

## Do Pediatricians Recognize Cognitive Developmental Problems in Preterm Children at Age 5 Years?

Meta Sondaar · Boudien J. M. van Kessel ·  
Martin J. K. de Kleine · Judy M. Briët ·  
A. Lya den Ouden · Anneloes van Baar

Published online: 9 June 2007

© Springer Science + Business Media, LLC 2007

**Abstract** Often developmental psychologists see children only after referral from physicians. Do pediatricians recognize which children in a known risk group are in need of a cognitive evaluation? A judgment by pediatricians, based on an assessment using a parent questionnaire, the Denver Developmental Screening Test (DDST) and the overall impression during a systematic examination, was compared in 368 very preterm children, not yet diagnosed with cognitive impairments, to the results of an intelligence test. The pediatric assessment moderately identified preterm children with cognitive developmental problems. A low score on the DDST part of the pediatric assessment and information on extra assistance at school or grade retention, best indicated which children should be referred. Preterm born children known to need extra assistance at school at 5 years of age, should be referred for cognitive

---

M. Sondaar

Department of Medical Psychology, Deventer Hospital, Deventer, The Netherlands

B. J. M. van Kessel

Department of Medical Psychology, University Medical Centre Nijmegen,  
Nijmegen, The Netherlands

M. J. K. de Kleine

Department of Neonatology, Máxima Medical Centre, Veldhoven, The Netherlands

M. J. K. de Kleine · A. L. den Ouden

TNO Prevention and Health, Leiden, The Netherlands

J. M. Briët

Department of Neonatology, Academic Medical Centre, Amsterdam, The Netherlands

A. L. den Ouden

Health Care Inspectorate, Ministry of Health, The Hague, The Netherlands

A. van Baar (✉)

Department of Pediatric Psychology, Tilburg University, Tilburg, The Netherlands  
e-mail: A.vanBaar@uvt.nl

assessment by a child psychologist, in view of their risk for difficulties in cognitive functioning and associated problems.

**Keywords** Cognitive disabilities · Developmental assessments · Follow-up studies · Preterm and very low birth weight children

Preterm children are at risk for developmental and school problems (Anderson and Doyle 2003; Davis 2003; Saigal *et al.* 2003). Long term follow-up shows that up to 55% of these children experience cognitive and learning problems during school age. In a meta-analysis of case-control studies, Bhutta *et al.* (2002) showed that the mean intelligence quotient (IQ) of very preterm children during school age was, depending on gestational age and birth weight, approximately 10 points below that of healthy controls. Systematic developmental and cognitive assessment of all survivors of neonatal intensive care is necessary to evaluate perinatal care as well as to identify as early as possible which children may need extra stimulation or intervention. In The Netherlands a national working party on neonatal follow-up designed a multidisciplinary and standardized follow-up program that would offer postnatal care as well as standardized follow-up figures for all very preterm infants. The regular, standardized and multidisciplinary assessments that the program entails, is found difficult to implement. Costs involved by appointing for instance enough psychologists for this work in follow up clinics play a role, and perhaps also doubts if such a relatively extensive program of standardized assessments is really necessary for early identification of children with developmental problems.

We evaluated an assessment tool for the age of 5 years that could help a pediatrician to identify which survivors of neonatal intensive care have developmental disturbances that may interfere with progress in normal education and functioning in daily life (De Kleine *et al.* 2003). Such an instrument could be sufficient in itself for evaluation of perinatal care and secondary prevention purposes, or it could serve as a filter by identifying children that need further assessment.

In this paper we present results of a comparison of this pediatric assessment in 368 very preterm infants, not diagnosed with cognitive impairment before, with a formal intelligence test by a child psychologist. We studied if pediatric assessment enables the physician to identify children that should be referred to a child psychologist for a more extensive cognitive assessment. We also evaluated which items of the pediatric assessment most clearly identify children that need further cognitive assessment.

## Methods

### Study Population

The total study population consisted of 764 infants of less than 32 weeks of gestation or weighing less than 1,500 grams that were born in a time period between October 1992 and December 1994. The children were treated in three Dutch neonatal

intensive care units: the University Medical Centre Nijmegen (UMCN); the Academic Medical Centre (AMC), Amsterdam; and the Máxima Medical Centre (MMC), Veldhoven. Mortality before the age of 5 years was 17.2% ( $n=131$ ). Forty-six patients with gestational ages below 30 weeks (6.0%) were excluded in the AMC because they participated in another study (van Wassenaer *et al.* 1997) and 21 (3%) were excluded because severe handicaps were present such as cerebral palsy, blindness, severe mental retardation, chromosomal abnormalities or inborn error of metabolism and it was obvious that they could not perform the tests at 5 years of age. Of the 566 eligible children 431 (76%) responded and 400 (71%) had a full assessment by a pediatrician as well as an intelligence test by a psychologist (de Kleine *et al.* 2003).

