Preterm Born Children at the Age of 5

Potharst, E.S.

2012

document version
Publisher's PDF, also known as Version of record

Chapter 4

Prediction of cognitive abilities at the age of 5 using developmental follow-up assessments at the age of 2 and 3 in very preterm children

Published as: Potharst, E.S., Houtzager, B.A., Van Sonderen, L., Tamminga, P., Kok, J.H., Last, B.F., Van Wassenaer, A.G. Prediction of cognitive abilities at the age of 5 using developmental follow-up assessments at the age of 2 and 3 in very preterm children.

D*evelopmental Medicine and Child Neurology 2012; 54: 240-6*

ABSTRACT

Objectives

This study investigates prediction of separate cognitive abilities at age 5 by cognitive development at the ages of both 2 and 3, and the agreement between these measurements, in very preterm born children.

Methods

Preterm born children (n = 102, 44 males and 58 females, with a gestational age < 30 weeks and/or birth weight < 1000 grams) were assessed at age 2 and 3 using the second version of the Bayley Scales of Infant Development, Child Behavior Checklist and a neurological examination, and at age 5 the third version of the Wechsler Preschool and Primary Scale of Intelligence.

Results

Cognitive development at age 2 and 3 explained 44% and 57% respectively of full scale intelligence at age five. Adding psychomotor, neurological and behavioural outcomes and perinatal and sociodemographic characteristics to the regression model increased explained variance to 57% and 64% respectively. These percentages were comparable for verbal intelligence. Processing speed quotient and especially performance intelligence were predicted less accurately.

Interpretation

Not all aspects of intelligence are predicted sufficiently by MDI at age 2 and 3. Follow-up of very preterm children until at least age 5 is needed to distinguish between different aspects of cognitive development.

Follow-up (FU) assessments of very preterm (VPT) born children are necessary in order to guarantee continued patient care and evaluation of treatments during the perinatal period.[1] Because FU rates are highest in the first years after the neonatal intensive care unit (NICU) period and because FU is expensive and time consuming, outcomes at 18-24 months of age are frequently reported as an endpoint of neurodevelopmental FU.[2] Content and timing of FU have to be attuned as much as possible to the problems and disabilities often seen in VPT children.[3]

Outcomes of several FU studies suggest that evaluation at 18-24 months should rather be seen as an intermediate point.[4,5] Developmental disabilities can only become fully apparent when the child starts to develop complex functions that can be assessed.[6] Roberts et al.[4] reported low agreement between assessments at 24 months and 8 years, especially on the cognitive domain ($\kappa = 0.11$), in their cohort of extremely preterm born children. Hack et al.[5] also found a poor predictive validity of cognitive development at the age of 20 months in a preterm born sample. Agreement appears to be more optimal in predicting outcome of children with severe disabilities.[7]

From preschool age on, different cognitive abilities can be distinguished.[8] Moreover, studies reveal that besides global cognitive disabilities, specific learning deficits are present in preterm born infants.[6] To our knowledge, the only study that reports on the prediction of specific aspects by early cognitive development is a study by Crowe et al.,[9] in which a more heterogeneous group was studied and older versions of developmental tests were used. Crowe et al concluded that cognitive abilities can only be predicted to a small extent.

This study focuses on the prediction of distinct cognitive abilities at the age of 5 from a viewpoint of early FU in VPT children with a gestational age of < 30 weeks and/or a birth weight (BW) of < 1000 grams who were able to complete an intelligence test. Other studies have already shown that cognitive assessment at toddler age does not predict later cognitive functioning well enough. In the current study we investigate to what extent cognitive functioning at the age of both 2 and 3 predicts not only full scale, but also verbal and performance intelligence and processing speed at age 5, and what the agreement between these measurements is. Because the younger a child is, the harder it is to distinguish between different developmental domains, we hypothesise that domains of development other than cognition at toddler age (psychomotor, neurological and behaviour) will improve the prediction of cognitive development at age 5. It is not clear whether perinatal morbidities already have its full impact on development at toddler age or alternatively, still have an additional impact on outcomes at school age.[10] We therefore aim to explore whether perinatal morbidities improve the prediction of cognitive outcomes at 5 years when development at age 2 or 3 is already accounted for, and which consequences of perinatal events manifest themselves after toddler age. Finally, sociodemographic factors tend to become more important as children grow up;[6] additional effects of these factors will therefore be investigated.

