Neonatal assessments for the preterm infant up to 4 months corrected age: a systematic review

YOLANDE NOBLE[1,2] | ROSLYN BOYD[2]

1 Physiotherapy Department, Royal Children's Hospital, Brisbane, Australia. **2** Queensland Cerebral Palsy and Rehabilitation Research Centre, School of Medicine, Faculty of Health Sciences, University of Queensland, Herston, Australia.

Correspondence to Mrs Yolande Noble at 7th floor, Block 6, Royal Brisbane and Womens Hospital, The University of Queensland, Queensland 4029, Australia. E-mail: yolande_noble@health.qld.gov.au

PUBLICATION DATA
Accepted for publication 2nd November 2010.
Published online 5th December 2011.

ABBREVIATIONS

APIB	Assessment of Preterm Infants' Behaviour
NBAS	Brazelton Neonatal Behavioural Assessment Scale
NAPI	Neurobehavioural Assessment of the Preterm Infant
NMBA	Neuromotor Behavioural Assessment
NNNS	Neonatal Intensive Care Unit Network Neurobehavioural Scale
GMs	Prechtl's Assessment of General Movements
TIMP	Test of Infant Motor Performance

AIM The aim of this study was to systematically review the clinimetric properties of longitudinal neonatal neurobehavioural and neuromotor assessments for preterm infants.
METHOD Twenty-seven assessment measures were identified. The following eight measures met the study inclusion criteria: Assessment of Preterm Infants' Behaviour (APIB), Neonatal Intensive Care Unit Network Neurobehavioural Scale (NNNS), Test of Infant Motor Performance (TIMP), Prechtl's Assessment of General Movements (GMs), Neurobehavioural Assessment of the Preterm Infant (NAPI), Dubowitz Neurological Assessment of the Preterm and Full-term Infant (Dubowitz), Neuromotor Behavioural Assessment (NMBA), and the Brazelton Neonatal Behavioural Assessment Scale (NBAS). The primary purposes included prediction (TIMP, GMs, Dubowitz), discrimination (all assessments), and evaluation of change (TIMP, NAPI). Measures of assessment were included in the study if they were (1) primarily neurobehavioural or neuromotor assessments that were suitable for use with preterm infants (<37 weeks gestation) up to 4 months corrected age and were discriminative, predictive, or evaluative; (2) standardized procedures designed for serial/longitudinal use; or (3) criterion or norm referenced. However, all assessment tools that were not published in English in a peer-reviewed journal or were primarily neurological assessments, one-time evaluations, screening tools, or not commercially available were not used.
RESULTS All of the measures included in the review demonstrated adequate content and construct validity. Concurrent validity was reported for APIB, NNNS, Dubowitz, and GMs. Predictive validity was high for GMs with studies reporting up to 100% senstivity for predicting cerebral palsy at the age of 12 to 24 months. Interrater reliability was strong for the TIMP (intraclass correlation=0.95), GMs (K=0.8), and moderate for the NAPI (r=0.67–0.97). Clinical utility was variable for ease of scoring, interpretability, cost, and access.
INTERPRETATION In the absence of a criterion standard for neonatal neuromotor assessments, the NNNS and APIB have strong psychometric qualities with better utility for research. Similarly, the GMs, TIMP, and NAPI have strong psychometric qualities but better utility for clinical settings. The GMs has best prediction of future outcome and the TIMP has best evaluative validity.

The survival rate for the majority of very preterm infants (born <32 weeks gestation) now exceeds 85%.[1] However, more than 50% of these infants experience later neurobehavioural impairments, including motor incoordination, cognitive impairment, attention deficits, or behavioural problems.[2] In Australia there are approximately 2600 very-low-birthweight (<1500g) or very preterm (<30 weeks gestational age) survivors per annum. Approximately 10 to 15% of infants born very preterm will be diagnosed with cerebral palsy (CP). Furthermore, there is increasing evidence for sustained adverse outcomes into school age, adolescence, and adulthood.[3,4] This makes preterm birth a major public health issue.[5]

