Are Outcomes of Extremely Preterm Infants Improving? Impact of Bayley Assessment on Outcomes

Betty R. Vohr, MD[1], Bonnie E. Stephens, MD[1], Rosemary D. Higgins, MD[2], Carla M. Bann, PhD[3], Susan R. Hintz, MD, MS Epi[4], Abhik Das, PhD[5], Jamie E. Newman, PhD, MPH[3], Myriam Peralta-Carcelen, MD, MPH[6], Kimberly Yolton, PhD[7], Anna M. Dusick, MD, FAAP[8], Patricia W. Evans, MD[9], Ricki F. Goldstein, MD[10], Richard A. Ehrenkranz, MD[11], Athina Pappas, MD[12], Ira Adams-Chapman, MD[13], Deanne E. Wilson-Costello, MD[14], Charles R. Bauer, MD[15], Anna Bodnar, MD[16], Roy J. Heyne, MD[17], Yvonne E. Vaucher, MD, MPH[18], Robert G. Dillard, MD[19], Michael J. Acarregui, MD[20], Elisabeth C. McGowan, MD[21], Gary J. Myers, MD[22], and Janell Fuller, MD[23], for the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development Neonatal Research Network*

Objectives To compare 18- to 22-month cognitive scores and neurodevelopmental impairment (NDI) in 2 time periods using the National Institute of Child Health and Human Development's Neonatal Research Network assessment of extremely low birth weight infants with the Bayley Scales of Infant Development, Second Edition (Bayley II) in 2006-2007 (period 1) and using the Bayley Scales of Infant and Toddler Development, Third Edition (Bayley III), with separate cognitive and language scores, in 2008-2011 (period 2).

Study design Scores were compared with bivariate analysis, and regression analyses were run to identify differences in NDI rates.

Results Mean Bayley III cognitive scores were 11 points higher than mean Bayley II cognitive scores. The NDI rate was reduced by 70% (from 43% in period 1 to 13% in period 2; $P < .0001$). Multivariate analyses revealed that Bayley III contributed to a decreased risk of NDI by 5 definitions: cognitive score <70 and <85, cognitive or language score <70; cognitive or motor score <70, and cognitive, language, or motor score <70 ($P < .001$).

Conclusion Whether the Bayley III is overestimating cognitive performance or whether it is a more valid assessment of emerging cognitive skills than the Bayley II is uncertain. Because the Bayley III identifies significantly fewer children with disability, it is recommended that all extremely low birth weight infants be offered early intervention services at the time of discharge from the neonatal intensive care unit, and that Bayley scores be interpreted with caution. *(J Pediatr 2012;161:222-8)*.

Extremely preterm (PT) infants are at increased risk for cognitive impairments. The most common individual severe impairment identified using the standard composite outcome of neurodevelopmental impairment (NDI) for extremely low birth weight (ELBW) infants at age 18 and 30 months is cognitive impairment, defined as a score >2 SD below the mean (<70).[1-5] Early developmental/cognitive function of PT children has traditionally been assessed using the Bayley Scales of Infant Development.[6,7] Updates of developmental tests such as the Bayley Scales are standard because of steadily increasing scores over

Bayley II	Bayley Scales of Infant Development, Second Edition	GMFCS	Gross Motor Function Classification System
Bayley III	Bayley Scales of Infant and Toddler Development, Third Edition	IVH	Intraventricular hemorrhage
		MDI	Mental Developmental Index
		NDI	Neurodevelopmental impairment
BPD	Bronchopulmonary dysplasia	NEC	Necrotizing enterocolitis
CA	Corrected age	NRN	Neonatal Research Network
Cog	Cognitive composite	PDI	Psychomotor Developmental Index
CP	Cerebral palsy		
ELBW	Extremely low birth weight	PT	Preterm
GA	Gestational age	PVL	Periventricular leukomalacia
		ROP	Retinopathy of prematurity