PATIENTS AND METHODS

This is a single centre longitudinal cohort study. In the Netherlands, NICU-care is centralized in only 10 specialized hospitals.

Participants

The study sample consisted of a cohort of 5-year old VPT children, of whom earlier FU data were available. Children born before 30 weeks gestation and/or with birth weights below 1000 grams, who reached the corrected age of five between December 2007 and June 2009, were included. Other inclusion criteria were 1) hospitalization in the NICU of our hospital, 2) the availability of at least one developmental FU score at age 2 and/or 3 and 3) residency in The Netherlands. Exclusion criteria were: 1) participation in one of two other studies[11,12] because of the use of different instruments and different timing of FU, 2) having a genetic syndrome, or 3) being too disabled to be assessed. Disability was defined by a considerable cognitive delay (Mental Developmental Index (MDI) score of below 55 at the age of 2 and/or 3, when age was corrected for prematurity), cerebral palsy with the inability for the child to point to pictures, or severe hearing problems making regular intelligence assessment impossible.

Test protocol

Ethical approval of this study was given by the Medical Ethical Committee of the Academic Medical Centre. Informed consent to the research and publication of results was given by parents of all children. At age 5, the child's intelligence was assessed by a trained child psychologist. Data concerning the child's functioning at 2 and 3 of age (neurological, cognitive and psychomotor as assessed by a trained paediatrician and child psychologist, and behaviour measured with questionnaires completed by parents) were available. For reasons of comparability over time,[4,13] all scores were calculated on the basis of the child's corrected age. Cognitive scores were considered mildly or severely abnormal if more than 1 or 2 standard deviations (SD) below the mean respectively.

Measures

Intelligence at age 5 was assessed using the Dutch standardization of the third version of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-III-NL).[14] Four composite scores were calculated: full scale, verbal, and performance intelligence quotient and processing speed quotient (from now on referred to as FSIQ-5 VIQ-5, PIQ-5 and PSQ-5).

The Dutch standardization of the Bayley Scales of Infant Development Second Edition (BSID-II-NL)[15] was used to assess MDI and the Psychomotor Development Index (PDI) at age 2 and 3. MDI

and PDI at the age of 2 and 3 will be referred to as MDI-2, PDI-2, MDI-3 and PDI-3 respectively. Emotional and behavioural problems were assessed at the age of 2 and 3 years using the Dutch version of the Child Behavior Checklist (CBCL 1 ½-5).[16] The total problem T-score was used.

Neurological development was assessed at age 2 and 3 using the Hempel neurological examination.[17] Neurological development was classified as normal, suspect or abnormal.

Baseline characteristics

Perinatal data were abstracted from an ongoing prospective database used for all infants admitted to our NICU. Level of parental education was used as a measure of socioecomic status. Parents who had the lowest type of college of less are rated 'low level of education' (total years post elementary schooling: < 6.) Parents who graduated form middle level of college were rated with 'middle level of education' (total years post elementary schooling: 6-8). Parents who had highest level college or university were rated with 'high level of education' (total years post elementary schooling: > 8). The combined parental education score was low if one or both parents had low level of education, middle if both parents had middle, and high if one or both parents had high education.