As the objective of this study was to verify whether the pediatrician's assessment could identify possible cognitive defects in children not yet known to have developmental difficulties, children with an already existing diagnosis of cognitive impairment and children in special education were excluded from the analyses ( $n=32$ , 8%).

## Procedure

At 5 years of age the children were assessed on the same day by a pediatrician, a child psychologist and a physiotherapist. Appointments were scheduled in a random order, and were limited to a fixed time. The professionals were blinded for each other's findings. Only the pediatrician was aware of the detailed medical and perinatal history. No correction for preterm birth was applied.

The institutional medical ethical review board of each of the participating hospitals approved the study and written parental consent was obtained.

## Measurements

### *Pediatric Assessment*

The pediatric assessment consisted of a questionnaire sent a priori to the parents and a standardized examination by a pediatrician specifically trained for this purpose. For the present study we analyzed the part of the questionnaire addressing school performance, ethnic origin (Dutch or non-Dutch) and maternal education. School performance was redefined in two categories: normal (no problems) or deviant (grade retention, remedial teaching or other forms of extra help).

The pediatric assessment started with a check and further exploration of the data from the parental questionnaire, and continued with a formal pediatric examination, a neurological assessment according to Touwen (1989), and the Dutch version of the Denver Development Screening Test (DDST; Cools and Hermans 1979). The DDST addresses four domains, adaptive behavior, social development, language development and motor development, in children between 1 month and 6;00 years (years; months). It consists of 105 test items with a cut-off point for each at an age that 90% of Dutch children are able to perform it. Each domain is separately categorized as normal (no delays or one delay compensated by one early item) or abnormal (one delay without compensation or two or more delays). Overall DDST classification is

normal (all four domains normal), at risk (one domain abnormal) or abnormal (two or more domains abnormal). Other parts of the pediatric assessment referred to language and speech, using a formal Dutch language screening test, the Dutch Taal Screeningstest (TST; Gerritsen 1988). The TST is more elaborate than the language development part of the DDST. It consists of nine subtests used to examine different language abilities (naming objects with the same function, plural forms of nouns, repeating sentences, pointing out parts of the body, repeating words, complete sentences with conjunctions, knowledge of prepositions, analogies and antithesis and understanding and insight). At age 5 language development is categorized as normal (17 errors or less), moderately delayed (18–25 errors) or severely delayed (26 errors or more).

At the end of the assessment the pediatrician gave an overall judgment of the cognitive development, neurological functioning and behavior of the child based on the performance on all tests and their overall impression of the child. These overall judgments were classified as: (1) normal, (2) re-assessment necessary in due time, or (3) referral for further examination or treatment necessary.

### *Intelligence Test*

Trained child psychologists administered the short version of the RAKIT, a Dutch intelligence test devised for children between 4;02 and 11;02 years of age (Bleichrodt *et al.* 1984). This short version takes approximately 50 min and has a correlation of 0.93 with the full-scale test (Bleichrodt *et al.* 1987). The short version contains the subtests Exclusion, Verbal Comprehension, Discs and Idea Production. For the children beneath 5;02 years the short version also contains the subtest Closure, whereas for the older children the subtests Learning Names and Hidden Figures are to be used. The subtests measure verbal capacities, perceptual and executive capacities, word fluency, memory and reasoning. The subtests are designed in a way to prevent cultural bias, which was done by reducing the influence of language. The concurrent validity with the WISC-R is 0.86 for total IQ. The mean score is 100 with a standard deviation (SD) of 15. All scores  $\geq 85$  ( $-1$  SD) are classified as normal, while all scores below 1 standard deviation ( $<85$ ) are classified as impaired.

