

Underestimation of Developmental Delay by the New Bayley-III Scale

Peter J. Anderson, PhD; Cinzia R. De Luca, PhD; Esther Hutchinson, DPsych; Gehan Roberts, PhD; Lex W. Doyle, MD, FRACP; and the Victorian Infant Collaborative Group

Objective: To assess the ability of the third edition of the Bayley Scales of Infant and Toddler Development (Bayley-III) to detect developmental delay in 2-year-old children who were extremely preterm and those carried to term.

Design: Prospective cohort study.

Setting: The state of Victoria, Australia.

Participants: Subjects were consecutive surviving children who were born either at less than 28 weeks' gestational age (extremely preterm) or with less than 1000 g birth weight (extremely low-birth-weight; n=221) in the state of Victoria, Australia, in 2005 and randomly selected controls who were both carried to term and of normal birth weight (n=220).

Main Outcome Measure: Children were assessed by psychologists blinded to knowledge of group at 2 years of age, corrected for prematurity with the new Bayley-III scale.

Results: Follow-up rates of both cohorts were high (>92%). Mean values for all composite and subtest scores for the extremely preterm/extremely low-birth-weight group were significantly below those of the control group ($P < .001$), with the magnitude of all group differences being in excess of two-thirds SD. Mean values for the extremely preterm/extremely low-birth-weight group approached the normative mean, but in contrast, the mean values for the control group were higher than expected, with composite scores being between 0.55 and 1.23 SD above the normative mean. Proportions of children with developmental delay were grossly underestimated using the reference values, but were within the expected range when computed relative to the mean (standard deviation) for the controls.

Conclusion: The Bayley-III scale seriously underestimates developmental delay in 2-year-old Australian children.

Arch Pediatr Adolesc Med. 2010;164(4):352-356

STANDARDIZED DEVELOPMENTAL assessments are important in the early detection of developmental delay in children, determining eligibility for early intervention programs and the evaluation of perinatal, neonatal, and infant treatments.¹ For high-risk infants such as those born very (<32 weeks) or extremely (<28 weeks) preterm, close monitoring using developmental screeners or standardized developmental assessments should be standard practice.

For editorial comment see page 391

While there is no criterion standard for determining developmental delay,^{1,2} the Bayley Scales of Infant Development (BSID)³ and its revisions^{4,5} are the most widely reported measures. The second edition of the Bayley scales (BSID-II), in particular, has been used in many studies to determine rates of developmental delay in

very preterm children and⁶⁻⁸ perinatal factors associated with poor outcome⁹⁻¹⁵ and as an outcome measure in perinatal randomized controlled trials.¹⁶⁻²¹ The BSID-II has also been applied in studies involving other high-risk conditions such as severe combined immunodeficiency,²² human immunodeficiency virus,²³ prenatal cocaine exposure,²⁴ cerebral palsy,²⁵ neurotoxin exposure,^{26,27} gastroschisis,²⁸ and Prader-Willi syndrome.²⁹

The primary scales from the BSID and BSID-II are the Mental Developmental Index (MDI) and the Psychomotor Developmental Index (PDI). In brief, the MDI evaluates early cognitive and language development, while the PDI evaluates early fine and gross motor development. The broad natures of both the MDI and PDI are the main limitations of the BSID and BSID-II.¹ For example, low MDI scores may reflect a specific delay in communication skills, cognitive abilities, or both. The third edition of the Bayley scales (Bayley-III) attempts to address this limitation by refining the measure to include separate com-

Author Affiliations: Murdoch Childrens Research Institute (Drs Anderson and Doyle); University of Melbourne (Drs Anderson, De Luca, Hutchinson, and Doyle); Royal Women's Hospital (Drs De Luca, Hutchinson, Roberts, and Doyle); and the Royal Children's Hospital (Dr Roberts), Melbourne, Australia.

Group Information: The Victorian Infant Collaborative Group members and their author affiliations are listed at the end of this article.

posite scores for cognitive, language, and motor domains. In addition, scale scores can be calculated to assess receptive communication, expressive communication, and fine and gross motor development. Parent-report questionnaires are incorporated into the Bayley-III to assess social-emotional and adaptive behavior. Thus, the structure of the new Bayley-III has the potential to provide more clinically useful information relating to early development, improving our capacity to discriminate specific developmental problems and helping to target early intervention programs to more specific areas of weakness. From a research perspective, the Bayley-III may improve understanding of early development in high-risk populations and may be a more sensitive outcome measure for clinical trials.

