Evaluation of neuromotor function in infancy – a systematic review of available methods

Kirsten Heineman¹,² and Mijna Hadders-Algra¹

¹Department of Paediatrics, Institute of Developmental Neurology
²Department of Neurology
University Medical Center Groningen

Chapter 2

ABSTRACT

Background: Neuromotor function in infancy can be evaluated in various ways. Assessment instruments are used for early detection of children with a high risk for developmental disorders. Early detection enables clinicians to provide intervention at a young age when plasticity of the nervous system is high. The assessments may also be used to monitor intervention. The present paper will review the psychometric properties of methods to assess neuromotor function in infancy.

Methods: A literature search was performed in PubMed, Medline and PsycINFO (1966 to March 2007) on instruments to assess neuromotor functioning of infants.

Results: Fifteen instruments were included and classified into four groups: 1) Comprehensive neurological examinations (n=4). These techniques are widely used, though little is known about their reliability. Their validity in predicting major developmental disorders such as cerebral palsy is good; their predictive validity for minor motor disorders is moderate at best. 2) Procedures with standardized scoring (n=7). These have good reliability, but only moderate predictive validity for major developmental disorders. No data available for prediction of minor developmental disorders. 3) Observation of milestones (n=2). Its predictive validity for major developmental disorders is only moderate, while reliability is good. 4) Assessment of quality of motor behavior or motor patterns (n=2). These instruments have the best predictive validity for major and minor developmental motor disorders, but current methods are only useful under the age of four months.

Conclusion: Prediction of developmental outcome at an early age is difficult. In medical evaluations of high-risk infants the best predictions are achieved through a combination of multiple, complementary tools, that is, achieved milestones, neurological examination and assessment of the quality of motor behavior.
INTRODUCTION

As the chances of survival of preterm and high-risk full-term infants have increased\cite{1,2}, extensive follow-up programs have been developed to determine which of these infants need intervention. Recent studies suggest that intervention may be most effective when it is applied during infancy when there is high plasticity of the brain\cite{3,4}. A prerequisite for early intervention is early detection of infants with a high risk for major developmental disorders such as cerebral palsy (CP) and minor motor disorders such as developmental coordination disorder (DCD) and minor neurological dysfunction (MND). It appears that parents of children with developmental disorders are concerned significantly later than physicians are about the developmental status of their children and therefore, in general, cannot be relied on for early recognition of infants who are likely to benefit from early intervention\cite{5}. Physiotherapists, occupational therapists, pediatricians and other clinicians in primary health care settings play an important role in early detection. They have a heterogeneous group of instruments at their disposal for the detection of early evidence of motor dysfunction in high-risk infants. In general, these instruments are not only used for detection but also for the evaluation of the effectiveness of an intervention. Usually the instruments are chosen based on habit and for practical reasons, and not on the basis of information regarding test accuracy and utility and theoretical basis\cite{6}. In fact, a primary selection criterion should be: “Has the instrument been designed for the task at hand?” Kirshner and Guyatt\cite{7} classified health measure instruments into three categories according to the goals they served. The first one is discrimination. In the field of neuromotor assessment this implies making a distinction between children who show features of a deviant neuromotor function compared to the general, healthy population. The second purpose is prediction; that is, instruments are used as a diagnostic tool to predict developmental outcome, for example, the likelihood that a child will develop CP. The third purpose is evaluation, the measurement of longitudinal change of an individual or group over time, for example, changes in motor function of infants enrolled in early intervention programs. Instruments are generally validated for only one of the three goals. This means that the instruments cannot automatically be used for other purposes\cite{6}.

The aim of this paper is to present a systematic review of the instruments used for the evaluation of neuromotor function and motor behavior in infancy. The contents of the methods will be reviewed, while special attention will be paid to psychometric properties, that is, reliability and validity.

METHODS

Selection procedure

A literature search in the following databases was performed: PubMed, Medline (1966 to March 2007) and PsycINFO (1967 to March 2007). Keywords used were “neuromotor,” “motor development,” “motor behavior,” “assessment,” “neurological examination,” “evaluation,” “instrument,” “method,” “infants,” “neonatal,” “preschool” and “review.” All articles with a name of an assessment in the title and/or abstract and reviews on one or more methods were selected. Further searches with names
of assessments and authors were performed and references to the articles were studied to find information on reliability and validity. Manuals were obtained when available in the Netherlands.