Statistical analysis

Univariate analyses were carried out to study differences in baseline characteristics between participants and excluded children and non-participants. Mean differences between measurement occasions were tested using repeated measures ANOVA and post-hoc paired samples t-tests. The agreement between MDI at the age of 2 and 3 and cognitive measures at the age of 5 was determined with a linearly weighted kappa. Categories were interpreted as proposed by Landis and Koch [18] ($0 < \kappa < 0.2$: slight; $0.2 < \kappa < 0.4$: fair; $0.4 < \kappa < 0.6$: moderate; $0.6 < \kappa < 0.8$: substantial; $0.8 < \kappa < 1$: almost perfect). Hierarchical regression analyses were performed to predict the continuous scores of FSIQ-5, VIQ-5, PIQ-5 and PSQ-5. In each step an independent variable or group of variables was added. If the new model added significantly ($p < .05$) to the explained variance of the dependent variable, the new model was accepted and the next step was analyzed. MDI was entered in the first step. In the following five steps, PDI, neurological examination, behaviour, perinatal and sociodemographic characteristics were added consecutively. Only perinatal characteristics that were significantly associated with the dependent variable in univariate analyses ($p < .10$) were added. Via a backwards procedure, perinatal characteristics were removed from the model until the model was significant with a p-value of < .05, and the individual variables had p-values of < .10. For sociodemographic characteristics the same procedure was followed. Regression analyses were performed separately for the assessments performed at 2 and 3 years of age.

RESULTS

Participants

One hundred-fifty children, born before 30 weeks gestation and/or BW below 1000 grams, turned 5 years old in the period of recruitment. Thirty-seven children were excluded; sixteen participated in another study, nine children had no developmental score available from earlier FU, four children were too disabled, three had a genetic syndrome and five families had moved abroad. One hundred thirteen children remained eligible and were invited. Eleven children were lost to FU or parents refused participation. With respect to baseline characteristics, displayed in Table 1, participants (n = 102) received indomethacin more often than non-participants and excluded children (n = 48, p = .02).

Of 102 participating children, 101 children were assessed at age 2 and 96 at age 3. MDI, PDI, neurological examination, and CBCL scores at the age of 2 were obtained for respectively 100, 100, 102 and 95 children and at the age of 3 for respectively 96, 94, 94 and 90 children. WPPSI-III scores were obtained for all children.

Course of development

Table 2 displays mean scores on developmental tests and the proportions of children with abnormal scores.

Significant differences were found between MDI-2, MDI-3 and FSIQ-5 (p = .000). MDI-3 scores were on average 11 and 9 points higher than MDI-2 and FSIQ-5 respectively (p = .000 and .000).

Table 1 Perinatal and sociodemographic characteristics

	Participants *n* = 102	Excluded children and non-participants *n* = 48	*p*
Perinatal and child characteristics			
Gestational age (wk), mean (SD)	28 5/7 (1 4/7)	28 6/7 (1 5/7)	.58
Birth weight (g), mean (SD)	1040 (253)	1093 (260)	.23
Sex: male (%)	44 (43.1%)	27 (56.3%)	.13
Gestational age ≥ 30 weeks and birth weight < 1000 grams (%)	14 (13.7%)	5 (10.4%)	.57
Very small for gestational age (BW < p 2.3 for gestational age) (%)	15 (14.7%)	8 (16.7%)	.76
Postnatal dexamethasone (%)	4 (3.9%)	4 (8.3%)	.44
Indomethacin for patent ductus (%)	28 (27.5%)	5 (10.4%)	.02*
Oxygen support at 36 weeks pma (%)	15 (14.7%)	5 (10.4%)	.47
Sepsis and meningitis (%)	28 (27.5%)	16 (36.4%)	.46
Necrotizing enterocolitis, stage 2 and 3 (%)	2 (2.0%)	1 (2.1%)	1.00
Periventricular leucomalacia (PVL) 1[b] (%)	5 (4.9%)	4 (8.3%)	.47
Subependymal hemorrhage (%)	25 (24.5%)	13 (27.1%)	.74
Intraventricular hemorrhage 2 (%)	3 (2.9%)	2 (4.2%)	1.00
Intraventricular hemorrhage 3[b] (IVH 3) (%)	3 (2.9%)	0 (-)	.55
Post hemorrhagic hydrocephalus (4 mm > p 97 of Levene curves) (%)	4 (3.9%)	0 (-)	.31
Severely abnormal ultrasound (PVL 2-4, IVH 3-4 and/ or PHH) (%)	4 (3.9%)	0 (-)	.31
Cerebral palsy at the age of 5	5 (4.9%)	4 (10.5%)[c]	.23
Sociodemographic characteristics			
Parental foreign country of birth[d] (%)	31 (30.4%)	-	-
Parental education at child's age of 5			-
High (%)	44 (43.1%)	-	
Low - middle (%)	58 (56.9%)	-	

Data are presented as n(%) or M±SD.