To date, few published studies have used the Bayley-III, and the original enthusiasm for this measure may have waned, with many clinicians suggesting that it overestimates development and, as such, underestimates delay. This article will examine this issue by contrasting developmental scores and rates of delay in a large regional cohort of 2-year-olds who are extremely preterm/extremely low-birth-weight and those carried to term who were born in 2005.

METHODS

SUBJECTS

The extremely preterm/extremely low-birth-weight (EP/ELBW) group comprised all children born at fewer than 28 completed weeks of gestation or with birth weights of less than 1000 g born in the state of Victoria in 2005 who survived to 2 years of age. Gestational age was determined by the best obstetric estimate, based on fetal ultrasound, before 20 weeks in most cases. The control group participants were born at 37 weeks' gestation or later and weighed more than 2499 g. They were randomly selected from each maternity unit associated with the 3 level-III perinatal centers in the state, stratified to balance with extremely preterm survivors for sex, mother's health insurance status, and the language spoken primarily in her country of birth (English or other).

DEVELOPMENTAL ASSESSMENT AT 2 YEARS OF AGE

Development was assessed in survivors at 2 years of age, corrected for prematurity, using the Bayley-III scale. Blinded psychologists administered the Cognitive, Language, and Motor scales but not the Social-Emotional or Adaptive Behavior scales. The Cognitive scale assesses abilities such as sensorimotor development, exploration and manipulation, object relatedness, concept formation, memory, and simple problem solving. The Language scale consists of Receptive Communication (verbal comprehension, vocabulary) and Expressive Communication (babbling, gesturing, and utterances) subtests, while the Motor scale consists of Fine Motor (grasping, perceptual-motor integration, motor planning, and speed) and Gross Motor (sitting, standing, locomotion, and balance) subtests.

The Composite scores for the Cognitive, Language, and Motor scales are age-standardized with a mean (SD) score of 100 (15). The Receptive Communication, Expressive Communication, Fine Motor, and Gross Motor subtest scores are age-standardized with a mean (SD) score of 10 (3). Percentile ranks, developmental age

equivalents, and growth scores are not reported. The standardization sample for the Bayley-III comprised 1700 children divided across 17 age bands from 1 to 42 months, with 100 children in each age band. The sample was reported to be representative of the 2000 US Bureau of the Census population survey data in terms of parent education, ethnicity, and geographic region. The original standardization sample included only typically developing children carried to term but later, children with cognitive, physical, and behavioral issues were added to constitute approximately 10% of the total sample.

Children were also assessed by blinded pediatricians for neurosensory impairments including cerebral palsy (CP), blindness (visual acuity <20/200 in the better eye), and deafness (hearing loss requiring amplification, or worse). The criteria for the diagnosis of CP included abnormal tone and delays in motor control and function.

Developmental delay was calculated according to (1) the Bayley-III norms, and (2) the control group mean (standard deviation). Mild cognitive/language/motor delay comprised a score on the relevant composite scale from -2 SD to less than -1 SD; moderate delay, from -3 SD to less than -2 SD; and severe delay, a score of less than -3 SD. Children who were unable to complete psychological testing because of severe developmental delay were assigned a score of -4 SD.

Data were analyzed using SPSS for Windows version 17.0 (SPSS Inc, Chicago, Illinois). Means were contrasted by mean difference and 95% confidence intervals, and by linear regression analysis to adjust for confounding variables (family structure and maternal education). Analyses were also performed excluding children with neurological impairments (CP, blindness, or deafness). Rates of impairment between groups were compared by χ^2 analysis or Fisher exact test with small cell sizes. $P < .05$ were statistically significant.

The Research and Ethics Committees at the Royal Women's Hospital, Mercy Hospital for Women, and Monash Medical Centre, Melbourne, Australia, approved this follow-up study. Written informed consent was obtained from parents of controls carried to term. Follow-up was considered routine clinical care for the very preterm infants.