Instruments were included when they could be applied to infants aged three to eighteen months and also when the age range of application was more extensive. We focused on the age range from three to eighteen months, as it has been relatively neglected. Reviews on neonatal neurological evaluation\textsuperscript{8,9}, instruments to assess motor function of children diagnosed with CP\textsuperscript{6,10} and assessments of motor development and function in preschool children aged eighteen months to four years\textsuperscript{11} were available. Methods were selected if they focused on neurological condition or motor performance, or if they combined items on neuromotor function with items on other developmental domains such as mental development, speech or behavior. In the latter case, only data on reliability and validity of the motor subscale were reviewed. Instruments were only included if they had been described in at least two English-language peer-reviewed papers.

Instruments used only for screening purposes were excluded, since 1) an overview of frequently used general developmental screening instruments was provided by Glascoe\textsuperscript{12} and 2) our main interest was “full” assessment of the infant’s neuromotor functioning. Screening was defined as “the application to all children born of certain procedures that can be carried out in a short time by the less specialized members of staff and that will give indication of the presence and absence of certain disabilities” (WHO 1967)\textsuperscript{13}. Screening is important in clinical practice. This is illustrated, for instance, by the considerable power of infant motor screening tests to predict CP, such as the Early Motor Pattern Profile\textsuperscript{14} and Capute’s motor quotient\textsuperscript{15}. Instruments were considered screening instruments if the words screen or screening were part of the instrument’s name and/or if the authors mentioned “screening” as the main purpose of the instrument. Fifteen instruments fulfilled the selection criteria and were included in the review.

Evaluation procedure
The selected instruments were systematically evaluated with a focus on population, age, purpose of instrument (discrimination, prediction or evaluation), type of instrument, test description, type of test data (categorical vs. continuous), test construction and standardization, training required to become an assessor and time needed to administer the test. For the classification of the type of instrument, the instrument’s dominant features were used.

The various types of validity and reliability were evaluated by the criteria presented in Table I. Validity is the extent to which an instrument measures what it is intended to measure. We addressed construct validity, concurrent validity and predictive validity. Construct validity is the extent to which items reflect the theoretical construct of interest, in this case neuromotor functioning. For tests of neuromotor function, it may be assumed, for instance, that relationships of pre-, peri- and neonatal adversities, such as preterm birth, and results of brain imaging with test scores may contribute to construct validity. Concurrent validity is the extent to which scores relate to scores on another measure on the same construct, ideally a gold standard. If there is no gold standard available,
correlation with other established instruments is assessed. Predictive validity of an instrument is the extent to which the scores on the instrument now predict future outcome \(^{11,17}\). In the present review, we concentrated on predictive validity for developmental motor disorders; we distinguished predictive validity for major motor disorders such as CP from predictive validity for minor disorders such as DCD and MND. Factors that were taken into consideration in the judgment of the predictive validity of the various instruments were type of population (typically developing children vs. high-risk population), age at follow-up and tests used at follow-up. Reliability is the ability of a

Table I: Evaluation criteria

<table>
<thead>
<tr>
<th>Construct validity</th>
<th>++ very good</th>
<th>+ good</th>
<th>± moderate</th>
<th>– poor</th>
<th>nda no data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score relationship</td>
<td>Scoring “good” twice, that is, good for relationship between scores and pre, peri- and neonatal adversities, and good for correlation of scores with results of brain imaging indicates very good construct validity</td>
<td>Scores negatively affected by pre-, peri- and neonatal adversities such as PT birth, IUGR, other medical complications, or</td>
<td>Scores are correlated with visible pathology on brain imaging (US, MRI), or with electrophysiological parameters of brain function (CFM, EEG)</td>
<td>Scores are not related to any of the above-mentioned factors</td>
<td>No data available</td>
</tr>
</tbody>
</table>

Concurrent validity and predictive validity

<table>
<thead>
<tr>
<th>++ very good</th>
<th>+ good</th>
<th>± moderate</th>
<th>– poor</th>
<th>nda no data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohen’s kappa or Spearman’s ( \rho &gt; 0.80^{\text{a}} )</td>
<td>Cohen’s kappa or Spearman’s ( \rho = 0.40 – 0.60 )</td>
<td>Cohen’s kappa or Spearman’s ( \rho &lt; 0.40 )</td>
<td>Cohen’s kappa or Spearman’s ( \rho = 0.40 – 0.60 ); agreement 60-80%</td>
<td>No data available</td>
</tr>
</tbody>
</table>

