SPECIAL FEATURE

Prognostic neurodevelopmental testing of preterm infants: do we need to change the paradigm?

HW Kilbride, GP Aylward, LW Doyle, LT Singer and J Lantos


INTRODUCTION

Modern neonatal intensive care began in the 1960s with efforts to provide long-term positive pressure ventilation for infants with respiratory distress. Early efforts to save preterm infants were sometimes successful but often were associated with iatrogenic complications. These ‘therapeutic misadventures’ prompted concern about long-term outcomes for the infants whose lives were being saved.1,2 In fact, providing support for preterm infants at lower gestational ages was considered by some critics to be ‘experimental’.

Concerns about long-term quality of life led to recommendations for routine monitoring of later outcomes for high-risk newborns who survived intensive care.3,4 Assessments for ‘neurodevelopmental impairment’ (NDI) at about 2 years of age is now common, to assess the prevalence of functional disabilities in neonatal intensive care unit (NICU) survivors, and to understand the costs (human and economic) of survival. In 1990, the National Institute for Child Health and Human Development (NICHD) Network promulgated a policy stating that ‘...follow-up should be a standard aspect of all trials involving newborns.’5

The primary reason for such careful attention to follow-up is that serious illness in the perinatal period can result in life-long complications, including functional impairments that become evident later in childhood. Assessments at 2 years of age for these infants provide an early opportunity to determine complex brain function. This, in turn, enables prediction of later (school age to adult) functional outcomes.

There are at least three good reasons to perform such early assessments: (1) the results can help determine which children should receive early intervention or enhanced educational services; (2) the assessments can be used as outcome measures in research protocols to determine whether specific neonatal interventions lead to better results and (3) such information may also be used to inform clinicians and parents about the appropriateness of providing care for certain groups of infants. Thus, the results of these early assessments may have future life and death implications for neonates with similar medical courses. While these three reasons for careful follow-up are compelling, each is more complex than it might first seem.

Follow-up is very important for neonatal care, and neurodevelopmental testing is an integral part of that process. However, the various ways in which results of neurodevelopmental testing are used raise questions about how the results should be interpreted. What are the sensitivity and specificity of early childhood findings for prognosticating later functional outcomes? Should the same model be used for all situations or should measures and outcome definitions be altered depending on the purpose of the follow-up? This early testing does not take into account emotional and behavioral disorders, which may influence early test abilities but are not commonly recognized until later ages. Psychosocial issues, which potentially affect functional outcomes more than many perinatal factors, are not generally considered in prognostication models. Measures of adaptation and quality of life have only more recently been given attention and need more study before they can be included for prognostication.

The purpose of this commentary is to examine these issues to better understand limitations and expectations of follow-up and the implications for how we report and use the results.

Testing challenges at early ages

There is no established optimal age for follow-up assessments, but much of the work published by the NICHD Neonatal Research Network has used 18 to 22 months’ corrected age data, and, more recently, 22 to 26 months.6 At 18–24 months, cognitive and motor functions diverge, while language and reasoning skills develop. The direct effect of a perinatal event on outcome is more apparent than at later ages, although environmental effects also become influential. From a practical standpoint, earlier ages allow closer tracking of the families, and offer opportunity for prediction of school age function and possible implementation of intervention strategies, which would not be as useful with later testing. However, the largest component of NDI is comprised of low cognitive scores and such testing in infancy and early childhood is challenging. Intelligence cannot be measured early on because the tests are developmental and only precursors of cognitive function may be evaluated. Infancy and early childhood are periods of qualitative and quantitative developmental change that are rapid and uneven. Testing at this age is strongly affected by child characteristics that include variability in performance, temperament, short attention span, physiological issues and a changing behavioral repertoire. As a result, what is measurable at any given age varies and often does not predict subsequent cognitive performance. This also precludes an absolute ‘gold