RESULTS

The EP/ELBW group comprised 221 survivors at 2 years' corrected age, of whom 211 participated in the developmental assessment (95% retention rate). The control group comprised 220 survivors aged 2 years, of whom 202 participated in this study (92% retention rate). The perinatal and demographic characteristics of the 2 groups are displayed in **Table 1**. The EP/ELBW children were less likely to be from intact families at 2 years of age ($\chi^2 = 12.2$; $P < .001$), and their mothers were less likely to have completed secondary school ($\chi^2 = 15.5$; $P < .001$). The rate of CP was elevated in the EP/ELBW group (9.0% [19 of 211] vs 0% [0 of 202]; $P < .001$, Fisher exact test), but the rate of deafness (1.9% [4 of 211] vs 0.5% [1 of 202]; $P = .37$, Fisher exact test) was low and did not differ between groups. No children were blind in either group. The groups did not differ regarding the corrected age at assessment (EP/ELBW mean [SD], 24.4 [2.5] months; control mean [SD], 24.2 [1.7] months; mean difference, 0.1; 95% confidence interval, -0.3 to 0.6).

Table 2 lists the descriptive statistics and mean group differences for the Bayley-III composite and subtest scores for the EP/ELBW and control groups. The means for all composite and subtest scores for the EP/ELBW group were sig-

nificantly lower than those of the control group ($P < .001$), with the magnitude of all group differences being in excess of two-thirds SD. However, it is important to note that the means for the EP/ELBW group approached the normative mean and were within the reference ("average") range. In contrast, the means for the control group were higher than expected, with the composite scores being between 0.55 and 1.23 SD above the normative mean. Analyses were repeated, adjusting for family structure and maternal education. While maternal education was a significant predictor of these developmental outcomes, the mean group differences remained substantial and statistically significant ($P < .001$). The magnitude of the mean group differences declined marginally when children with CP and/or deafness were excluded, although no statistical conclu-

sions were altered ($P < .001$), and all group differences remained in excess of 0.5 SD.

The main purpose of the Bayley-III is to detect developmental delay. The rates of mild, moderate, and severe delay determined according to reference value and compared with controls are presented in **Table 3**. Using normative criteria, the proportions of children in the EP/ELBW group with cognitive, language, and motor delay were only 13%, 21%, and 16%, respectively. The rates for the control group were well below those expected for normally distributed data: 13.6%, 2.0%, and 0.3%, for mild, moderate, and severe developmental delay, respectively. Furthermore, the rate of children in the total cohort with moderate to severe delay was minimal.

When delay was calculated on the basis of the control distribution, the rates rose considerably and were more in line with expectations (Table 3). Using this approach, one-third of the EP/ELBW group exhibited cognitive delay, and even higher proportions had language and motor delay. The proportion of children in the EP/ELBW group with moderate to severe delay was consistent with clinical impressions and previous studies. For the control group, the rate of delay varied from 12% for cognitive development to 17% for motor development.

Table 1. Perinatal and Demographic Characteristics of the EP/ELBW and Control Groups

Characteristic	No./Total (%)	
	EP/ELBW (n=211)	Controls (n=202)
Perinatal		
Gestational age, mean (SD), wk	26.8 (1.9)	39.7 (1.3)
Birthweight, mean (SD), g	868 (178)	3586 (485)
Male	95 (45)	91 (45)
Multiple births	48 (23)	2 (1)
Antenatal corticosteroids	186 (88)	1 (0.5)
Bronchopulmonary dysplasia ^a	102 (48)	0
Postnatal corticosteroids	40 (19)	0
Proven infection ^b	96 (46)	0
Patent ductus arteriosus	141 (67)	0
Grade 3 or 4 IVH	16 (8)	0
Cystic PVL	6 (3)	0
Demographic		
Intact family	185 (88)	197 (98)
Completed high school	124/189 (66)	163/195 (84)
English speaking ^c	170 (81)	167 (83)
Private health insurance	52 (25)	52 (26)

Abbreviations: EP/ELBW, extremely preterm (<28 weeks' gestational age) and/or extremely low birthweight (<1000 g); IVH, intraventricular hemorrhage determined by cranial ultrasound; PVL, cystic periventricular leukomalacia determined by cranial ultrasound.

^aDefined as oxygen dependency at 36 weeks' gestational age.

^bDefined as definite pathogen from normally sterile site or a less virulent organism in association with inflammatory markers.