From the [1]Department of Pediatrics, Women & Infants' Hospital, Brown University, Providence, RI; [2]*Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, MD; [3]Statistics and Epidemiology Unit, RTI International, Research Triangle Park, NC; [4]Department of Pediatrics, Stanford University School of Medicine, Palo Alto, CA; [5]Statistics and Epidemiology Unit, RTI International, Rockville, MD; [6]Division of Neonatology, University of Alabama at Birmingham, Birmingham, AL; [7]Department of Pediatrics, University of Cincinnati, Cincinnati, OH; [8]Department of Pediatrics, Indiana University School of Medicine, Indianapolis, IN; [9]Department of Pediatrics, University of Texas Medical School at Houston, Houston, TX; [10]Department of Pediatrics, Duke University, Durham, NC; [11]Department of Pediatrics, Yale University School of Medicine, New Haven, CT; [12]Department of Pediatrics, Wayne State University, Detroit, MI; [13]Department of Pediatrics, Emory University School of Medicine and Children's Healthcare of Atlanta, Atlanta, GA; [14]Department of Pediatrics, Rainbow Babies & Children's Hospital, Case Western Reserve University, Cleveland, OH; [15]University of Miami Miller School of Medicine, Miami, FL; [16]Department of Pediatrics, Division of Neonatology, University of Utah School of Medicine, Salt Lake City, UT; [17]Department of Pediatrics, University of Texas Southwestern Medical Center, Dallas, TX; [18]Division of Neonatology, University of California San Diego, La Jolla, CA; [19]Wake Forest University School of Medicine, Winston-Salem, NC; [20]Department of Pediatrics, University of Iowa, Iowa City, IA; [21]Department of Pediatrics, Division of Newborn Medicine, Floating Hospital for Children, Tufts Medical Center, Boston, MA; [22]University of Rochester School of Medicine and Dentistry, Rochester, NY; and [23]University of New Mexico Health Sciences Center, Albuquerque, NM

*A list of members of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development Neonatal Research Network is available at www.jpeds.com (**Appendix**).

Supported by grants from the National Institutes of Health and the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development. Data collected at participating sites of the National Institute of Child Health and Human Development's Neonatal Research Network were transmitted to RTI International, the data coordinating center for the network, which stored, managed, and analyzed the data for this study. The authors declare no conflicts of interest.

 10.1016/j.jpeds.2012.01.057

time (3 points per year), a phenomenon known as the Flynn effect.[8] There is some evidence of this effect. A report from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development Neonatal Research Network (NRN) on Bayley scores of infants born at gestational age (GA) <32 weeks demonstrated increasing scores between 1993 and 1998.[3] A limitation of the Bayley Scales of Infant Development, Second Edition (Bayley II) is its inclusion of only 2 developmental scores, the Mental Developmental Index (MDI), a composite of cognitive, receptive language, and expressive language tasks, and the Psychomotor Developmental Index (PDI), a composite of fine and gross motor skills. This limitation contributed to the development of the Bayley Scales of Infant and Toddler Development, Third Edition (Bayley III),[7] which contains 3 individual developmental scores: a cognitive composite score (Cog), a language composite score (with receptive and expressive subscores), and a motor composite score (with gross and fine motor subscores), in addition to social-emotional and adaptive behavior domains. This allows the examiner to identify a deficit in a specific developmental domain as well as relative strengths and challenges. It was expected that separating the language scores from the Cog score might result in higher cognitive scores.

The National Institute of Child Health and Human Development's NRN has been reporting on the outcomes of ELBW infants (birth weight <1000 g) with the Bayley II since 1993. The NRN converted to the Bayley III in January 2008. This change in protocol provides an opportunity to compare rates of low (<2 SD) Bayley II MDI scores and low Bayley III Cog scores.