### Statistical Analysis

Statistical analyses were performed in SPSS 14.0 (SPSS). Differences in the normal distribution were compared with one-sample Kolmogorov–Smirnov. Differences between means were compared with the Student *t* test and agreement in  $2 \times 2$  tables with Cohen's Kappa. Two-tailed comparisons were used and *p* values  $<0.05$  were considered statistically significant. Diagnostic efficiency of the assessments by the pediatrician was defined with diagnostic efficiency statistics: the sensitivity, the specificity, the positive predictive value (PPV), the negative predictive value (NPV), the positive likelihood ratio (LR+), the negative likelihood ratio (LR-), and odds ratio (OR; Bouter and Van Dongen 2000).

## Results

### Participants

A total of 368 patients was included in this study. Mean gestational age was  $30.2 \pm 1.9$  weeks; mean birth weight was  $1,272 \pm 329$  g. Of the study group, 54% were male, 35% multiple births, 48% had been artificially ventilated, 14% sustained a sepsis and 18% had an intraventricular hemorrhage in the neonatal period. Mean duration of neonatal intensive care was 28 days.

Mean age at the time of assessment was  $5;02 \pm 1.9$  months. Only 2% of the participants were from non-Dutch origin. Of the mothers 27% had a low educational level and 20% a high educational level.

### Intelligence Test

The RAKIT revealed that 83% of the children in this study had IQ's of 85 or higher, 13% had borderline scores (70–84) and 4% scored deviant ( $<70$ ). Mean IQ was  $98.5 \pm 14.6$ , range 56–142. The IQ scores were normally distributed (Kolmogorov–Smirnov 0.811, *df* 367,  $p=0.53$ ). In comparison to the Dutch population norm of 100 the difference was marginally significant (*t* test  $-1.932$ , *df* 367,  $p=0.05$ ).

### Outcome of Cognitive Developmental Screening by Pediatricians

The overall judgment of the cognitive development by the pediatrician showed that they considered 92% of the children normal. On the DDST part 71% of the children were classified as normal (Table 1). Most delays were found on the language (12%) and social (18%) scales of the DDST. The TST identified slightly more language problems than the language scale from the DDST, classifying 17% of the children as moderately or severely delayed. School performance was found to be delayed in 37% of the children.

### Agreement Between Pediatric Assessment of Cognitive Development and IQ-Test

Correspondence in overall judgement by the pediatricians and the RAKIT IQ was only fair (kappa 0.39). A moderate match (kappa between 0.40 and 0.60) was found between the TST and the IQ (Table 1). The agreement between the DDST-motor scale and the IQ was poor (kappa  $<0.20$ ), as expected. The correspondence between the other parts of the pediatric assessment of cognitive development and the IQ was fair (kappa between 0.20 and 0.40).

### Diagnostic Efficiency of the Pediatric Assessments

A good screening instrument has a high sensitivity (children with impaired scores on the pediatric assessment do have an impaired IQ) and a high negative predictive value (few children with a normal scores on the pediatric assessment have an

**Table 1** Agreement between pediatric assessments of cognition and the intelligence test

	Impaired IQ N=61 (17%)	Normal IQ N=307 (83%)	Total N=368	Kappa
Overall judgment of cognitive development				
Re-evaluation/referral	21	9	30 (8%)	0.395
Normal	40	298	338 (92%)	
DDST				
Borderline/deviant	44	65	109 (30%)	0.387
Normal	17	242	259 (70%)	
DDST-Adaptation				
Borderline/deviant	17	4	21 (6%)	0.360
Normal	44	303	347 (94%)	
DDST-Language				
Borderline/deviant	19	24	43 (12%)	0.265
Normal	42	283	325 (88%)	
DDST-Motor				
Borderline/deviant	11	12	23 (6%)	0.188
Normal	50	295	345 (94%)	
DDST-Social				
Borderline/deviant	22	43	65 (18%)	0.215
Normal	39	264	303 (82%)	
TST <sup>a</sup>				
Borderline/deviant	33	29	62 (17%)	0.450
Normal	27	278	305 (83%)	
School performance				
Double/assistance	48	90	138 (37%)	0.318
Normal	13	217	230 (63%)	

<sup>a</sup> One less tested with the TST

impaired IQ). The complete DDST performed best of all parameters available for the pediatrician with a sensitivity of 0.72, a negative predictive value of 0.93 and a likelihood ratio of a negative test of 0.35. Information on school performance showed good concordance with IQ, with a sensitivity of 0.79, a negative predictive value of 0.94, and a likelihood ratio of a negative test of 0.35. All diagnostic efficiency values are presented in Table 2.