^cPrimary language spoken at home.

COMMENT

The Bayley-III is currently the most commonly applied measurement tool for assessing early development both in clinical practice and research settings but, to date, limited evidence exists supporting its construct and predictive validity. Our study used the Bayley-III to assess the developmental profile of a geographic cohort of EP/ELBW 2-year-olds and a randomly selected control group. Our findings were contrary to expectations in that the rate of developmental delay for the EP/ELBW group was well below that reported previously^{6,8,9,30} and the rate of delay in the control group was negligible.

Possible explanations for these findings include (1) the Bayley-III's overestimation of developmental outcomes in 2-year-olds and, as such, the underestimation of developmental delay; (2) substantial improvement in developmental outcomes for EP/ELBW children and the recruitment of a high-achieving control group; and (3) systematic error in administration and/or scoring. We are

Table 2. Group Differences on Bayley-III Scale

Variable	Mean (SD)		Mean Difference (95% CI)		
	EP/ELBW	Controls	Unadjusted	Adjusted ^a	Adjusted ^b
Cognitive composite	96.9 (13.8)	108.9 (14.3)	12.0 (9.2-14.8)	10.1 (7.3-13.0)	8.9 (6.1-11.7)
Language composite	94.2 (16.8)	108.2 (14.8)	14.0 (10.7-17.3)	12.2 (8.8-15.6)	10.2 (6.9-13.5)
Receptive communication	9.3 (2.9)	11.6 (3.0)	2.2 (1.6-2.8)	2.1 (1.4-2.7)	1.7 (1.1-2.4)
Expressive communication	8.9 (2.7)	11.2 (2.5)	2.3 (1.7-2.8)	2.1 (1.5-2.6)	1.8 (1.2-2.3)
Motor composite	100.4 (19.6)	118.4 (16.7)	18.0 (14.4-21.6)	16.2 (12.4-19.9)	13.5 (9.9-17.1)
Fine motor	11.3 (3.3)	13.9 (2.9)	2.6 (2.0-3.2)	2.4 (1.8-3.1)	2.1 (1.4-2.7)
Gross motor	9.1 (3.3)	12.1 (3.6)	3.0 (2.3-3.7)	2.9 (2.1-3.6)	2.4 (1.7-3.1)

Abbreviation: Bayley-III, Bayley Scales of Infant and Toddler Development, third edition; CI, confidence interval; EP/ELBW, extremely preterm (<28 weeks gestational age) and/or extremely low birthweight (<1000 g).

^aMeans adjusted for maternal education and family structure.

^bChildren with cerebral palsy and/or deafness excluded and means adjusted for maternal education and family structure.

confident that our findings are not owing to systematic error in administration/scoring, as our psychologists are experienced in conducting developmental assessments with the Bayley scales and all completed the accredited training program for the Bayley-III. The first 2 possible explanations, however, have important implications. We propose that the first explanation is the more likely and that standardized scores of the Bayley-III 2-year-old children underestimate developmental delay and need to be interpreted with great caution. This premise is supported by the finding that the means for the control group for all Bayley-III scales were substantially above the standardized mean, whereas a previous control sample recruited by our group 8 years earlier, assessed using the Bayley-II, had a mean (SD) MDI of 99 (15.4), indistinguishable from the expected mean value of 100. We used the same procedures to recruit the 2 control groups, and there have been no substantial demographic changes during such a short period to suggest that the control groups might be systematically different between eras. Thus, it is highly unlikely that these findings are owing to a high-achieving control group.