There is increasing evidence available regarding the impact of preterm birth on the developing brain, which is particularly vulnerable in the third trimester.[5] The early use of structural magnetic resonance imaging (MRI), combined with neurodevelopmental assessment, allows for the very early detection of brain lesions,[1,5] particularly for subtle injury in the periventricular white matter.[3] Despite advances in the care of preterm infants, it remains difficult to predict adverse neurodevelopmental outcome and subsequent disability accurately. Recently, the combination of structural brain MRI with standardized assessment of neuromotor development has demonstrated better prediction of future motor dysfunction.[6] Informed decisions regarding determination and measurement of early brain injury and the relationship between assessment domains and structural neural markers could, potentially, enable earlier targeted intervention.

Neonatal neurobehavioural and/or neuromotor assessments need to be valid, reliable, and designed for longitudinal use from the prenatal to the early postnatal period. Assessments that are suitable for use in the neonatal intensive care unit (NICU) setting on fragile and unstable infants are highly relevant to both clinician and researcher, making clinical utility a major factor in deciding which tool to use. A standardized assessment tool that has a consistent, documented set of procedures for administering, criterion testing, and scoring should be used to ensure that all infants are assessed under similar conditions.[7]

Neurobehavioural and neuromotor examinations are administered for a variety of purposes. These include the examination of the relationship between motor, neurological, and behavioural functioning, the detection of early central nervous system dysfunction, the prediction of future outcome, the evaluation of longitudinal development, and the impact of interventions.[7] The accurate discrimination of atypical development is essential to targeting early interventions to those most at risk and to prevent unnecessary intervention for those who are unlikely to have any neurodevelopmental impairments.

Traditional neonatal examinations vary according to gestational age, stability of the infant, and theoretical construct.[8] Some involve observing spontaneous motility or antigravity postures, with minimal or no handling and offline assessment from video recordings. Others include handling to elicit responses and to assess innate infant reflexes and muscle tone. Another domain of interest involves the behavioural state of the infant, social/attentional responses, and autonomic responses.

Longitudinal assessments are more predictive and useful than assessments administered at one time-point, as they provide information on maturation, recovery from injury, and reorganization. Multiple variables, especially physiological status, may have an impact on the reliability of performance, so a single snapshot of an infant's repertoire may not be helpful.[9] Longitudinal assessments help to build a picture of an infant's developmental trajectory and may give information on the effects of intervention.[10,11]

There have been several non-systematic reviews describing newborn or neonatal assessments,[12–14] and one systematic review evaluating motor assessments in the first 12 months of life from term.[7] However, there is no systematic review of longitudinal neurobehavioral and/or neuromotor assessments that has been specifically designed for use in the neonatal period (i.e. <37 weeks gestation) and into the early newborn period.

The purpose of this review was to identify systematically all published neonatal neurobehavioural and neuromotor assessments suitable for use in preterm infants up to 4 months corrected age or 4 months post term and to analyse the psychometric properties and clinical utility of these assessments. A secondary aim of major interest to clinicians was to determine which assessments have the best utility and feasibility for the clinical setting, depending on whether their primary purpose is discrimination, prediction, or evaluation.

What this paper adds

- To our knowledge, this is the first systematic review of the clinimetric properties of neuromotor assessments for preterm infants in the neonatal period.
- The validity, reliability, and clinical utility of this study's selected measures have been assessed by two independent raters who provided a comprehensive evaluation for clinicians and researchers.
- The results of this should aid clinicians and researchers in deciding which assessment is most suitable for their needs.

METHOD

Search strategy

In order to identify the key papers on this topic, a comprehensive search was undertaken of the following computerized databases: MEDLINE Advanced (1950–May 2010), CINAHL (1982–May 2010), PsycINFO (1806–May 2010), EMBASE (1980–May 2010), the Cochrane Library (May 2010), and PEDro (May 2010). The search strategy used included the MeSH terms and text words for 'infant-premature' OR 'infant-low birthweight' OR 'neonate' OR 'preterm' AND 'neurobehavioral or neuromotor or motor' AND 'outcome assessment' OR 'scale' OR 'test' OR 'evaluation' OR 'tool' OR 'neurological examination' AND 'clinimetric' OR 'psychometric' OR 'validity' OR 'reliability' OR 'reproducibility of results' OR 'sensitivity and specificity' OR 'clinical utility'.

