Neurodevelopmental Outcome of Preterm Babies of 1999-2009

Abstract:

D Huggard, M Slevin, C Vavasseur
National Maternity Hospital, Holles St, Dublin 2

Abstract

The Bayley scale of infant development is employed as the performance indicator at 2 years corrected gestational age for high risk paediatric groups. We compare neurodevelopmental outcomes in two cohorts of VLBW infants born in 1999 to a cohort born a decade later, while also examining the challenges of direct comparison of modified scales: BSID-II (2nd edition of the scales) with Bayley III. BSID â II was used in the 1999 group and Bayley-III for the 2009 cohort. We demonstrated that over a ten year period there was an improvement in neurodevelopmental scores achieved in VLBW babies. Overall there was almost an 8 point increase in the cognitive scores from the 2009 cohort compared with the 1999 cohort in this time period. The mean motor score increased by 6 points when comparing 1999 and 2009. However we highlight the difficulties in comparing developmental scales, and consider whether Bayley-III overestimates developmental ability?

Introduction

Intact long-term survival is the ultimate aim of the neonatologist. With advancing techniques in the NICU there has been a decrease in perinatal mortality of pretermbabies[1], and over time the emphasis has shifted from improving mortality to morbidity in this vulnerable group. Therefore developmental followup is crucial in the assessment of the quality of outcome. How these babies fare in terms of cognitive & motor skills is clearly important and has become an integral part of childcare[2]. Formal assessment allows us to identify children with developmental delay and thus determine areas for appropriate early intervention services. The Bayley scale of infant and toddler development is the gold standard of assessing high risk paediatric groups. It is an assessment tool to screen for developmental delay in preterm children, and identifies areas of strength or weakness when planning an intervention programme for a child. In 1993 the 1st edition of the scales[3] was revised as the Bayley Scales of Infant Development II (BSID-II)[4] it measured two scores: the Mental Developmental Index (MDI) used to assess cognitive and language skills, and the Psychomotor Developmental index (PDI) for the evaluation of motor development. Although widely used, the format of BSID-II was criticized due to an absence of separate scales to appraise language and cognitive development. In 2006 the scales were restructured and restandardized as the Bayley Scales of Infant and Toddler Development Third edition (Bayley-III)[5]. This led to a separation of the pre-existing MDI into distinct cognitive, receptive and expressive language scales, and the PDI into fine & gross motor categories.

In recent years concerns have been raised with regard to the Bayley-III and the scores being achieved. Several studies[6-8] have demonstrated that there has been be an increase in the cognitive and language scores of the Bayley-III versus MDI scores using the previous version (BSID-II). Indeed it has been reported in the US[7] and Australia[8] that there have been unexpectedly high Bayley-III results studying developmentally normal children and high risk populations. This indicates that the new edition may overestimate developmental ability. There have been a couple of reasons postulated for the discrepancies in scores between the two editions. Firstly, the change in test format means we are no longer comparing like with like across the two scales. Could the increased scores seen be a result of the change in assessment tool rather than advances in neonatology? The second issue concerns the restandardization of the Bayley-III. In the US[6] it is thought that the population demographic changed over time, especially in relation to levels of parental education and ethnicity, between 1988 and 2000 when the two tests were standardized. Furthermore, 10% of the normative population for restandardization had existing conditions that incur a risk of developmental delay (Downs, Cerebral Palsy), however this may lead to an overestimation of abilities when using the assessment to detect those with suboptimal development. Our aim was to compare the cognitive, language, and motor skills, at 2 years CGA, of VLBW infants born before the new millennium to a cohort born a decade later. We also wanted to examine the challenges of direct comparison of modified scales, namely BSID-II with Bayley-III.

