	23	43
Expressive communication	9	15	25	46	62
Gross motor	14	19	24	36	47
Fine motor	6	6	12	20	30
Total detection number (rate)	22 (31.88)	30 (43.48)	36 (52.17)	52 (75.36)	68 (98.55)

Values are presented as number (%).
Total number of infants with neurodevelopmental impairment is 69.
Bayley-III, Bayley Scales of Infant and Toddler Development, 3rd edition.

Table 3. Detection rate using modified criteria for the DDST-II

	Criteria for abnormal DDST-II	
	≥2 cautions or ≥1 delay	≥1 caution
Personal-social	6 (8.7)	15 (21.7)
Language	13 (18.8)	21 (30.4)
Gross motor	22 (31.9)	33 (47.8)
Fine motor-adaptive	7 (10.1)	16 (23.2)
Total detection number (rate)	35 (50.7)	48 (69.6)

Values are presented as number (%).
Total number of infants with neurodevelopmental impairment is 69.
DDST-II, Denver Developmental Screening Test II.

With corrected ages set at 6-month intervals, a corrected age of 7–12 months showed the highest detection rate in 'cognition', 'gross motor,' and 'fine motor' using the Bayley-III, whereas a corrected age of 13–18 months showed the highest detection rate in 'language' (receptive communication and expressive communication). In contrast, a corrected age of 19–24 months showed the highest detection rate in 'cognition (personal-social) and language' using the DDST-II, whereas a corrected age of 13–18 months showed the highest detection rate in 'gross motor' and 'fine motor' (Table 4).

For the level of agreement between the DDST-II and the Bayley-III, the correlation coefficient between the language score on the DDST-II and the language (expressive communication) score on the Bayley-III was 0.355 (p=0.003), indicating fair agreement. Comparing language on the DDST-II and language (receptive communication) on the Bayley-III, the correlation coefficient was 0.193 (p=0.03), indicating a slight agreement. All results showed statistically significant correlation. The correlation coefficient between the 'personal-social' score on

the DDST-II and the 'cognitive' score on the Bayley-III was 0.112 (p=0.352), fine motor-adaptive and fine motor was 0.066 (p=0.580), and the gross motor scores was 0.04 (p=0.731), indicating a slight agreement. However, none of these results were statistically significant. The overall agreement between the DDST-II and the Bayley-III was 0.048 (p=0.664), indicating a slight agreement, but these results were not statistically significant. 'Language' on the DDST-II and 'language' (expressive communication) on the Bayley-III showed fair agreement, but the overall agreement between the remaining items and between these two tests did not exceed 'fair agreement' (Table 5).

DISCUSSION

Preterm and low birth weight infants are at a risk of behavioral, motor, and cognitive impairment, and early detection and subsequent treatment are critical to prognosis [5-7]. Clinical practice in Korea uses several tests to predict developmental impairment, including the DDST-II, the Sequenced Language Scale for Infants (SELSI), the

Table 4. Detection rates of neurodevelopmental impairment by corrected age group using the conventional criteria of the Bayley-III and DDST-II

		Corrected age (mo)			
		0–6 (n=17)	7–12 (n=29)	13–18 (n=14)	19–24 (n=9)
Cognition					
Cognitive[a]	Normal	16	26	13	9
	Abnormal	1 (6)	3 (10.3)	1 (7.1)	0 (0)
Personal-social[b]	Normal	17	27	12	7
	Abnormal	0 (0)	2 (6.9)	2 (14.3)	2 (22.2)
Language					
Receptive communication[a]	Normal	17	27	13	9
	Abnormal	0 (0)	2 (6.9)	1 (7.1)	0 (0)
Expressive communication[a]	Normal	16	26	11	8
	Abnormal	1 (6)	3 (10.3)	3 (21.4)	1 (11.1)
Language[b]	Normal	17	26	8	5
	Abnormal	0 (0)	3 (10.3)	6 (42.9)	4 (44.4)
Motor					
Gross motor[a]	Normal	14	20	12	9
	Abnormal	3 (17.6)	9 (31.0)	2 (16.7)	0 (0.0)
Gross motor[b]	Normal	17	20	5	5
	Abnormal	0 (0)	9 (31.0)	9 (64.3)	4 (44.4)
Fine motor[a]	Normal	16	25	13	9
	Abnormal	1 (6)	4 (13.8)	1 (0.7)	0 (0)
Fine motor-adaptive[b]	Normal	17	28	10	7
	Abnormal	0 (0)	1 (0.34)	4 (28.6)	2 (22.2)
Total					
Bayley-III	Normal	16	20	5	5
	Abnormal	1 (6)	9 (31.0)	9 (64.3)	4 (44.4)
DDST-II	Normal	17	17	0	0
	Abnormal	0 (0)	12 (41.4)	14 (100)	9 (100)

Values are presented as number (%).