Intra-observer agreement and inter-observer agreement

<table>
<thead>
<tr>
<th>++ very good</th>
<th>+ good</th>
<th>± moderate</th>
<th>– poor</th>
<th>nda no data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohen’s kappa or Spearman’s ( \rho &gt; 0.80^{\text{a}} )</td>
<td>Cohen’s kappa or Spearman’s ( \rho = 0.40 – 0.60 ); agreement &gt;80%</td>
<td>Cohen’s kappa or Spearman’s ( \rho = 0.40 – 0.60 ); agreement 60-80%</td>
<td>Cohen’s kappa or Spearman’s ( \rho &lt; 0.40 ); agreement &lt; 60%</td>
<td>No data available</td>
</tr>
</tbody>
</table>

\(^{a}\) Criteria for reliability according to Landis and Koch\(^{16}\), extended to Spearman’s \( \rho \) and % agreement

PT = preterm, IUGR = intrauterine growth restriction, US = ultrasound, MRI = magnetic resonance imaging, CFM = cerebral function monitor, EEG = electroencephalography and PPV = positive predictive value
measurement to give consistent scores on repeated assessments in the absence of change in the characteristics being studied18. Intra-observer agreement is the stability of the observer’s ratings on the same behavior. Often videotapes are used, which are scored twice by the same observer after a pre-determined time interval. Interobserver reliability is the stability of ratings across different evaluators8. The ages used in the review imply that preterm infants are assessed at ages corrected for preterm birth.

RESULTS

Description of instruments

Fifteen methods to assess neuromotor function in infancy fulfilled the inclusion criteria and were included in the review. Their main characteristics are described in Table II. The age range in which the instruments can be applied varied between birth to four months and zero to six years. Four types of instruments were discerned: 1) comprehensive neurological examination, 2) procedures with standardized scoring (i.e. condensed neurological assessment with or without observation of motor behavior with standardized scoring), 3) observation of milestones and specific aspects of motor behavior, and 4) quality of motor behavior. Four instruments were classified as comprehensive neurological examinations with a focus on cranial nerves, posture, muscle tone, reflexes and reactions19-22. Seven instruments were procedures with standardized scoring24,25,29,31-34. Two instruments focused on observation of motor milestones and specific aspects of motor behavior35,36. Two instruments focus on assessing the quality of motor behavior, including postural adjustments40,41,37.