Methods

Our study was a retrospective review of neurodevelopmental outcome of a cohort of VLBW infants born 1999, compared to the outcome of a cohort born in 2009. The infants included were neonates admitted to NICU, National Maternity hospital during 1999 and 2009. The babies studied all had birth weights less than 1500g and gestational ages ranging from 24 weeks up to 35 weeks gestation. There was data available for a total of 47 babies from the 1999 cohort, and 88 patients from 2009. The Bayley scale of infant development was used as the performance indicator at 2 years corrected gestational age. The Bayley scale of infant development (BSID-II), measured mental developmental index (MDI) and psychomotor developmental index (PDI) for the 1999 cohort. The Bayley-III composite scores; language, cognition and motor, were used for those infants born in 2009. As the 1999 group received a single MDI score, and those from2009 got separate composite scores for cognitive and language scales, for comparison the mean of the cognitive and language scales was calculated and called the combined Bayley-III score (CB-III). The PDI from 1999 was compared directly with the composite motor score of 2009 as an approximation. The assessments of both cohorts were carried out by a single examiner (MS). Due to deficiencies in the Bayley-III highlighted by some commentators, a previously tested conversion[6] of language and cognitive results from Bayley-III was made to allow approximation and direct comparison with the MDI of BSID-II. Due to the differing scoring classifications between the two scales we decided on a categorization that we felt would optimise comparisons. The degree of delay was categorized and scored as extremely delayed (<70), below average (70-79), average (80-109), above average (>110). Potential confounding variables were also compared. SPSS 18 was used for data analyses, & independent t-test analysis was tested at 95% C.I.

Results

The outcome data was available for 47 infants born 1999 and 88 infants born 2009. The table shows direct comparison of the mean BSID-II MDI 1999 with the average of the combined cognitive and language scores (CB-III) from Bayley-III 2009. The mean MDI 1999 was92.7 (SD 16.9) compared with a CB-III 2009 score of 100.5 (SD 17.8). The Bayley-III cognitive and language scores were 100.7 (SD 16.6), and 100.3 (SD 18.9) respectively. This demonstrates that CB-III 2009 scores were nearly 8 points higher than MDI 1999 scores. However, despite the significant improvement in cognition, it is not significant when cognitive outcome in 2009is assessed by adjusted MDI and compared to 1999 MDI using the Moore et al algorithm[6]. The adjusted score was 93.7 in 2009 compared to MDI of 92.7 in 1999 (p=0.758). A direct comparison of the mean PDI 1999 with the composite motor score from 2009 is also shown. This revealed a mean PDI 1999 score of 96.7 (SD 17.7) and a mean composite motor score 2009of 102.9 (SD 16.9). This reveals an improvement over time in the motor skills category, with scores from 2009 6 points higher than a decade previously.

Figure 1 compares MDI and CB-III outcomes1999 and 2009. The percentage of children scoring in the low average and extremely delayed categories was higher in 1999 compared with 2009. In 1999, the percentage was 23.1%, compared with 13.9% in 2009. Figure 2, the proportion of patients with a score of <70 i.e., extremely delayed (MDI/CB-III and PDI/Composite motor score) in the two cohorts were compared with and without adjustments of CB-III. In 1999 there was 10.6% (5/47) with MDI and PDI in the extremely delayed category. Ten years on in 2009 and this figure fell to 4.5% (4/88). When the cognitive and language scores from Bayley-III are converted to an approximate MDI and the percentages of patients receiving scores <70 in MDI and PDI/Motor were looked at, it revealed a rise in those in the extremely delayed categories from 4.5% to 10.2%. Thus the improvement in extremely delayed category as demonstrated with unadjusted CB-III, disappears completely when CB-III scores are converted to an adjusted MDI.

Table 2 compares potential confounding variables between the two cohorts. Notably the babies born in 2009 tended to be more premature (mean GA 28.4 weeks vs. 29.5 weeks) and consequently had lower birth weights (mean BW 1086g vs.1173g). In 1999 51.1% of the cohort was male; however in 2009 that percentage was 59.1%. Over a decade the numbers outborn increased from 2.1% to 15.9%.There was also an increase in antenatal steroid use, from 53.2% to 90.9%. Statistical analysis was performed on the potential confounders; an ANCOVA with Cohort as dependent variable; MDI/CBIII as independent variable; and BWGT, GA, Inborn, ANS & PNS as covariates. This demonstrated that ANS are a significant

predictor of 'cohort' (p< .001) and MDI/CBIII is still significant (p = .002).When logistic regression analysis is utilised the only predictors that are significant are MDI/CBIII (p= .034), ANS (p< .001) and Inborn (p=.042). Unfortunately data was not available for other potentially significant confounders such as, ventilation, CPAP, chronic lung disease, intraventricular hemorrhage. It is acknowledged that this is a limitation of this study.