* p < .05

[b] There were no cases of PVL 2-4 or intraparenchymal hemorrhage 4.

[c] Neurological examination was done for 38 of all excluded children and non-participants, either at the age of 5 (12 children) or at younger age. Ten children were not included in the denominator; 7 were lost to follow-up and 3 had a genetic syndrome.

[d] Of parents born outside the Netherlands, 23% was from Suriname, 19% from Turkey, 14% from Marocco, 44% from other countries.

Table 2 Outcome of assessment of cognitive development at 2, 3 and 5 years of age of VPT born children

	Preterm group *n* = 102
Corrected age of 2 years (*n* = 101)	
MDI (BSID-II) (*n* = 100)	
Mean (SD)	91 (18)
Mildly abnormal (%)	20 (20.0%)
Severely abnormal (%)	13 (13.0%)
Corrected age of 3 years (*n* = 96)	
MDI (BSID-II) (*n* = 96)	
Mean (SD)	102 (14)
Mildly abnormal (%)	11 (11.5%)
Severely abnormal (%)	3 (3.1%)
Corrected age of 5 years (*n* = 102)*	
FSIQ (WPPSI-III) (*n* = 102)	
Mean (SD)	93 (17)
Mildly abnormal (%)	16 (15.7%)
Severely abnormal (%)	10 (9.8%)
VIQ (WPPSI-III) (*n* = 102)	
Mean (SD)	95 (17)
Mildly abnormal (%)	21 (20.6%)
Severely abnormal (%)	9 (8.8%)
PIQ (WPPSI-III) (*n* = 102)	
Mean (SD)	93 (15)
Mildly abnormal (%)	19 (18.6%)
Severely abnormal (%)	8 (7.8%)
PSQ (WPPSI-III) (*n* = 102)	
Mean (SD)	93 (18)
Mildly abnormal (%)	19 (18.6%)
Severely abnormal (%)	12 (11.8%)
All (I)Q-scores	
Normal (%)	56 (54.9%)
At least one abnormal (I)Q-score (mild or severe) (%)	46 (45.1%)
At least one severely abnormal (I)Q-score (%)	17 (16.7%)

Data are presented as *n* (%), or *M* ± *SD*.

Scores are considered mildly abnormal when < - 1 SD and severely abnormal when < - 2 SD.

Agreement in cognitive scores

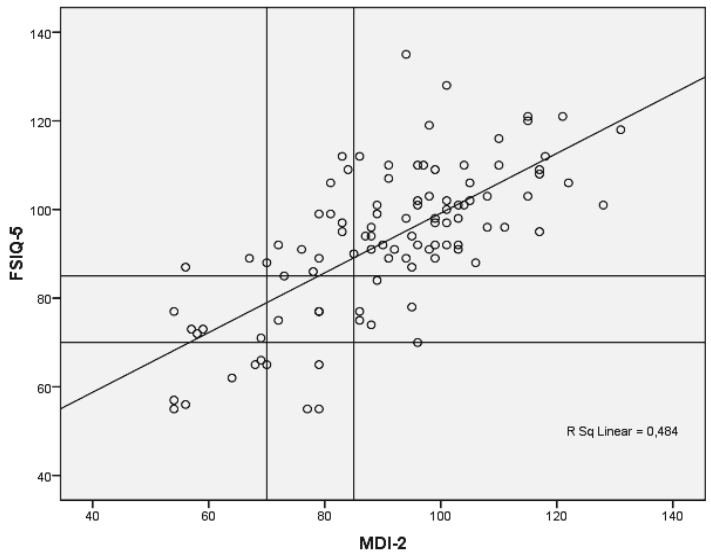
Figure 1 shows the correlation between MDI-2 and FSIQ-5. Children with an MDI-2 of 97 or higher all had FSIQ-5 scores in the normal range. Only in children with an MDI-2 score of 118 or higher (n = 5), 100% had scores in the normal range on all 5-year cognitive outcomes.