1Division of Neonatology, Department of Pediatrics, Children’s Mercy-Kansas City and the University of Missouri-Kansas City School of Medicine, Kansas City, MO, USA; 2Division of Developmental and Behavioral Pediatrics/Psychology, Southern Illinois University School of Medicine, Springfield, IL, USA; 3Research Office, Royal Women’s Hospital, Department of Obstetrics & Gynaecology, University of Melbourne, Clinical Sciences, Murdoch Children’s Research Institute, Parkville, VIC, Australia; 4School of Medicine, Department of Epidemiology and Statistics, Case Western Reserve University, Cleveland, OH, USA and 5Children’s Mercy Bioethics Center, Children’s Mercy-Kansas City and the University of Missouri-Kansas City School of Medicine, Kansas City, MO, USA. Correspondence: Dr HW Kilbride, Section of Neonatal-Perinatal Medicine, Children’s Mercy Hospitals and Clinics, 2401 Gillham Road, Kansas City, MO 64108, USA.
E-mail: kilbride@cmh.edu

This manuscript is based on a symposium presentation at the Pediatric Academic Societies (PAS) meeting in Baltimore, MD, 3 May 2016.
Received 7 December 2016; revised 5 January 2017; accepted 18 January 2017
standard' in developmental assessment; instead we have reference standards.

The behavioral repertoire of an infant and toddler evolves from the neurologic to the motor, to sensorimotor and then to cognitive. Early in development, motor and cognitive functions are intertwined to the degree that a motor problem will prevent adequate assessment of cognitive skills. These functions diverge as development progresses, and complex and purposeful behaviors as well as learning can be evaluated. However, many components of early infant tests are 'canalized,' meaning these are 'pre-wired,' basic sensorimotor skills that are less susceptible to the effects of disruption and damage. Conversely, complex, integrated neural networks that underlie later intelligence are more easily affected and the true impact may not be apparent until later demands are placed on behaviors mediated by these networks. It is for this reason that developmental assessment of NICU graduates is particularly difficult.

The Bayley Scales are typically the reference standard for infant and toddler assessment. The test has changed in format from the first two versions, namely, Bayley Scales' and the Bayley Scales of Infant Development (Bayley-II) to the current Bayley Scale of Infant and Toddler Development (Bayley-III). It is challenging to understand how to use developmental quotients to predict cognitive outcomes, but these testing changes have added to the complexity and have made longitudinal comparisons even more problematic. A Mental Developmental Index (MDI) and a Psychomotor Developmental Index (PDI) were derived in the first two versions, but, cognitive, language (receptive and expressive communication) and motor (fine and gross motor) subscales replaced the MDI/PDI on the Bayley-III. Essentially, the MDI was split into the Cognitive and Language subscales, while the PDI was transformed into the two Motor subscales.

Comparison of the Bayley Scales to the Bayley-II produced a mean Bayley Scales-II MDI and PDI that were 12 and 7 points lower than the Bayley Scales MDI and PDI, respectively. This trend was reversed in a later comparison of children who were given the Bayley-II and Bayley-III: the Bayley-III Cognitive Composite was 6 points higher than the MDI, while the Motor Composite was 8 points higher than the PDI. Other studies indicated that the Cognitive score was as much as 10 points higher and the Motor 18 points greater in comparison to the MDI and PDI, which raised significant concern. The fact that the Bayley III norms included 10% of 'at-risk' infants and toddlers (the premise was that mixed sampling reflects the diversity found in the normal population) could at least partially explain the inflation in norms. This situation is more problematic if the at-risk children were not evenly distributed across all age ranges of the normative sample.

The Bayley-II/Bayley-III discrepancies result in two problems. In longitudinal studies the higher Bayley-II scores raise concerns when comparing different cohorts. Differences in scores could be attributable to improvements in medical care or other medical, biologic factors, test issues or a combination of the two. Second, clinically, the Bayley-III scores may be inflated and lead to under-identification of children, needing intervention services. The differences between Bayley-II and Bayley III scores are particularly problematic at the lower end of the normal distribution. These disparities are especially important for long-term follow-up studies in which previous results using the earlier scales are combined with more contemporary data.

There have been numerous attempts to make the scores obtained on the Bayley-II and Bayley-III more comparable. These include combining the Bayley-II Cognitive and Language scores into a Composite score, use of nonlinear regression to produce a predicted MDI from Bayley-III scores, application of a least-squares regression to convert Bayley-II to Bayley-III scores, employing a developmental quotient score; and the use of cutoff scores of 80 or even 85, versus 70. At this point, only the last proposal seems feasible. Moving forward, one way to circumvent this problem is to employ controls who would be given the same tests as the children under study. In so doing, differences on the same test between groups would not be attributable to testing issues.