The authors of all the instruments stated that the main purpose of their instrument was to discriminate between infants with a deviant neuromotor condition and infants falling within the range of typical development. Additional purposes have been described for six instruments: three40,41,21,22 aimed at prediction of future neuromotor development and another three29,32,34 at evaluation of changes in motor function over time or during intervention. Test construction differed considerably for the various methods, but most assessments took elements from pre-existing methods, adapted these and/or combined them with other test elements. For four instruments test construction was not described. Data on standardization was available for seven out of the fifteen instruments. These standardization procedures were carried out in strikingly heterogeneous ways with respect to sample sizes (ranging from 35 to 2202 children) and types of populations (typically developing infants, neonatal intensive care unit (NICU) graduates or infants with a high risk for developmental delay). Assessors were pediatricians, neonatologists, psychologists, occupational therapists, nurses or other clinicians working in NICU follow-up programs. Time needed to administer the test varied from a few minutes to one hour.
<table>
<thead>
<tr>
<th>Assessment (Author)</th>
<th>Short name</th>
<th>Population</th>
<th>Age group</th>
<th>Purpose</th>
<th>Test description - assessment of:</th>
<th>Type of variables</th>
<th>Test construction / standardization</th>
<th>Assessors / time to administer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Touwen infant neurological examination (Touwen\textsuperscript{19})</td>
<td>Touwen infants</td>
<td>0 mos. to independent walking</td>
<td>D</td>
<td>Cranial nerves, posture, tone, reflexes and reactions, trunk coordination and gross and fine motor functions</td>
<td>cat.</td>
<td>Based on traditional neuropediatric concepts (&quot;Groningen school&quot;) / nda</td>
<td>Pediatricians and others with specific neurodevelopmental training / 15 minutes</td>
<td></td>
</tr>
<tr>
<td>Amiel-Tison neurological examination (Amiel-Tison and Gosselin\textsuperscript{20})</td>
<td>Amiel-Tison at-risk infants</td>
<td>0 to 6 yrs.</td>
<td>D</td>
<td>Active and passive muscle tone, cranial nerves, motor milestones, spontaneous motor activity, reflexes and reactions, qualitative abnormalities</td>
<td>cat.</td>
<td>Based on traditional neuropediatric concepts (&quot;French school&quot;) / nda</td>
<td>Neonatologists, pediatricians with neurodevelopmental training, occupational therapists / 10 minutes</td>
<td></td>
</tr>
<tr>
<td>Active and passive muscle power (de Groot\textsuperscript{21})</td>
<td>Muscle power high-risk infants</td>
<td>3 to 12 mos.</td>
<td>D, P</td>
<td>Special emphasis on relationship between active and passive muscle power; both components should be in balance in order to create a stable posture and fluent motility</td>
<td>nda / nda</td>
<td>nda / nda</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hammersmith Infant Neurological Examination (Haataja et al.,\textsuperscript{22})</td>
<td>HINE infants</td>
<td>2 to 24 mos.</td>
<td>D, P</td>
<td>Cranial nerve function, posture, movements, tone, reflexes and reaction (26 items), developmental milestones (8 items), state of behavior (3 items)</td>
<td>cont.</td>
<td>Based on Dubowitz and Dubowitz method for neurologic assessment of the newborn\textsuperscript{13} / standardized in a low-risk population of 135 infants at age 12 and 18 mos.</td>
<td>Routine clinical practice, relatively inexperienced staff / nda</td>
<td></td>
</tr>
<tr>
<td>Assessment (Author)</td>
<td>Short name</td>
<td>Population</td>
<td>Age group</td>
<td>Purpose</td>
<td>Test description - assessment of:</td>
<td>Type of variables</td>
<td>Test construction / standardization</td>
<td>Assessors / time to administer</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------</td>
<td>---------</td>
<td>----------------------------------</td>
<td>-------------------</td>
<td>------------------------------------</td>
<td>--------------------------------</td>
</tr>
<tr>
<td>Primitive Reflex Profile (Capute et al.24)</td>
<td>PRP</td>
<td>infants</td>
<td>0 to 2 yrs.</td>
<td>D</td>
<td>9 primitive reflexes</td>
<td>cont.</td>
<td>Reflexes were selected from clinical observations of pathologically persisting primitive reflexes in children with CP / nda</td>
<td>Developmental pediatricians / nda</td>
</tr>
<tr>
<td>Infant Neurological International Battery (Ellison et al.25)</td>
<td>INFANIB</td>
<td>at-risk infants in NICU follow-up programs</td>
<td>1 to 18 mos.</td>
<td>D</td>
<td>20 items grouped into 5 categories: spasticity, vestibular function, head and trunk control, resting tone and description of motor behavior of the legs</td>
<td>cat.</td>
<td>Based on factor analyses of neuromotor behavior on 32 items selected from four existing assessments 24,26-28 in 308 NICU graduated infants / standardization based on these 308 infants</td>
<td>Professionals working in NICU follow-up programs / a few minutes</td>
</tr>
<tr>
<td>Bayley Scales of Infant Development, 2nd/3rd Ed. (Bayley29,30)</td>
<td>BSID-II/III</td>
<td>children</td>
<td>1 mo. to 3.5 yrs.</td>
<td>D, E</td>
<td>Motor scale (81 items, gross and fine motor behavior) (in addition: mental scale and behavior rating scale)</td>
<td>cont.</td>