Correlation coefficient (*r*) between MDI-2 and FSIQ-5 (.715) and MDI-3 and FSIQ-5 (.737) were not statistically different.

Appendix 1 displays the number of children with (ab)normal MDI-2/3 and IQ-5 scores, and the movement of children across categories over time.

Agreement between MDI and FSIQ was moderate (κ = 0.48 for MDI-2 and 0.40 for MDI-3). Of all different cognitive abilities measured at age 5, MDI-2 had the highest agreement with VIQ-5 (κ = 0.63). PIQ-5 and PSQ-5 showed lower agreement with MDI-2 (κ = 0.36 and 0.47 respectively). Agreement of MDI-3 with VIQ, PIQ and PSQ was 0.42, 0.41 and 0.37 respectively.

Figure 1 Correlation between MDI-2 and FSIQ-5 scores.



Data are presented as *n* (%).
Agreement (κ with linear weighting) for categories of scores between the age of 2 and 5 is 0.48, and between the age of 3 and 5 is 0.40.

Prediction of 5 year outcome

Table 3 displays the prediction models for cognitive outcomes at age 5. MDI-2 and MDI-3 were positively associated with FSIQ-5 (44% and 56% explained variance respectively). PDI-2, PDI-3 and neurological functioning did not improve the prediction of FSIQ-5, with MDI already in the model. Adding behaviour problems at age 2 somewhat improved the prediction. Neonatal sepsis and/or meningitis in both models and postnatal dexamethasone in the MDI-3 model increased the amount of explained variance and were

Table 3 Hierarchical regression analyses predicting cognitive outcomes at 5 years of age.

	FSIQ-5		VIQ-5		PIQ-5		PSQ-5	
	Developmental assessment at:		Developmental assessment at:		Developmental assessment at:		Developmental assessment at:	
	2 yrs	3 yrs	2 yrs	3 yrs	2 yrs	3 yrs	2 yrs	3 yrs
Independent variables	*n* = 94	*n* = 96	*n* = 100	*n* = 96	*n* = 100	*n* = 96	*n* = 100	*n* = 94
Step 1. Cognitive development (MDI)	.662**	.747**	.734**	.765**	.490**	.530**	.547**	.623**
R^2	.44	.56	.54	.59	.24	.28	.30	.39
Step 2. Psychomotor development (PDI)	---	---	---	---	---	---	.254*	.296**
R^2 (ΔR^2)							.34 (.04)	.44 (.05)
Step 3. Neurological functioning	---	---	---	---	---	---	-.173*	---
R^2 (ΔR^2)							.36 (.03)	
Step 4. Behaviour	-.190*	---	---	---	---	---	---	---
R^2 (ΔR^2)	.47 (.03)							
Step 5. Perinatal characteristics					---	---		
- Gestational age	-	-	-	-			-	-
- Very small for gestational age	-	-	-	-.109			-	-
- Sex	-	-	-	-			-.199*	-.269**
- Postnatal dexamethasone	-	-.135*	-	-			-	-.158*
- Oxygen support at 36 weeks pma	-	-	-	-			-	-
- Sepsis and meningitis	-.195*	-.122	-.176*	-.143*			-.153	-
- Severely abnormal ultrasound	-	-	-	-			-	-
R^2 (ΔR^2)	.51 (.04)	.59 (.03)	.57 (.03)	.62 (.03)			.43 (.07)	.52 (.08)
Step 6. Sociodemographic characteristics						---	---	---
- Parental foreign country of birth	-.137	-.205**	-.157*	-.194**	-.167			
- Parental education	-.235**	-.134	-.176*	-	-.234*			
R^2 (ΔR^2)	.57 (.06)	.64 (.05)	.62 (.05)	.65 (.03)	.31 (.07)			

Data are presented as Standardized Regression Coefficients β, as explained variance and added explained variance R^2 and ΔR^2, and as F-values.
* $p < .05$, ** $p < .01$
--- Step eliminated from analysis because of non-significance of F-value ($p > .05$)-
Variable removed from model via backward method because of non-significance of t-value ($p > .10$)

negatively associated with FSIQ-5. Middle/low parental education and foreign parental country of birth were negatively associated with FSIQ-5 and improved the amount of explained variance in the final step. A total of 57% and 64% of the variance of FSIQ-5 was explained in the MDI-2 and MDI-3 model respectively.