Furthermore, we doubt that the higher-than-expected standard scores of the EP/ELBW cohort reflect improved outcome, as the rate of delay judged according to the control group mirrors previous research. Most previous studies examining developmental outcomes in very preterm cohorts have used the BSID-II. In extremely preterm children, the rates of developmental delay determined using BSID-II reference values in cohorts born in the 1990s are high. For example, in cohorts of children born earlier than 25 weeks' gestation, Hintz et al⁸ reported rates of moderate to severe cognitive delay ranging from 40% to 47% at 18 to 22 months' corrected age, while moderate to severe motor delay ranged from 31% to 32%. The EPICure study assessed a geographic cohort of children with gestational ages of fewer than 26 weeks at 30 months, corrected, and based on MDI/PDI reference values, classified 64% of their cohort as delayed (34%, mild; 11%, moderate; 19%, severe).³⁰ Hack et al⁹ have also reported high rates of mild to severe cognitive (68%) and motor (71%) delay in a hospital-based ELBW cohort born from 1992 to 1995. In cohorts from the state of Victoria with gestational ages of less than 28 weeks, the rates of mild, moderate, and severe developmental delay defined by the MDI relative to the mean (standard deviation) for randomly selected controls were 23%, 11%, and 7%, respectively, for those born in 1991 to 1992, and 22%, 9%, and 15%, respectively, for those born in 1997.³¹ As expected, lower rates of delay are reported in cohorts that include more mature infants. In a recently described New Zealand cohort of children with gestational ages of fewer than 33 weeks born from 2001 to 2002, one-third exhibited cognitive delay and 30% exhibited motor delay.⁶ Given previous developmental studies of EP/ELBW children, we had expected rates of overall delay in the 40% to 45% range, consistent with what we observed when delay was based on our control group, but much lower than when delay was based on Bayley-III reference values.

The structural differences between the Bayley-III and BSID-II mean that the scale scores from the 2 tests are

Table 3. Rates of Cognitive, Language, and Motor Delay in EP/VLBW and Control Groups According to the Bayley-III References and Control Distribution

Variable	Percentage			
	Bayley-III Reference		Control Distribution	
	EP/ELBW	Controls	EP/ELBW	Controls
Cognitive delay				
Mild	10	1	20	11
Moderate	2	0	10	1
Severe	1	0	3	0
Total	13	1	33	12
Language delay				
Mild	14	4	34	14
Moderate	5	0	7	1
Severe	2	0	2	0
Total	21	4	43	15
Motor delay				
Mild	8	2	30	14
Moderate	5	0	10	3
Severe	3	0	8	0
Total	16	2	48	17

Abbreviations: Bayley-III, Bayley Scales of Infant and Toddler Development, third edition; EP/ELBW, extremely preterm (<28 weeks gestational age) and/or extremely low birth weight (<1000 g).

not comparable, and direct comparisons with earlier studies that have used the BSID-II are problematic. Theoretically, rates of delay or impairment should increase rather than decrease with the introduction of new standardized measures such as the Bayley-III owing to the creeping phenomena of developmental/intelligence quotient scores over time, often referred to as the Flynn effect.^{2,32} For example, we observed an increased sensitivity in detecting developmental delay when the BSID-II replaced the original BSID.⁷

One limitation of the current study is that our observations are restricted to reference values for 2-year-old children. Further research is needed to assess the appropriateness of reference values in other age bands; we therefore stress that our results should not be extrapolated to other ages prior to receiving the results of such studies. We recognize that there are cultural and other differences between Australia and the United States, where the Bayley-III was standardized; however, this was not an issue for previous Australian cohorts using the BSID-II, which was also standardized in the United States.

In conclusion, the Bayley-III seriously overestimated the developmental progress of 2-year-old Australian children. Given the extent of the overestimation that we observed, we have similar reservations regarding the Bayley-III's sensitivity to detect developmentally delayed children in other countries including the United States, Canada, and England but clearly, further research is needed to confirm our suspicions. Also, the appropriateness of the measure and its reference values for children in other age bands needs to be studied. Our findings have important implications for clinical services, follow-up programs, and clinical trials that rely on the Bayley-III for the assessment of

developmental delay, and we recommend caution in the interpretation of Bayley-III scores for high-risk children in the absence of appropriate control groups.

Accepted for Publication: September 23, 2009.

Corresponding Author: Peter Anderson, PhD, Victorian Infant Brain Studies, Royal Children's Hospital, Flemington Rd, Parkville, Victoria, Australia 3052 (peter.anderson@mcri.edu.au).

Author Contributions: All authors had access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. *Study concept and design:* Anderson, De Luca, Hutchinson, Roberts, and Doyle. *Acquisition of data:* Anderson, Hutchinson, and Roberts. *Analysis and interpretation of data:* Anderson, De Luca, Hutchinson, and Doyle. *Drafting of the manuscript:* Anderson and Hutchinson. *Critical revision of the manuscript for important intellectual content:* Anderson, De Luca, Hutchinson, Roberts, and Doyle. *Statistical analysis:* Anderson and Doyle. *Obtained funding:* Anderson. *Administrative, technical, and material support:* Anderson, De Luca, Hutchinson, and Roberts. *Study supervision:* Anderson.