**Table 2** Diagnostic efficiency statistics for pediatric assessments of cognition

	Sensitivity	Specificity	PPV	NPV	LR+	LR-
Overall judgment of cognitive development	0.34	0.97	0.70	0.88	11.74	0.68
DDST	0.72	0.79	0.40	0.93	3.41	0.35
DDST-Adaptation	0.28	0.99	0.81	0.87	21.39	0.73
DDST-Language	0.31	0.92	0.44	0.87	3.98	0.75
DDST-Motor	0.18	0.96	0.48	0.86	4.61	0.85
DDST-Social	0.36	0.86	0.34	0.87	2.58	0.74
TST	0.55	0.91	0.53	0.91	5.82	0.50
School performance	0.79	0.71	0.35	0.94	2.68	0.30

## Discussion

The objective of this study was to investigate if a standardized pediatric follow-up instrument for 5-year-old preterm children adequately identifies which children need further assessment of cognitive development. Results of our study show that the agreement between the pediatric assessment of cognition and the formal intelligence test varies from poor to moderate. Sensitivity of the overall judgment of cognition by the pediatrician and of the DDST subscales was low and overestimated the cognitive abilities of the children and hence underestimated the need for formal cognitive testing. The complete DDST combined with information on school performance provided the best identification. Pediatricians often use the development of language as an indicator of cognition, but the sensitivity of the TST was also rather low. Knuijt *et al.* (2004) documented that the performance of this test may be improved by lowering the cut off point from 17 to 16. In a population at high risk for developmental problems, it is necessary to detect as many children with developmental problems as possible. Therefore, assessments should be aimed at high negative predicting values and negative likelihood ratios close to zero. The complete DDST, school performance, and TST were found to have high negative predicting values. School performance at 5 years apparently and not surprising is a good indicator of cognitive difficulties. However, aim of early detection of cognitive problems is to detect these before causing problems in daily functioning. Problematic school performance in preterm children was also found in an international study of four countries, but to a larger extent as it concerned older and extremely low birth weight children (Saigal *et al.* 2003). Results of the IQ-test differ favorably from other studies on cognitive development of preterm children (Davis 2003; Wolke and Meyer 1999). Our results are heavily influenced by our inclusion criteria, in that we studied a group of relatively healthy preterm children. In line with the objective of this validation study, children known to be seriously handicapped were not invited and children with known cognitive impairments were not analyzed for this study. Furthermore, it is also known that actually examined children perform better than “non-response” children, who also have worse perinatal characteristics (Wolke *et al.* 1995; Tin *et al.* 1998). Therefore, our results do not reflect cognitive outcome for the total population of preterm survivors of neonatal intensive care.

Another explanation for the rather high mean score on the intelligence test is that it underestimates mild developmental delay. Flynn has pointed out that IQ increases approximately 0.3 per year (Flynn 1987, 1999). As a result, standardized tests get outdated in time (Wolke *et al.* 1994). The RAKIT was standardized in 1984. With a yearly increase of 0.3 point, the mean IQ in a contemporary control group (tested between 1997–1999) may well be 3–4.5 points higher than the norm we used. This Flynn effect however, forms a problem for all IQ test norms. As use of norm scores follows convention and allows comparability between studies, we did not use an adapted norm score. Furthermore, results of all kinds of assessments are influenced by estimation errors and intra-individual variation in accomplishments. This can be solved by repeated measurement and by the use of multiple sources and instruments when important individualized decisions concerning the development of preterm born children, are made.

The decision to refer a preterm born child for further assessment by a pediatrician using the systematic assessment may be based on results of the combination of the DDST and the information on school performance. Information on school performance can be obtained easily and without further constraint and costs. It may therefore be used as a filter and reduce the costs of a full developmental assessment by a psychologist for all preterm children. When the aim is to remediate cognitive disabilities or even prevent school problems, referral of the child when such problems are obvious, is too late and even then 30% of the impaired children will be missed. Since 37% of the children in this study already have school assistance or grade retention, an earlier preschool cognitive assessment is necessary. We therefore recommend to avoid waiting until cognitive deficits have led to problems at school, but to test all preterm born children with a thorough, formal cognitive assessment at toddler or preschool age.

**Acknowledgements** The authors thank Louis A. A. Kollée, Maria W. G. Nijhuis-van der Sanden and Simone Knuijt (University Medical Centre Nijmegen), Adri Ilzen, Rinske Breur-Pieterse and Karin Koldewijn (Academic Medical Centre, Amsterdam), Jan Buijs en Marion Wit (Máxima Medical Centre Veldhoven), Ina Kloosterboer-Boerriqter (TNO Prevention and Health, Leiden) for their participation in the Validation Study Group.