MDI-2 and MDI-3 were positively associated with VIQ-5 (54% and 59% explained variance respectively). Psychomotor development, neurological functioning and behaviour problems did not improve the prediction of VIQ-5. Neonatal sepsis and/or meningitis in both models and very small for gestational age in the MDI-3 model added to the explained variance and were negatively associated with VIQ-5. Sociodemographic characteristics that were subsequently negatively associated with VIQ-5 were foreign parental country of birth, and only in the 2-year model middle/low parental education.

MDI-2 and MDI-3 were positively associated with PIQ-5 (24% and 28% explained variance respectively). Middle/low parental education and foreign country of birth increased the amount of explained variance of PIQ-5 in the 2-year model only.

MDI-2 and MDI-3 were positively associated with PSQ-5 (30% and 39% explained variance respectively). With MDI in the model, PDI-2 and PDI-3 were also positively associated with PSQ-5. An abnormal neurological outcome at age 2 was negatively associated with PSQ-5. Perinatal and child characteristics that improved the amount of explained variance and were negatively associated with PSQ-5, were male sex in both models, sepsis and/or meningitis in the MDI-2 model, and postnatal dexamethasone in the MDI-3 model.

DISCUSSION

In this study, the association between cognitive development at the age of 2 and 3 and full scale, verbal and performance intelligence and processing speed at the age of 5 was investigated in VPT born children who were able to complete an intelligence test. Agreement between cognitive development at the age of 2 and 3 and cognitive functioning at the age of 5 was moderate to substantial for full scale and verbal intelligence and fair to moderate for performance intelligence and processing speed.

We found a relatively high mean cognitive score at age 3, compared to the scores at age 2 and age 5 (all corrected for gestational age). Apparently, the relatively high cognitive scores we found in an earlier study at age 3[13] did not persist until age 5. The question arises whether the differences in mean test scores represent true development, or that the favourable outcome at age 3 is an artefact of the test. General stability of developmental scores[19-21] lends support to the latter, in the

absence of a theoretical foundation supporting the specifically high developmental scores at age 3 in preterm born children. Regarding the preferability of test age (2 or 3 years), the results of our study are not univocal. Agreement and correlation statistics point to different preferred test ages, but the differences are too small for a well founded recommendation. In spite of this, the age of 2 appears to be preferable, because 1) the number of false negatives will be much lower and 2) priority should be given to the earliest possible age to start intervention in case of developmental problems.

Research has shown that agreement between the BSID-II MDI and later intelligence is poor and the predictive validity of MDI is low.[4,5] In the study of Crowe et al,[9] in which children were followed up after NICU treatment at age 2 and 4 1/2, explained variance of verbal and performance intelligence by MDI (first edition of the BSID) was 21% and 17% respectively. The current study reports higher proportions of explained variance, especially for verbal intelligence. An explanation for this difference may be the use of different versions of the BSID. A substantial part of the cognitive scale of the second edition consists of items that require verbal comprehension or response.

Performance intelligence and processing speed predominantly require non-verbal responses. It is not surprising, therefore, that only one quarter of the variance of performance intelligence was explained by MDI. The prediction of performance intelligence was not improved by other domains of toddler development. Lags in performance and processing speed development occur equally frequent as lags in verbal development in our study group. A study of Mulder et al.[22] showed that deficits in processing speed explained significant group differences in academic attainment between VPT and term children in middle childhood. The present study results add to this that early developmental assessments only marginally predict performance and processing speed development. Therefore, FU timed at toddler age is insufficient to reveal the full impact of preterm birth on all affected areas of cognitive development. In line with the above, we also found that only children with an MDI-2 of 118 or higher had normal scores (>85) on every aspect of intelligence at age 5. Moreover, almost half of the children in our study group had a (mildly) abnormal score for one of the areas of intelligence at age 5, while at the age of 2 and 3, only 33% and 15% respectively had a (mildly) abnormal MDI score. Note that children too disabled to be assessed, are not included in these percentages.