Victorian Infant Collaborative Group: Catherine Callanan, RN, Noni Davis, FRACP, Julieanne Duff, FRACP, Elaine Kelly, MA, Marion McDonald, RN, Michael Stewart, FRACP, Linh Ung, BSc, Royal Women's Hospital; Elaine Kelly, MA, Gillian Opie, FRACP, Andrew Watkins, FRACP, Amanda Williamson, MA, Heather Woods, RN, Mercy Hospital for Women; Elizabeth Carse, FRACP, Margaret P. Charlton, MEd(Psych), PhD, Marie Hayes, RN, Monash Medical Center; Rod Hunt, PhD, FRACP, Michael Stewart, FRACP, Royal Children's Hospital, Melbourne, Australia.

Financial Disclosure: None reported.

Funding/Support: This study was supported in part by a project grant 454413 from the National Health and Medical Research Council, Australia.

REFERENCES

1. Johnson S, Marlow N. Developmental screen or developmental testing? *Early Hum Dev.* 2006;82(3):173-183.
2. Aylward GP. Developmental screening and assessment: what are we thinking? *J Dev Behav Pediatr.* 2009;30(2):169-173.
3. Bayley N. *The Bayley Scales of Infant Development.* San Antonio, TX: The Psychological Corporation; 1969.
4. Bayley N. *The Bayley Scales of Infant Development-II.* San Antonio, TX: The Psychological Corporation; 1993.
5. Bayley N. *Bayley Scales of Infant and Toddler Development.* San Antonio, TX: The Psychological Corporation; 2006.
6. Darlow BA, Horwood LJ, Wynn-Williams MB, Mogridge RNN, Austin NC. Admissions of all gestations to a regional neonatal unit versus controls: 2-year outcome. *J Paediatr Child Health.* 2009;45(4):187-193.
7. Doyle LW; Victorian Infant Collaborative Study Group. Evaluation of neonatal intensive care for extremely low birth weight infants in Victoria over two decades I: effectiveness. *Pediatrics.* 2004;113(3 pt 1):505-509.
8. Hintz SR, Kendrick DE, Vohr BR, Poole WK, Higgins RD; National Institute of Child Health and Human Development Neonatal Research Network. Changes in neurodevelopmental outcomes at 18 to 22 months' corrected age among infants of less than 25 weeks' gestational age born in 1993-1999. *Pediatrics.* 2005;115(6):1645-1651.
9. Hack M, Wilson-Costello D, Friedman H, Taylor GH, Schluchter M, Fanaroff AA. Neurodevelopment and predictors of outcomes of children with birth weights of less than 1000 g: 1992-1995. *Arch Pediatr Adolesc Med.* 2000;154(7):725-731.
10. Jeng SF, Hsu CH, Tsao PN, et al. Bronchopulmonary dysplasia predicts adverse developmental and clinical outcomes in very-low-birthweight infants. *Dev Med Child Neurol.* 2008;50(1):51-57.
11. Kiechl-Kohlendorfer U, Ralser E, Pupp Peglow U, Reiter G, Trawöger R. Adverse neurodevelopmental outcome in preterm infants: risk factor profiles for different gestational ages. *Acta Paediatr.* 2009;98(5):792-796.
12. Miller SP, Ferriero DM, Leonard C, et al. Early brain injury in premature newborns detected with magnetic resonance imaging is associated with adverse early neurodevelopmental outcome. *J Pediatr.* 2005;147(5):609-616.
13. O'Shea TM, Kuban KCK, Allred EN, et al; Extremely Low Gestational Age Newborns Study Investigators. Neonatal cranial ultrasound lesions and developmental delays at 2 years of age among extremely low gestational age children. *Pediatrics.* 2008;122(3):e662-e669.
14. Shah DK, Doyle LW, Anderson PJ, et al. Adverse neurodevelopment in preterm infants with postnatal sepsis or necrotizing enterocolitis is mediated by white matter abnormalities on magnetic resonance imaging at term. *J Pediatr.* 2008;153(2):170-175.e1.