## References

- Anderson, P., & Doyle, L. W. (2003). Neurobehavioral outcomes of school-age children born extremely low birth weight or very preterm in the 1990s. *JAMA*, *289*, 3264–3272.
- Bhutta, A. T., Cleves, M. A., Casey, P. H., Cradock, M. M., & Anand, K. J. (2002). Cognitive and behavioral outcomes of school-aged children who were born preterm: A meta-analysis. *JAMA*, *288*, 728–737.
- Bleichrodt, N., Drenth, P. J. D., Zaal, J. N., & Resing, W. C. M. (1984). *Revisie amsterdamse kinder intelligentie test*. Lisse: Swets & Zeitlinger.
- Bleichrodt, N., Resing, W. C. M., Drenth, P. J. D., & Zaal, J. N. (1987). *Intelligentie-meting bij kinderen: Empirische en methodologische verantwoording van de geReviseerde amsterdamse kinder intelligentie test*. Lisse: Swets & Zeitlinger.
- Bouter, L. M., & Van Dongen, M. C. I. M. (2000). *Epidemiologisch onderzoek: Opzet en interpretatie*. Houten: Bohn Stafleu Van Loghum.
- Cools, A. T. M., & Hermans, J. M. A. (1979). *DOS-handleiding: Denver ontwikkeling screeningstest*. Lisse: Swets & Zeitlinger.
- Davis, D. W. (2003). Cognitive outcomes in school-age children born prematurely. *Neonatal Network*, *22*, 27–38.
- Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ test really measure. *Psychological Bulletin*, *101*, 171–191.
- Flynn, J. R. (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist*, *54*, 5–20.
- Gerritsen, F. M. E. (1988). *VTO Taalscreening 3- tot 6-jarigen: de ontwikkeling van een taalscreening-instrument voor gebruik in de Jeugdgezondheidszorg*. Amsterdam/Lisse: Swets & Zeitlinger.
- De Kleine, M. J. K., Den Ouden, A. L., Kollee, L. A., Nijhuis-van der Sanden, M. W., Sondaar, M., Kessel-Feddema, B. J., et al. (2003). Development and evaluation of a follow up assessment of preterm infants at 5 years of age. *Archives of Disease in Childhood*, *88*, 870–875.
- Knuijt, S., Sondaar, M., De Kleine, M. J. K., & Kollée, L. A. A. (2004). Validation of a Dutch language screening instrument for 5-year-old preterm infants. *Acta Paedagogica*, *93*, 1372–1377.
- Saigal, S., den Ouden, A. L., Wolke, D., Hoult, L., Paneth, N., Streiner, D. L., et al. (2003). School-age outcomes in children who were extremely low birth weight from four international population-based cohorts. *Pediatrics*, *112*, 943–950.
- Tin, W., Fritz, S., Wariyar, U., & Hey, E. (1998). Outcome of very preterm birth: Children reviewed with ease at 2 years differ from those followed up with difficulty. *Archives of Disease in Childhood. Fetal and Neonatal Edition*, *79*, F83–F87.



- Touwen, B. C. (1989). Examination of the child with minor neurological dysfunction. *Clinics in Developmental Medicine*, *11*, 353–363.
- Van Wassenaer, A. G., Kok, J. H., De Vijlder, J. J., Briet, J. M., Smit, B. J., Tamminga, P., *et al.* (1997). Effects of thyroxine supplementation on neurologic development in infants born at less than 30 weeks' gestation. *New England Journal of Medicine*, *336*, 21–26.
- Wolke, D., & Meyer, R. (1999). Cognitive status, language attainment, and prereading skills of 6-year-old very preterm children and their peers: The Bavarian longitudinal study. *Developmental Medicine & Child Neurology*, *41*, 94–109.
- Wolke, D., Ratschinski, G., Ohrt, B., & Riegel, K. (1994). The cognitive outcome of very preterm infants may be poorer than often reported: An empirical investigation of how methodological issues make a big difference. *European Journal of Pediatrics*, *153*, 906–915.
- Wolke, D., Sohne, B., Ohrt, B., & Riegel, K. (1995). Follow-up of preterm children: Important to document dropouts. *Lancet*, *345*, 447.