In addition to changes in test content that have made cohort comparisons difficult, even if the process of testing was stable, test norms will change over time. The Flynn effect, as this is known, results in an average test score increasing 0.3 – 0.5 points per year, thereby raising the mean by as much as 5 points per decade, equivalent to one-third of a standard deviation. A test normed 20 years ago therefore could have a mean score that now is 10 points higher.

Another issue that confounds testing is use of adjusted age rather than chronologic age for assessments. Children born preterm arguably may never totally catch up to their full-term peers. It appears that correction for gestational age should be employed for preterm children at 2 years for functional levels or cognitive outcome. There is evidence that at 3 years of age, the difference between adjusted and unadjusted scores in those born 3 or 4 months early is 5 points. Moreover, although the magnitude of difference between scores decreases at older ages, differences still exist and continue to place those born very preterm and younger at a disadvantage, leading some to suggest that scores should continue to be adjusted for research evaluations throughout childhood. For clinical use however, many developmental pediatricians age-adjust only up to 24 months.

Adolescent/adult outcomes of extreme preterm infants

Data on neurodevelopmental outcomes in adolescence/adulthood have been reported from both longitudinal cohort studies and national registers that link perinatal data with subsequent health data. There are many neurodevelopmental outcomes that might be of interest, but the most common include cerebral palsy, blindness, deafness and intellectual impairment (the last being the most prevalent). Cerebral palsy, blindness and deafness are diagnosed in early childhood and seldom change beyond the early school years. In one longitudinal cohort study of 242 20-year-olds born with very low birthweight (VLBW, <1500 g) in the late 1970s from Cleveland, Hack et al. reported that 6% had cerebral palsy, 2% were blind and 1% were deaf, whereas these impairments were present in none of the 233 normal birthweight (>2499 g) controls, except for one who was deaf. In the same cohort study, Hack et al. reported that mean IQ scores were 6 points lower for VLBW males and 3 points lower for VLBW females compared with controls, and tests of academic achievement were similarly lower in the VLBW cohorts compared with controls. Overall, the rates of high-school graduation were lower in VLBW participants (74%) compared with controls (83%;odds ratio 0.6, 95% confidence interval 0.4, 1.0, P = 0.04).

In a more recent cohort of 18-year-olds born in the early 1990s who were <28 weeks' gestational age at birth from Victoria, Australia, brain volumes were smaller in adolescents born preterm compared with term-born controls (mean difference (95% confidence interval) –5.9% (–8.0, –3.7%) for total brain tissue volume). The largest relative differences were noted in the thalamus and hippocampus. The extreme preterm group had lower IQs (–11.9 (–15.4, –8.3)), and spelling (–8.0 (–11.5, –4.6)), math the Bayley (–12.7), verbal (–7.3, –6.9) and word reading (–6.7, –2.1) scores than controls (all P-values <0.001). Volumes of total brain tissue, and other brain tissues and structures correlated positively with IQ and educational skills, a relationship that was similar for both the EP and controls. Total brain tissue volume explained between 20–40% of the IQ and educational outcome differences between preterm and control participants.
In a study of 317,761 male army conscripts born in 1969–1979 linking data obtained from birth with IQ testing at 18 years of age, the odds for a low IQ (score <16 th centile) were 1.93 (95% confidence interval 1.55, 2.41) higher for those born at 26–29 weeks gestation compared with those born at 39–41 weeks; there was a gradual decrease in the odds of low IQ with increasing gestational age.25

Predicting adolescent/adult outcomes from assessments earlier in childhood

Is it possible to predict outcome status at school age or later based on testing in the first 2 years of life? To address this question, most follow-up investigators of extreme prematurity have developed models using a composite metric, which includes cognition, motor and neurosensory measures. These outcomes are usually expressed categorically to provide groupings of infants that define severity of NDI. Children are characterized as impaired based on an abnormal finding in any of these domains. Thus, a child with a severe neurosensory finding such as deafness would be categorized as 'severe NDI' even if the child exhibits 'normal' cognitive and motor findings. The implication is that this designation in early childhood would predict later functional outcomes. However, this approach creates heterogeneous outcomes that may have markedly different functional implications.