15. Wood NS, Costeloe K, Gibson AT, Hennessy EM, Marlow N, Wilkinson AR. The EPICure study: associations and antecedents of neurological and developmental disability at 30 months of age following extremely preterm birth. *Arch Dis Child Fetal Neonatal Ed.* 2005;90(2):f134-f140.
16. Kaarens PI, Ronning JA, Tunby J, Nordhov SM, Ulvund SE, Dahl LB. A randomized controlled trial of an early intervention program in low birth weight children: outcome at 2 years. *Early Hum Dev.* 2008;84(3):201-209.
17. Maguire CM, Walther FJ, van Zvieten PHT, Le Cessie S, Wit JM, Veen S. Follow-up outcomes at 1 and 2 years of infants born less than 32 weeks after newborn individualized developmental care and assessment program. *Pediatrics.* 2009;123(4):1081-1087.
18. Mestan KKL, Marks JD, Hecox K, Huo D, Schreiber MD. Neurodevelopmental outcomes of premature infants treated with inhaled nitric oxide. *N Engl J Med.* 2005;353(1):23-32.
19. O'Shea TM, Nageswaran S, Hiatt DC, et al. Follow-up care for infants with chronic lung disease: a randomized comparison of community and center-based models. *Pediatrics.* 2007;119(4):e947-e957.
20. Schmidt B, Roberts RS, Davis P, et al; Caffeine for Apnea of Prematurity Trial Group. Long-term effects of caffeine therapy for apnea of prematurity. *N Engl J Med.* 2007;357(19):1893-1902.
21. Tan M, Abernethy L, Cooke R. Improving head growth in preterm infants: a randomized controlled trial II: MRI and developmental outcomes in the first year. *Arch Dis Child Fetal Neonatal Ed.* 2008;93(5):F342-F346.
22. Lin M, Epport K, Azen C, Parkman R, Kohn DB, Shah AJ. Long-term neurocognitive function of pediatric patients with severe combined immune deficiency (scid): pre- and post-hematopoietic stem cell transplant (HSCT). *J Clin Immunol.* 2009;29(2):231-237.
23. Mekmullica J, Brouwers P, Charurat M, et al. Early immunological predictors of neurodevelopmental outcomes in HIV-infected children. *Clin Infect Dis.* 2009;48(3):338-346.
24. Richardson GA, Goldschmidt L, Willford J. The effects of prenatal cocaine use on infant development. *Neurotoxicol Teratol.* 2008;30(2):96-106.
25. Enkelaar L, Ketelaar M, Gorter JW. Association between motor and mental functioning in toddlers with cerebral palsy. *Dev Neurorehabil.* 2008;11(4):276-282.
26. Davidson PW, Strain JJ, Myers GJ, et al. Neurodevelopmental effects of maternal nutritional status and exposure to methylmercury from eating fish during pregnancy [published online ahead of print June 11, 2008]. *Neurotoxicology.* 2008;29(5):767-775.
27. Tofail F, Vahter M, Hamadani JD, et al. Effect of arsenic exposure during pregnancy on infant development at 7 months in rural Matlab, Bangladesh [published online ahead of print October 24, 2008]. *Environ Health Perspect.* 2009;117(2):288-293.
28. South AP, Marshall DD, Bose CL, Laughon MM. Growth and neurodevelopment at 16 to 24 months of age for infants born with gastroschisis [published online ahead of print July 10, 2008]. *J Perinatol.* 2008;28(10):702-706.
29. Festen DAM, Wevers M, Lindgren AC, et al. Mental and motor development before and during growth hormone treatment in infants and toddlers with Prader-Willi syndrome [published online ahead of print November 19, 2007]. *Clin Endocrinol (Oxf).* 2008;68(6):919-925.
30. Wood NS, Marlow N, Costeloe K, Gibson AT, Wilkinson AR. Neurologic and developmental disability after extremely preterm birth. *N Engl J Med.* 2000;343(6):378-384.
31. Doyle LW; Victorian Infant Collaborative Study Group. Neonatal intensive care at borderline viability: is it worth it? *Early Hum Dev.* 2004;80(2):103-113.
32. Flynn J. Searching for justice: the discovery of IQ gains over time. *Am Psychol.* 1999;54(1):5-20.