Discussion

Using two separate scales we have demonstrated that over a ten year period there was an improvement in neurodevelopmental scores achieved in VLBW babies (Table 1.) Overall there was an improvement of both cognitive and motor scores over a decade. Our study also highlighted that over a decade, there were less children being placed in the below average and extremely delayed categories. However given that we are not comparing like with like by employing two distinct editions of the scales we utilised a tested algorithm[6,10] converting cognitive and language scores from Bayley-III to an MDI equivalent. This allowed us to calculate and compare outcomes between the two scales. Using this formula we saw that there was a difference of only 1 point between the two groups; MDI 1999 and CB-III 2009. Indeed this demonstrates the absence of any significant difference between the 1999 MDI score and 2009 adjusted MDI score. Furthermore the same proportions of infants are placed in the extremely delayed category if the 1999 MDI and 2009 adjusted MDI scores are compared. The algorithm: Predicted MDI = 88.8 â {61.6 x (language score/100)-1 + (0.67 x cognitive score)}.

The inflation of scores using the 3rd edition of the scales has been replicated in other papers[6-8]; Moore et al[6] recommended caution in the interpretation of Bayley-III scores in population studies as the correlation with the previous edition appears worse at lower test score values. As with our study, they found that CB-III values were increasingly higher than MDI at lower scores. This leads us to consider whether the Bayley-III is overestimating developmental ability? or are advances in neonatology the reason for the improvements in scoring? Vohr et al[7] have also advised discretion when utilising the third edition of the Bayley scales. They feel it identifies significantly less children with disability. They go as far as recommending that all extremely low birth weight infants be offered early assistance and follow up at the time of discharge from the neonatal intensive care unit. This raises an important clinical issue. With fewer children placed in the low average and extremely delayed categories going forward, relying on Bayley scores alone may lead to at risk children missing out on timely early intervention services? The weight of literature as discussed would say that interpreting Bayley-III should be done with caution. Indeed, perhaps the cut off for identifying developmental delay using Bayley-III scores should be raised, to avoid missing those who maybe lost to follow up otherwise?

In conclusion, demonstrating improvements in neurodevelopmental outcomes over time is extremely challenging, yet it is these longer term functional outcomes for our preterm infants that are the more important key performance indicators of the quality of our newborn services and as such must warrant greater resources for their assessment.

Correspondence: D Huggard
National Maternity Hospital, Holles St, Dublin 2
Email: dean.huggard@gmail.com

References

1. Stoelhorst GM, Rijken M, Martens SE, Brand R, den Ouden AL, Wit JM, Veen S; Leiden Follow-Up Project on Prematurity. Changes in neonatology: comparison of two cohorts of very preterm infants (gestational age <32 weeks): the Project On Preterm and Small for Gestational Age Infants 1983 and the Leiden Follow-Up Project on Prematurity 1996-1997. Pediatrics. 2005 Feb;115:396-405.

2. Petrou S, Abangma G, Johnson S, Wolke D, Marlow N. Costs and health utilities associated with extremely preterm birth: evidence from the EPICure study. Value Health. 2009 Nov-Dec;12:1124-34.

3. Bayley N. Bayley Scales of Infant Development. New York: Harcourt Assessment; 1969.

4. Bayley N. Manual for the Bayley scales of infant development. 2nd ed. San Antonio (TX): The Psychological Corporation; 1993.

5. Bayley N. Bayley scales of infant and toddler development. 3rd ed. San Antonio (TX): Harcourt Assessment; 2006.

6. Moore T, Johnson S, Haider S, Hennessy E, Marlow N. Relationship between test scores using the second and third editions of the Bayley Scales in extremely preterm children. J Pediatr. 2012 Apr;160:553-8.

7. Vohr BR, Stephens BE, Higgins RD, Bann CM, editors. Are outcomes of extremely preterm infants improving? Impact of Bayley assessment on outcomes. J Pediatr. 2012 Aug;161:222-8.

8. 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 Apr;164:352-6.

9. Acton BV, Biggs WS, Creighton DE, Penner KA, Switzer HN, Thomas JH, Joffe AR, Robertson CM. Overestimating neurodevelopment using the Bayley-III after early complex cardiac surgery. Pediatrics. 2011 Oct;128:e794-800.

10. Lowe JR, Erickson SJ, Schrader R, Duncan AF. Comparison of the Bayley II Mental Developmental Index and the Bayley III Cognitive Scale: are we measuring the same thing? Acta Paediatr. 2012 Feb;101:e55-8.

11. Flynn J. Searching for justice, the discovery of IQ gains over time. American Psychologist 1999;54:5-20.