Neurosensory findings may be most constant over time, providing the clearest understanding of functional impairment, but even for those findings the impact on quality of life will vary based on availability of support systems and the individual's ability to adapt.26 As mentioned previously, the most frequent component of NDI at 2 years is a developmental score <70. The relationships between early developmental scores and later IQ at school age in preterm survivors are not strong.27,28 The developmental quotients reflect developmental levels at the age of testing, but they do not define function and only loosely correlate with potential cognitive abilities. Hack et al.29 showed that for extremely low birth weight (ELBW) children, a 20-month MDI <70 had a positive predictive value of only 0.37 overall and, among those with normal neurosensory status, only 0.20.27

In general, the ability to predict cognitive outcomes at school age from infancy and preschool ages has been described as a conundrum.29 The elusive nature of estimates of IQ stability may be due to differences in sample selection, data analytic approaches, the presence of appropriate control groups as well as validity of assessment instruments, as discussed earlier. Even in the best of testing circumstances, defining impairment in early childhood is imprecise and is likely to over-estimate level of disability.

Social and biological variables associated with neurodevelopmental outcomes

Motor abilities are considered to be more neurologically based and 'hard-wired' than cognitive outcomes, which are known to be significantly influenced by environmental factors such as socioeconomic status (SES), parental education, and quality of parenting. However, in both domains, biologic and environmental influences continually interact.

Several large studies addressed the question of stability of cognitive outcomes in preschool children. In a cohort of mostly full-term infants, Breslau et al. reported that IQs of urban children declined across early school age, whereas those of suburban children did not. Although predictive of IQ scores at 6 years, birthweight was unrelated to these IQ changes over time indicating the effects of low SES on outcome.30 In another longitudinal study, Ment et al. found that between 3 and 8 years, LBW children improved in IQ; higher maternal education, early intervention and absence of intraventricular hemorrhage predicted improved performance. Forty-two percent of VLBW children experienced no to little change in IQ, whereas 58% showed an increase in IQ across time. Children with intraventricular hemorrhage and central nervous system injury had lower scores initially that declined over time related to environmental factors.31 Overall, neurological factors were associated with negative change and positive environmental factors with improvement.

Specific and sometimes differential effects of risk factors on outcome domains have also been noted in other studies of preterm groups. Three-year MDI scores were negatively affected by severity of neurologic risk in a longitudinal study of VLBW children from 3 to 14 years of age.32 Neurologic risk lowered MDI scores by 10 points, while minority race was associated with an 8-point decrease, independent of low SES which also decreased scores by 4 points. In contrast, motor outcomes at 3 years were unaffected by environmental factors, but were negatively affected by neurologic risk (~14 points) and the presence of bronchopulmonary dysplasia (~12 points). Similar findings were evident in a study comparing preschool ELBW and term children with their full-term siblings.33 In that study, both preterm status and SES had equivalent effects on IQ scores, such that the mean IQ for high SES ELBW children was equivalent to the mean IQ for low SES, term children. Motor scores were negatively affected by preterm status, but not SES. However, this study excluded ELBW infants with severe neurologic impairment.

In a study of all survivors born at <28 weeks' gestation or weighing <1000 g in the state of Victoria during 1991–1992, severe intraventricular hemorrhage and postnatal corticosteroid therapy were the biologic variables most associated with worse outcomes.34 Of social variables, being reared in a multilingual household was disadvantageous early, with SES and maternal education becoming more important for later outcomes. However, the strength of the biological associations on cognitive and academic outcomes mostly equaled or exceeded those of social exposures, even in late adolescence, contrary to what was expected. Thus, these data indicate that severe neurologic injury will have a dominant or ceiling effect on later functioning. In the absence of such a degree of identifiable injury, environmental factors related to SES and parenting become more important determinants of cognitive outcomes than perinatal factors.

Other factors that may disproportionately affect preterm infants' development should also be considered, particularly maternal depressive symptoms, maternal-infant interactions and social supports and coping mechanisms, all of which have been demonstrated to affect cognitive outcomes in healthy term children, and all of which have particular impact for preterm children. Evidence for greater stress and elevated levels of depression and other psychological symptoms for mothers of VLBW infants in the neonatal period is high.35 Ten percent of mothers of VLBW infants report severe levels of psychological distress after infant birth, five-fold the rate for term mothers, of whom one-third had moderate, clinically significant symptoms.36 Better social support differentially reduced the occurrence of these symptoms in the mothers of VLBW infants,37 suggesting that early intervention to ease parenting stress may positively affect child outcomes.