In this study, it was shown that developmental domains other than cognition at age 2 and 3 could not or only marginally improve the prediction of cognitive functioning at age 5. The hypothesis that a broad developmental assessment can be used to substantially contribute to a prediction of cognition at later age was not confirmed.
A study by the NICHD Neonatal Research Network[23] in extremely low BW infants showed that

neonatal infections are associated with poor neurodevelopmental outcomes at the age of 18-22 months. The present study demonstrates that the age of 2 is too early to show the full impact of infections. As complex cognitive functions develop until young adulthood,[24] longer term FU might reveal further impact of perinatal events.

In an earlier paper on outcomes in this cohort, we described the interaction of level of parental education with preterm birth on cognitive development at age 5.[25] In the present study it was, as expected, shown that sociodemographic factors have additional impact on 5 year outcome even if MDI at age 2/3 is already accounted for.

Strengths of the present study were the focus on different cognitive abilities and the availability of longitudinal developmental assessment at three time points. A study limitation was the lack of a term born control group. Another limitation was the exclusion of children too disabled to be assessed. Results cannot be generalized to all low functioning children. As the third version of the BSID is increasingly being used, the predictive value of that version also requires further examination. Nevertheless, still many papers are being published using the BSID-II as main outcomes.[2] This study can be of help to interpret these studies with respect to later outcomes.

CONCLUSION

In a cohort of VPT born children without severe disabilities, early cognitive development predicted verbal intelligence more accurately than performance intelligence and processing speed. FU of VPT children until at least age 5 is needed to distinguish between different aspects of cognitive development and find effects of interventions aiming to improve cognitive outcomes.

REFERENCE LIST

1. Harrold J, Schmidt B. Evidence-based neonatology: making a difference beyond discharge from the neonatal nursery. *Curr Opin Pediatr.* 2002;14:165-169.

2. Logan JW, O'Shea TM, Allred EN, Laughon MM, Bose CL, Dammann O, et al. Early postnatal hypotension and developmental delay at 24 months of age among extremely low gestational age newborns. *Arch Dis Child Fetal Neonatal Ed.* 2010.

3. American Academy of Pediatrics. Follow-up care of high-risk infants. *Pediatrics.* 2004;114:1377-1397.

4. Roberts G, Anderson PJ, Doyle LW. The stability of the diagnosis of developmental disability between ages 2 and 8 in a geographic cohort of very preterm children born in 1997. *Arch Dis Child.* 2010;95:786-790.

5. Hack M, Taylor HG, Drotar D, Schluchter M, Cartar L, Wilson-Costello D, et al. Poor predictive validity of the Bayley Scales of Infant Development for cognitive function of extremely low birth weight children at school age. *Pediatrics.* 2005;116:333-341.

6. Aylward GP. Cognitive and neuropsychological outcomes: more than IQ scores. *Ment Retard Dev Disabil Res Rev.* 2002;8:234-240.

7. Marlow N, Wolke D, Bracewell MA, Samara M. Neurologic and developmental disability at six years of age after extremely preterm birth. *N Engl J Med.* 2005;352:9-19.

8. Wechsler D. Wechsler preschool and primary scale of intelligence, 3rd edition (WPPSI-III). San Antonio, TX: Psychological Corporation; 2002.

9. Crowe TK, Deitz JC, Bennett FC. The relationship between the Bayley Scales of Infant Development and preschool gross motor and cognitive performance. *Am J Occup Ther.* 1987;41:374-378.

10. Aylward GP. Neurodevelopmental outcomes of infants born prematurely. *J Dev Behav Pediatr.* 2005;26:427-440.

11. Koldewijn K, Wolf MJ, van Wassenaer A, Meijssen D, van Sonderen L, van Baar A., et al. The Infant Behavioral Assessment and Intervention Program for very low birth weight infants at 6 months corrected age. *J Pediatr.* 2009;154:33-38.