<td>General maturationalist principles / standardization on 1700 US children aged 1 to 42 mos.</td>
<td>Psychologists / 15-20 minutes (motor scale only) total test: 25-60 minutes</td>
</tr>
<tr>
<td>Peabody Developmental Motor Scales, 2nd ed. (Folio and Fewell31)</td>
<td>PDMS-II</td>
<td>children</td>
<td>0 to 6 yrs.</td>
<td>D</td>
<td>Gross and fine motor scales subtests: reflexes, stationary, locomotion, object manipulation, grasping and visual-motor integration</td>
<td>cont.</td>
<td>nda / standardized on 2003 children</td>
<td>Occupational therapists / 45-60 minutes</td>
</tr>
<tr>
<td>Assessment (Author)</td>
<td>Short name</td>
<td>Population</td>
<td>Age group</td>
<td>Purpose</td>
<td>Test description - assessment of:</td>
<td>Type of variables</td>
<td>Test construction / standardization</td>
<td>Assessors / time to administer</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------</td>
<td>---------</td>
<td>----------------------------------</td>
<td>-------------------</td>
<td>------------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Movement Assessment of Infants (Chandler et al.³²)</td>
<td>MAI</td>
<td>high-risk infants in NICU follow-up clinics</td>
<td>0 to 12 mos.</td>
<td>D, E</td>
<td>65 test items grouped into 4 sections: muscle tone, primitive reflexes, automatic reactions and volitional movement</td>
<td>cont.</td>
<td>nda / a profile for typical motor behavior at 4 mos. has been developed based on 35 infants</td>
<td>Physical therapists, occupational therapists, physicians, nurses, psychologists and others who have specialized knowledge and experience in infant development / 20 to 30 minutes</td>
</tr>
<tr>
<td>Neuromotor Behavioral Inventory (Gorga and Stern³³)</td>
<td>NBI</td>
<td>infants</td>
<td>0 to 12 mos.</td>
<td>D</td>
<td>120 items grouped into categories on muscle tone, developmental motor abilities, quality of movement, reflexes and reactions, oral-motor behavior</td>
<td>cont.</td>
<td>nda / nda</td>
<td>nda / nda</td>
</tr>
<tr>
<td>Toddler and Infant Motor Evaluation (Miller and Roid³⁴)</td>
<td>TIME</td>
<td>children</td>
<td>4 mos. to 3.5 yrs.</td>
<td>D, E</td>
<td>Five primary subtests on mobility, motor organization, stability, social/emotional abilities and functional performance and three clinical subtests: quality rating, atypical positions and component analysis; focus on transitions between movement patterns</td>
<td>cont.</td>
<td>Subdomains developed by a panel of experts, field testing and expert consultation / standardization on 144 children with moderate to severe developmental delays and 731 TD children</td>
<td>Physical and occupational therapists or other clinicians with expertise in assessment of motor scales / 10-55 minutes</td>
</tr>
<tr>
<td>Assessment (Author)</td>
<td>Short name</td>
<td>Population</td>
<td>Age group</td>
<td>Purpose</td>
<td>Test description - assessment of:</td>
<td>Type of variables</td>
<td>Test construction / standardization</td>
<td>Assessors / time to administer</td>
</tr>
<tr>
<td>---------------------</td>
<td>------------</td>
<td>------------</td>
<td>-----------</td>
<td>---------</td>
<td>----------------------------------</td>
<td>------------------</td>
<td>-----------------------------------</td>
<td>----------------------------------</td>
</tr>
<tr>
<td><strong>Observation of motor behavior</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alberta Infant Motor Scale (Piper and Darrah&lt;sup&gt;35&lt;/sup&gt;)</td>
<td>AIMS</td>
<td>infants</td>
<td>0 to independent walking</td>
<td>D</td>
<td>58 items in four postural positions: prone, supine, sitting and standing. Test items are scored as observed or not observed based on drawings</td>
<td>cont.</td>
<td>Evaluation of the sequential development of postural control relative to four postural positions / standardization on 2202 sex and age stratified Canadian FT infants</td>
<td>Pediatric physical therapists / 15 minutes</td>
</tr>
<tr>
<td>Structured observation of motor performance (Persson and Strömberg&lt;sup&gt;36&lt;/sup&gt;)</td>
<td>SOMP-I</td>
<td>high-risk infants</td>
<td>0 to 10 mos.</td>
<td>D</td>
<td>13 ascending scales of motor development for each body part in supine and prone position and in the whole body with the infant sitting, standing and in locomotion</td>
<td>cont.</td>
<td>Assessment of level of motor development and quality of movements / nda</td>
<td>Physical therapists, neonatologists, paediatric neurologists / nda</td>
</tr>
<tr>
<td><strong>Quality of motor behavior</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Test of Infant Motor Performance (Campbell&lt;sup&gt;37&lt;/sup&gt;)</td>
<td>TIMP</td>
<td>infants</td>
<td>birth (32 wks. PMA) to 4 mos.</td>
<td>D</td>
<td>42 items grouped into two sections: observed and elicited sections</td>
<td>cont.</td>
<td>Some items from neurological, neurobehavioral and motor assessments&lt;sup&gt;33,37-39&lt;/sup&gt; / nda</td>
<td>Physical and occupational therapists / nda</td>
</tr>
<tr>
<td>General Movements (Einspieler et al.&lt;sup&gt;40&lt;/sup&gt;, Hadders-Algra&lt;sup&gt;41&lt;/sup&gt;)</td>
<td>GMs</td>
<td>infants</td>
<td>birth to 4 mos.</td>
<td>D, P</td>
<td>Assessment of variability and complexity of spontaneous motor behavior in supine position</td>
<td>cat.</td>
<td>Neuronal group selection theory principles / nda</td>
<td>No specific profession, but training courses are required to become a skilled observer / 3-minute video</td>
</tr>
</tbody>
</table>