Using follow-up data for counseling parents regarding ELBW decisions

Neonatologists are challenged when asked to counsel women who are prenatally at risk for delivery of an infant at an extreme preterm gestation. Clinicians are advised to use objective information, based on published data from longitudinal follow-up studies or their own institution's NDI outcomes. Some clinicians use a tool made available on line by the NICHD Neonatal Research Network.39 This calculator allows one to quickly determine estimated survival with or without impairment for
extreme preterm infants by gestational age, birth weight, gender, singleton or multiple gestation status and receipt of antenatal corticosteroids (http://www.nichd.nih.gov/about/org/cdbpm/pp/ prog-epbo/epbocase.cfm). These estimated predications are based on data from evaluation of ELBW children who were 18-22 months old and had been born in 1998-2003. Although a caveat accompanying the calculator states ‘these data are not intended to be predictive for individual outcomes’, the tool remains widely used as a source of information on ELBW outcomes. As noted, the basis for these predictions are early-aged assessments of infants born nearly 20 years ago.

These data have severe limitations as a tool for counselling parents of infants who are born now. They are at best only markers of risk and should not be used for clinical purposes to determine rates of impairment for infants born now. The calculator further confounds the issue by presenting data as a combined outcome of ‘death or profound/moderate to severe neurodevelopmental impairment’. Parents are usually less concerned with that outcome, but rather may wish to understand the risk of impairments if the infant was to survive. In that regard, the numbers presented are misleading. For instance, for a 23 week gestation female, the calculator may indicate a 21% rate of survival without profound NDI. However, since the overall survival is only 33%, the actual estimated percentage of survivors who would not have profound impairment requires further calculation and in this case would be 64% (21/33 x 100%).

Recent publications have shown the impact when such data are used to counsel parents regarding appropriateness of resuscitation and NICU care for extreme preterm infants.40,41 In addition to concerns about the validity of these early markers to predict later functional outcomes, some investigators have suggested many parents will not be able to process the information and may be overwhelmed with data when these medical decisions need to be made.42,43 These authors have suggested that parents’ decisions do not depend on physician prediction of medical or developmental outcomes for similar infants. Statistics are difficult to understand for both parents and health providers and the mode of how such data are presented will affect decision making more than the actual percentages.42

The issue of quality of life remains separate from accurate presentation of neurodevelopmental outcome findings. Quality of life has been defined by the World Health Organisation as ‘the individual’s perception of their position in life in the context of culture and value systems in which they live and in relation to their goals, expectations, standards and concerns’. Innovative studies by Saigal et al.45,46 have shown ELBW parents and ELBW children themselves, are more likely to view survivors with disabilities as less impaired than would health providers. In a study of survivors born <1001 g in Ontario, assessed at 23 years of age, self-reported health-related quality of life did not differ substantially between preterm survivors and controls, although their health status was objectively lower.47

It is also apparent that perceptions of quality of life greatly influence how the physician counsels regarding interventions for extreme preterm infants.48 As part of the counseling process, it is important that this divergence between health providers’ and patient families’ views of disabilities be acknowledged. These differences in perception of quality of life could bias presentation of information and may result in decisions that do not reflect parents’ values.

Summary
Longitudinal neurodevelopmental assessments of high-risk newborns remain an important component for evaluating the effectiveness of neonatal care. The tools used in the first 2 years of life assess developmental levels and indicate areas of divergence from expected developmental trajectories. These areas of divergence are particularly important for those at risk who display markers of perinatal neurologic injury. However, using those findings to predict functional disability in later childhood is problematic. Cognitive, neurosensory and severe motor deficits recognized within the first 2 years of life may or may not be accurate predictors of long-term outcomes. The functional limitations of those assessments are not clearly defined. Long-term outcomes are dependent on postnatal environmental factors, family support, genetics and other factors that may contribute to adaption and resilience. When early assessment data are used clinically or used to provide a predications for later functioning, we recommend the results not be dichotomized to define levels of NDI. Instead, we recommend presentation in descriptive form, which would facilitate individualized interventions to address specific findings. Categorization of children based on composite findings should be limited to outcome measurements for research purposes. Providers who counsel families prenatally regarding risk for extreme preterm or other difficult newborn conditions need to fully understand the implications of 24-month neurodevelopmental findings to avoid using terminology that overstates what is known. Finally, further investigation should focus on better understanding of what families want and need to make decisions about their infant born at borderline gestational age so that providers may move beyond data presentation in these settings.