12. Ganzevoort W, Rep A, Bonsel GJ, Fetter WP, van Sonderen L, de Vries JI, et al. A randomised controlled trial comparing two temporising management strategies, one with and one without plasma volume expansion, for severe and early onset pre-eclampsia. *BJOG.* 2005;112:1358-1368.

13. Houtzager BA, Gorter-Overdiek B, van Sonderen L, Tamminga P, van Wassenaer AG. Improvement of developmental outcome between 24 and 36 months corrected age in very preterm infants. *Acta Paediatr.* 2010;99:1801-1806.

14. Hendriksen J, Hurks P. WPPSI-III-NL Wechsler preschool and primary scale of intelligence-Third edition. Nederlandse bewerking. Amsterdam: Pearson Assessment and Information B.V.; 2009.

15. Van der Meulen BF, Ruiter SAJ, Lutje Spelberg HC, Smrkovsky M. Bayley Scales of Infant Development - II. Dutch version. Lisse (NL): Swets; 2002.

16. Verhulst FC, Koot JM, Akkerhuis GW, Veerman JW. Praktische handleiding voor de CBCL. Assen: Van Gorcum; 1990.

17. Hempel MS. Neurological development during toddling age in normal children and children at risk of developmental disorders. *Early Hum Dev.* 1993;34:47-57.

18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159-174.

19. Sajaniemi N, Hakamies-Blomqvist L, Makela J, Avellan A, Rita H, von Wendt L. Cognitive development, temperament and behavior at 2 years as indicative of language development at 4 years in pre-term infants. *Child Psychiatry Hum Dev.* 2001;31:329-346.

20. Koller H, Lawson K, Rose SA, Wallace I, McCarton C. Patterns of cognitive development in very low birth weight children during the first six years of life. *Pediatrics.* 1997;99:383-389.

21. Bhutta AT, Cleves MA, Casey PH, Cradock MM, Anand KJ. Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis. *JAMA.* 2002;288:728-737.

22. Mulder H, Pitchford NJ, Marlow N. Processing speed and working memory underlie academic attainment in very preterm children. *Arch Dis Child Fetal Neonatal Ed.* 2010.

23. Stoll BJ, Hansen NI, Adams-Chapman I, Fanaroff AA, Hintz SR, Vohr B, et al. Neurodevelopmental and growth impairment among extremely low-birth-weight infants with neonatal infection. *JAMA.* 2004;292:2357-2365.

24. Blakemore SJ, Choudhury S. Development of the adolescent brain: implications for executive function and social cognition. *J Child Psychol Psychiatry.* 2006;47:296-312.

25. Potharst ES, van Wassenaer A, Houtzager B, van Hus JWP, Last BF, Kok JH. High incidence of multi-domain disabilities in very preterm children at the age of 5. *J Pediatr.* 2011;159:79-85.

APPENDIX

Appendix 1 Number of children with (ab)normal MDI-scores at the age of 2 and 3 and IQ-scores at the age of 5

		Normal FSIQ	Mildly abnormal FSIQ	Severely abnormal FSIQ	Total
Age 2	Normal MDI	61 (61.0%)	6 (6.0%)	0 (0%)	67 (67%)
	Mildly abnormal MDI	13 (13.0%)	3 (3.0%)	4 (4.0%)	20 (20.0%)
	Severely abnormal MDI	2 (2.0%)	5 (5.0%)	6 (6.0%)	13 (13.0%)
	Total	76 (76.0%)	14 (14.0%)	10 (10.0%)	100 (100%)
Age 3	Normal MDI	69 (71.9%)	11 (11.5%)	2 (2.1%)	82 (85.4%)
	Mildly abnormal MDI	4 (4.2%)	2 (2.1%)	5 (5.2%)	11 (11.5%)
	Severely abnormal MDI	0 (0%)	1 (1.0%)	2 (2.1%)	3 (3.1%)
	Total	73 (76.0%)	14 (14.6%)	9 (9.4%)	96 (100%)

Data are presented as *n* (%).

Agreement (κ with linear weighting) for categories of scores between the age of 2 and 5 is 0.48, and between the age of 3 and 5 0.40.