cat. = categorical data, cont. = continuous data, D = discriminative, E = evaluative, FT = full-term, mos. = months, nda = no data available, NICU = neonatal intensive care unit, P = predictive, PT = preterm, TD = typically developing, USA = United States of America, yrs. = years.
Validity and reliability

In Table III data on the different kinds of validity and reliability of the selected methods are presented. Extended versions of Table III on validity and reliability can be found in Appendix II or on the journal’s website. Construct validity, the extent to which the items of the instruments reflect neuromotor function, was moderate to very good for most instruments. For three instruments no data were available. No data were available on concurrent validity with other methods to assess neuromotor function for eight instruments; for the other seven the range was from moderate to very good. Studies on the predictive validity of the instruments for CP or minor developmental disorders showed good predictive validity for six instruments, moderate predictive validity for four instruments and no data were described for five instruments. Data were available on intra-observer reliability for only four assessments. Two instruments had a very good and two had a moderate intra-observer reliability. For eleven instruments information on interobserver agreement was available: it ranged from moderate to very good.

Table III: Validity and reliability of the tests

<table>
<thead>
<tr>
<th>Assessment (short name)</th>
<th>Construct validity</th>
<th>Concurrent validity</th>
<th>Predictive validity</th>
<th>Intra-observer agreement</th>
<th>Inter-observer agreement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Touwen 19,74</td>
<td>nda</td>
<td>nda</td>
<td>+</td>
<td>nda</td>
<td>nda</td>
</tr>
<tr>
<td>Amiel-Tison 20,42-46</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>nda</td>
<td>±</td>
</tr>
<tr>
<td>Muscle power 47-49</td>
<td>++</td>
<td>nda</td>
<td>+</td>
<td>nda</td>
<td>nda</td>
</tr>
<tr>
<td>HINE 22,50-52</td>
<td>+</td>
<td>nda</td>
<td>+</td>
<td>nda</td>
<td>+</td>
</tr>
<tr>
<td>PRP 24,53</td>
<td>–</td>
<td>–</td>
<td>nda</td>
<td>nda</td>
<td>+</td>
</tr>
<tr>
<td>Infanib 25,54,55</td>
<td>+</td>
<td>nda</td>
<td>±</td>
<td>nda</td>
<td>nda</td>
</tr>
<tr>
<td>BSID-II 29,56</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>nda</td>
<td>+</td>
</tr>
<tr>
<td>PDMS-II 31,56</td>
<td>±</td>
<td>±</td>
<td>±</td>
<td>nda</td>
<td>++</td>
</tr>
<tr>
<td>MAI 32,57,62</td>
<td>nda</td>
<td>nda</td>
<td>±</td>
<td>±</td>
<td>±</td>
</tr>
<tr>
<td>NBI 54,63</td>
<td>+</td>
<td>nda</td>
<td>nda</td>
<td>nda</td>
<td>nda</td>
</tr>
<tr>
<td>TIME 64</td>
<td>±</td>
<td>nda</td>
<td>nda</td>
<td>nda</td>
<td>++</td>
</tr>
<tr>
<td>AIMS 66-67</td>
<td>±</td>
<td>++</td>
<td>±</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td>SOMP-f 68,69</td>
<td>+</td>
<td>nda</td>
<td>±</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>TIMP 77,70-73</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>nda</td>
<td>+</td>
</tr>
<tr>
<td>GMs 40,41,74-78</td>
<td>++</td>
<td>+</td>
<td>++</td>
<td>++</td>
<td>++</td>
</tr>
</tbody>
</table>

+++ = very good, + = good, ± = moderate, – = poor, nda = no data available, see Table I for description of evaluation criteria. Extended versions of Table III on validity and reliability can be found in Appendix II.
Chapter 2

DISCUSSION

We will discuss the fifteen instruments according to their classification into 1) comprehensive neurological examination, 2) procedures with standardized scoring, 3) observation of milestones and specific aspects of motor behavior and 4) assessment of the quality of motor behavior.