CONFLICT OF INTEREST
The authors declare no conflict of interest.

REFERENCES


OFFPRINT ORDER FORM

Please return this form by email to nyajproofs4@us.nature.com or by fax to +1 646 563 7084.

Journal: Journal of Perinatology

Article Title:

Manuscript Number:

Offprint Orders (see pricing table overleaf)

If you would like to order offprints, please complete the following details ensuring that the delivery address is correctly and clearly entered.

Number of offprints required: ___________________ Total Cost: ___________________

☐ I agree to pay the costs in full and authorize Nature Publishing Group to invoice me for the above amount.

Signature: ___________________ Date: ___________________

Print Name: ___________________ E-mail: ___________________

Delivery Address
Name: ___________________
Address: ___________________

______________________________________________

Post code/Zip code: ___________________
Country: ___________________

Invoice Address (if different)
Name: ___________________
Address: ___________________

______________________________________________

Post code/Zip code: ___________________
Country: ___________________

☐ Purchase order number: ___________________

☐ VAT number: ___________________ (if applicable)

VAT will be added to orders from UK based authors. Offprints are subject to VAT but at the reduced rate of 0%. Local taxes will be added where applicable.

Payment

Payment for both page charges and offprints can be made by bank transfer, bank cheque or credit card after the invoice is issued.
Pricing—Offprints

Authors in the US and Canada will be charged in US dollars; authors in the rest of the world will be invoiced in UK sterling.

<table>
<thead>
<tr>
<th>Number of copies</th>
<th>Number of pages</th>
<th>Add $538 (£355) per 100 copies if offprints contain color illustrations</th>
<th>Each additional 4pp</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1-4</td>
<td>5-8</td>
<td>9-12</td>
</tr>
<tr>
<td>100</td>
<td>$402 (£261)</td>
<td>$565 (£367)</td>
<td>$700 (£454)</td>
</tr>
<tr>
<td>200</td>
<td>$564 (£366)</td>
<td>$857 (£560)</td>
<td>$1,142 (£742)</td>
</tr>
<tr>
<td>300</td>
<td>$700 (£454)</td>
<td>$1,142 (£742)</td>
<td>$1,589 (£1,031)</td>
</tr>
</tbody>
</table>

Prices include postage.

Please note:

1. All offprint orders must be made on this order form. Offprint orders are not acknowledged on receipt.
2. OFFPRINT ORDERS CAN ONLY BE ACCEPTED BEFORE THE JOURNAL IS PRINTED.
3. SUBSEQUENT ORDERS INVOLVE REPRINTING AND REPRINT PRICES ARE SUBSTANTIALLY HIGHER THAN THOSE QUOTED BELOW. NATURE PUBLISHING GROUP DOES NOT ACCEPT RESPONSIBILITY FOR UNPROCESSED ORDERS DUE TO FORMS BEING INCOMPLETE OR WHICH ARRIVE AFTER THE ISSUE DEADLINE.
4. As only one author receives an offprint order/page charge form it is essential that adequate supplies for all authors of a paper be included on this form.
5. If a purchase order number is required to accompany your order, please ensure that the number is clearly written on the offprint order form. A purchase order number must be supplied for orders to be billed to US institutions.
6. Offprints are dispatched within four to eight weeks of publication of the journal.

The Publishers are pleased to help with any enquiries about post-publication reprints.

Special prices apply for quantities greater than 300. For a quote, please contact the reprints department.

USA/Canada/Northern Europe/UK/ROW:
Tel: +1 212-451-5713
E-mail: commercial-reprints@us.nature.com

Southern Europe / Latin America:
Tel: +34 91 447 4643
Fax: +34 91 447 4609
E-mail: reprintsnpqi@macmillan.es

Asia Pacific:
Tel: +61-3-3267-8775
Fax: +61-3-3267-8746
E-mail: commercial-reprints@natureasia.com

India:
Tel: +91 124 288 1053
Fax: +91 124 288 1052
E-mail: commercial-reprints-india@nature.com