The objectives of the present study were (1) to compare cognitive outcomes in infants born at a GA of $\leq 26^{6/7}$ weeks and a birth weight of ≤1000 g and assessed during period 1 (2006-2007) using the Bayley II and in infants assessed during period 2 (2008-2011) using the Bayley III; (2) to explore the utility of a threshold of <70 versus <85 on the Bayley III to reflect impairment; and (3) to compare the association of major neonatal morbidities with NDI rates between period 1 and period 2. We hypothesized that cognitive scores would be higher and NDI rates lower in period 2, and that the Bayley III would have comparable performance to the Bayley II in identifying an increased risk of NDI in infants with major neonatal morbidities.

Methods

This was a retrospective cohort study of ELBW infants with a birth weight of 401-1000 g and born at a GA of <27 weeks who were admitted to one of 20 neonatal intensive care units in the NRN and underwent comprehensive neurologic and developmental assessments at 18-22 months corrected age (CA) during calendar years 2006-2011. The infants were categorized into 2 groups, those assessed during period 1 using the Bayley II (n = 1012) and those assessed during period 2 using the Bayley III (n = 1616). In addition, to address potential differences in factors other than Bayley scores that could affect the rate of NDI, we used propensity score matching with SAS software (SAS Institute, Cary, North Carolina)[9] to select subsamples of children from the 2 time periods who were similar based on the following characteristics: maternal age, education, race, public insurance coverage, antenatal steroid use, cesarean delivery, birth weight, GA, multiple birth, child's sex, intraventricular hemorrhage (IVH)/periventricular leukomalacia (PVL), necrotizing enterocolitis (NEC), sepsis, bronchopulmonary dysplasia (BPD), retinopathy of prematurity (ROP), postnatal steroid use, number of days on ventilation, duration of hospital stay, adjusted age at follow-up assessment, cerebral palsy (CP), vision impairment, hearing impairment, and research center. These matched samples included 922 children from each time period for a total sample of 1844.

All maternal and neonatal data, treatments, and clinical outcomes are prospectively collected in the NRN generic database. Centers participating in the NRN received local Institutional Review Board approval for data collection. Trained research coordinators obtained data based on the definitions listed in the NRN Manual of Operations. The effects of 4 major neonatal risk factors (brain injury, defined as IVH grade 3-4 or cystic PVL; BPD; NEC; and sepsis) on NDI rates were examined. BPD was defined as treatment with supplemental oxygen at 36 weeks gestation. IVH grade 3-4 was based on the Papile classification scheme,[10] and represents the maximum grade noted on cranial ultrasound before discharge. Cystic PVL was based on the reported head ultrasound examination. Sepsis was combined early-onset sepsis (≤72 hours) and/or late-onset sepsis (>72 hours) based on positive blood cultures, and NEC was defined as modified Bell classification stage II A or higher.

The follow-up evaluation included neurologic, hearing, vision, and developmental assessment. The NRN has protocols in place for annual training of all examiners to ensure reliability of all study assessments.[11] The neurologic examination, based on the Amiel-Tison assessment scheme,[12] includes an evaluation of tone, strength, reflexes, angles, and posture. CP was defined as a nonprogressive central nervous system disorder characterized by abnormal muscle tone in at least one extremity and abnormal control of movement and posture. Hearing status was obtained by parental history, and hearing impairment was confirmed by audiologic testing. A history of eye examinations and procedures since initial discharge was obtained, and a standard eye examination was completed. Blindness was defined as bilateral corrected vision of worse than 20/200.

Development was assessed with the Bayley II in period 1 and the Bayley III in period 2. Both tests have a mean score of 100 ± 15. A score <70 (>2 SD below the mean) indicates significant delay, and a score of <85 (>1 SD below the mean) indicates at least mild to moderate delay. The Bayley III manual documents the establishment of content, criterion-related and construct validity. A standardization sample of 1700 children for the Bayley III was reported to be representative of the 2000 US Census population survey data for parent education, ethnicity, and geographic location. Children with a history of PT birth or cognitive, physical, or behavioral issues