Bayley-III, Bayley Scales of Infant and Toddler Development, 3rd edition; DDST-II, Denver Developmental Screening Test II.

[a] Bayley-III subtest (developmental disorder baseline: scaled score of less than 7).
[b] Sector of DDST-II.

Table 5. Agreement between the Bayley-III and the DDST-II

DDST-II	Bayley-III	Kappa value	p-value
Personal-social	Cognitive	0.112	0.352
Fine motor-adaptive	Fine motor	0.066	0.580
Language	Receptive communication	0.193	0.030
	Expressive communication	0.355	0.003
Gross motor	Gross motor	0.040	0.731
Total	Total	0.048	0.664

Bayley-III, Bayley Scales of Infant and Toddler Development, 3rd edition; DDST-II, Denver Developmental Screening Test II.

K-DST, and the K-ASQ [8]. The K-DST is used for patients 4–71 months old, and observes infants and children or conducts a questionnaire survey with parents; however, it is a recently developed test and is therefore not used widely. The K-ASQ has the advantage of allowing parents to observe their infants and reply to the questionnaire survey by post. However, the results may be subjective since the survey is conducted with the parents. The SELSI is for infants under 36 months old, and can only be used to evaluate delays in language development.

In the DDST-II and Bayley-III, medical personnel objectively observe the infants for the given criteria; these tests have the advantage of being able to determine quantified and age-corrected scores [8-10]. Due to their ability to evaluate the developmental status of infant patients in general (e.g., language, motor, cognition, and sociality), the DDST-II and Bayley-III are widely used for early developmental screening and developmental impairment diagnosis [11-13]. The DDST-II is intended to screen for developmental delays, and numerous studies have confirmed its validity and reliability in infants [4,14,15].

The BSID-II is composed of only two types of developmental scores: the Mental Developmental Index (MDI) consisting of cognitive and language development, and the Psychomotor Development Index (PDI) consisting of fine and gross motor skill development. The Bayley-III is also able to calculate 'cognition', 'language,' and 'motor' skills as separate composite scores, as well as determine the level of development by sub-dividing the sectors into 'receptive communication', 'expressive communication,' and 'fine and gross motor', using scaled scores. Additionally, it approaches the social-emotional and adaptive behaviors of infants by integrating parent questionnaires. For these reasons, the Bayley-III provides more useful clinical information regarding the early development of a high risk population.

Very few surveys have been conducted to determine the prevalence of developmental disabilities [16], particularly in Korea. Furthermore, there are difficulties in applying most of these testing methods to developmental delays for infants in Korea, because they were designed for infants in the US.

The results of this study can be summarized as follows: when the Bayley-III test with conventional criteria was conducted by a skilled rehabilitation specialist in infant patients with suspected neurological abnormalities, the detection rate was 31.88%. The detection rate for the DDST-II with conventional criteria was 50.70%. The detection rate was higher for the DDST-II than the Bayley-III, which is in agreement with the results of recent studies which demonstrate that the Bayley-III tends to overestimate development [11,17,18].

We estimated detection rate of Bayley-III and DDST-II in premature infants less than 34 weeks of gestational age, at 34 weeks, and above 34 weeks of gestational age.

Among the infants with neurological impairment in cognition and motor skills, the Bayley-III had a higher detection rate for those under 12 months, and the DDST-II had a higher detection rate for at age 12–24 months. In language, both tests had higher detection rates for infants 12–24 months old. These results suggest that the Bayley-III is superior for infants under 12 months, whereas the DDST-II should be recommended for 12–24 month infants suspected of neurological developmental impairment in cognition and motor skills. It is believed that the evaluation of language development (e.g., SELSI, etc.) is required for infants under 12 months old who are suspected of neurological developmental impairment in language.

Both the Bayley-III and DDST-II had relatively high detection rates for infants with neurological impairment above a corrected age of 13 months, yet they had rather low detection rates for infants under 13 months. In particular, the DDST-II presented a 100% detection rate for infants 13–24 months old.