Comprehensive neurological examinations are mainly based on traditional neuropediatric concepts. They have a good construct validity and a good predictive validity for the development of major motor disorders such as CP or the prediction of future locomotor function. Virtually no information was available on the predictive validity of neurological examinations for more subtle, minor developmental disorders such as MND and DCD. The exception to this rule was the Touwen examination for which some data were available showing that MND at school age may be predicted to a limited extent. It is interesting to note that little information was available on the reliability of these frequently used assessments.

The second group of instruments consists of procedures with standardized scoring, which are quite heterogeneous and comprise the assessment of primitive reflexes (Primitive Reflex Profile (PRP)), a compilation of tests of pre-existing neurological assessments (Infant Neurological International Battery (Infanib)) and more extensive test batteries like the Motor Scales of the Bayley (BSID-II/III) and the Peabody Developmental Motor Scales (PDMS-II). Three tests, the Movement Assessment of Infants (MAI), Neuromotor Behavioral Inventory (NBI) and Toddler and Infant Motor Evaluation (TIME), combine neurological test items with observation of specific aspects of motor behavior. In general, the procedures with standardized scoring have a poor to moderate construct validity and concurrent validity. Predictive validity was either only moderate or no data were available. This might be a point of concern, but it should be realized that these tests have not been developed with the aim of predicting future motor disorders. The Bayley Scales, MAI and TIME can be used for the evaluation of changes in neuromotor functioning. Another attractive feature of the Bayley and PDMS-II is their standardization for very large groups of children. Interobserver reliability of most of these assessments is good to very good.

Two instruments consist mainly of observation of milestones and specific aspects of motor behavior: the Alberta Infant Motor Scale (AIMS) and the Structured Observation of Motor Performance (SOMP). Construct validity for both instruments is acceptable. For the SOMP, additional validity data are lacking. Concurrent validity for the AIMS is good, but predictive validity for major developmental disorders is only moderate. Reliability of these observational instruments is satisfactory.

The last two assessments, the General Movement method (GM) and the Test of Infant Motor Performance (TIMP) share the feature that they both assess quality of motor behavior or motor patterns. The GM method assesses quality of spontaneous motor behavior in supine position. The TIMP differs from the GM method in that not only does it focus on spontaneous movements but also mainly assesses quality of postural adjustments elicited by handling the infant. Interestingly,
both methods have a good predictive validity, the TIMP for major developmental disorders such as CP, and the GM method for both major and minor developmental disorders such as MND. Construct and concurrent validity and reliability of the GMs and TIMP are also satisfactory. These last two assessments are only useful under the age of four months.

**CONCLUDING REMARKS**

The main issue in choosing a suitable instrument in a certain situation is defining the goal that the instrument needs to serve. If the main goal is discriminating between infants with deviant neuromotor function and infants falling in the range of typical development, all fifteen reviewed instruments can be used. However, only three of the instruments (BSID-II\(^{29,30}\), MAI\(^{32}\) and TIME\(^{34}\)) can be used for evaluation of the effect of intervention. Data on predictive validity are available for ten instruments. The predictive validity for most of them is moderate at best. Only the two instruments that assess qualitative aspects of motor behavior (TIMP\(^{37}\) and GM-method\(^{40,41}\)) show good predictive validity. Prediction of developmental outcome at an early age will never be perfect. This is inherent in the developmental characteristics of the young brain\(^4\). Therefore, in medical evaluation of high-risk infants, the best prediction is achieved when multiple, complementary clinical tools are used. A good combination is medical history including achieved milestones, physical and neurological examination, a specific assessment of the quality of motor behavior and results of neuroimaging such as ultrasound or MRI assessment.
REFERENCES

Evaluation of neuromotor function in infancy


71. Kolobe TH, Bulanda M, Susman L. Predicting motor outcome at preschool age for infants tested at 7, 30, 60, and 90 days after term age using the Test of Infant Motor Performance. Phys Ther 2004;84:1144-1156.