When conventional criteria were applied, the DDST-II and Bayley-III had detection rates as low as 50.7% and 31.88%, respectively. When modified criteria for developmental impairment were applied, with one or more cautions in the DDST-II, the detection rates increased to 69.6%, and with scores of below 8 or 11 points in the Bayley-III, the detection rates increased to 43.48% and 98.5%, respectively.

In this study, an agreement between the Bayley-III and DDST-II was very low, with varying detection rates of neurological impairment, depending on the respective tests and development sector. Furthermore, the detection rate of infants with neurological impairment was only 31.88%–50.7% when the conventional criteria were applied in the Bayley-III and the DDST-II. The use of modified criteria facilitated early detection of infants with

neurological impairment. Since the possibility of early rehabilitation and normal development increases when the neuroplasticity of the child is taken into consideration, careful determination of abnormalities should be made using the Bayley-III and DDST-II.

There are some limitations of this study. The sample size was small, hence there were difficulties in powering the desired effect. Furthermore, for infants with clinically determined neurological developmental impairment, the sectors were not subdivided into motor, cognition, and language. The maximum corrected age of the subjects was 24 months, which failed to represent a study population of a wider age range. Therefore, we recommend a large-scale research on infant patients diagnosed with neurodevelopmental impairments, to reestablish the criteria for screening tests that are appropriate for Korean infants.

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

1. Stoll BJ. Overview of mortality and morbidity. In: Kliegman RM, Behrman RE, Jenson HB, Stanton BM, editors. Nelson textbook of pediatrics. 18th ed. Philadelphia: Saunders; 2007. p. 671-5.
2. VOHR, B. Follow-up care of high-risk infants. Pediatrics 2004;114:2.
3. de Vries LS, Benders MJ, Groenendaal F. Imaging the premature brain: ultrasound or MRI? Neuroradiology 2013;55 Suppl 2:13-22.
4. Galascoe FP. Developmental screening and surveillance. In: Kliegman RM, Behrman RE, Jenson HB, Stanton BM, editors. Nelson textbook of pediatrics. 18th ed. Philadelphia: Saunders; 2007. p. 74-81.
5. Spittle AJ, Orton J, Doyle LW, Boyd R. Early developmental intervention programs post hospital discharge to prevent motor and cognitive impairments in preterm infants. Cochrane Database Syst Rev 2007;18: CD005495.
6. Spittle AJ, Ferretti C, Anderson PJ. Improving the outcome of infants born at <30 weeks' gestation-a randomized controlled trial of preventative care at home. BMC Pediatr 2009;9:73.
7. Doyle LW. Evaluation of neonatal intensive care for extremely low birth weight infants in Victoria over two decades. I. Effectiveness. Pediatrics 2004;113:505-9.
8. Kwun Y, Park HW, Kim MJ. Validity of the ages and stages questionnaires in Korean compared to Bayley Scales of Infant Development-II for screening preterm infants at corrected age of 18-24 months for neurodevelopmental delay. J Korean Med Sci 2015;30:450-5.
9. Kim MS, Kim JK. Assessment of children with developmental delay: Korean-ages & stages questionnaires (K-ASQ) and Bayley Scales of Infant Development Test II (BSID-II). J Korean Child Neurol Soc 2010;18:49-57.
10. Vohr BR, Msall ME. Neuropsychological and functional outcomes of very low birth weight infants. Semin Perinatol 1997;21:202-20.
11. 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;164:352-6.
12. Ga HY, Kwon JY. A comparison of the Korean-ages and stages questionnaires and Denver Developmental Delay Screening Test. Ann Rehabil Med 2011;35:369-74.
13. Lee JH, Lim HK, Park E, Song J, Lee HS, Ko J, et al. Reliability and applicability of the Bayley Scale of Infant Development-II for children with cerebral palsy. Ann Rehabil Med 2013;37:167-74.
14. Lee K. Denver II Developmental Screening Test and development of Seoul children. Korean J Pediatr 1996; 39:1210-5.
15. Jeon MC, Kim YH, Chung SY, Lee IG, Kim JW, Whang KT. Accuracy of Denver II in developmental delay screening. J Korean Child Neurol Soc 1997;5:111-8.
16. Boyle CA, Decoufle P, Yeargin-Allsopp M. Prevalence and health impact of developmental disabilities in US children. Pediatrics 1994;93:399-403.
17. Reuner G, Fields AC, Wittke A, Lopprich 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.
18. Acton BV, Biggs WS, Creighton DE, Penner KA, Switzer HN, Thomas JH, et al. Overestimating neurodevelopment using the Bayley-III after early complex cardiac surgery. Pediatrics 2011;128:e794-800.