After this initial search, a more specific search using the names of each identified measure and their authors were performed. References from key papers were also scanned to ensure that all key studies were included.

Inclusion criteria

Assessment tools were included if they were: (1) primarily neurobehavioural or neuromotor assessments suitable for use with preterm infants (<37 weeks gestation) up to 4 months corrected age that are discriminative, predictive, or evaluative; (2) standardized procedures designed for serial or longitudinal use; or (3) criterion or norm referenced.

Exclusion criteria

Assessment tools were excluded if they were: (1) not published in English in a peer-reviewed journal; (2) a manual that is not published or commercially available; (3) a screening tool or one-time evaluation; or (4) a primarily neurological assessment.

Data extraction and quality evaluation

The titles and abstracts of papers retrieved in the initial searches were screened independently by the two authors after removing duplicates. Assessments were included following agreement by both raters, and any conflicting viewpoints were discussed until a consensus was reached. A modified version of the Outcome Measures Rating Form (Appendix S1, published online)[15] including the International Classification of Functioning, Disability and Health (ICF) linking rules were used to determine the characteristics, clinical utility, and psychometric properties of the included assessment tools.[16,17] Assessment tools were classified as discriminative, predictive, or evaluative.[18]

Discriminative tools are used to distinguish between individuals with or without neurological and/or motor dysfunction, functional limitations, or disabilities at a point in time.[19] These are often norm-referenced tests designed to compare the infant's motor performance with a normative sample. Predictive tools are used to predict future neuromotor performance, condition, or outcome based upon current performance, and evaluative tools are used to document change over time or the efficacy of intervention.[7]

The psychometric properties of the assessment tools were evaluated for validity and reliability using previously reported definitions.[7,20] Validity included content, criterion, construct, prediction, and evaluation.[17,18,21] Reliability included test–retest, intrarater and interrater reliability, and internal consistency, with the latter defined as the extent to which multiple items contribute to a construct.[17,20] The appropriate statistics for reliability have been reported in earlier clinimetric reviews.[7,15,17,20]

The characteristics of the included measures were described for the purpose, age range, construct (criterion or norm referenced), normative sample, and domains tested. Clinical utility was rated for assessor and infant burden (time to administer, test procedure), training requirements and cost, availability of the manual and cost, as well as the time for scoring and interpretation. The feasibility of the assessment in a clinical setting on unstable preterm infants was of prime consideration when assessing clinical utility.

RESULTS

Twenty-seven measures were identified, eight of which met the study inclusion criteria (Table I). The included assessments were the Assessment of Preterm Infants' Behaviour (APIB),[22] the Neonatal Intensive Care Unit Network Neurobehavioural Scale (NNNS),[23,24] the Test of Infant Motor Performance, version 5 (TIMP),[25] Prechtl's Assessment of General Movements (GMs),[26] the Neurobehavioural Assessment of the Preterm Infant (NAPI),[27] the Dubowitz Neurological Assessment of the Preterm and Full-term Infant (Dubowitz),[9] the Neuromotor Behavioural Assessment (NMBA),[28] and the Brazelton Neonatal Behavioural Assessment Scale (NBAS).[10] Nineteen assessments were excluded because they did not meet all the inclusion criteria (Appendix S1).

Characterisitcs of included assessments

A summary of the characteristics of each assessment is reported in Table I. With respect to the ICF, the focus of measurement was on body functions, and items were linked to global mental functions (state of infant), sensory functions (responses to stimuli), and neuromuscular and movement-related functions (muscle tone, infant reflexes, and spontaneous and elicited movement patterns). The features of each measure varied depending on the primary construct. All of the assessments included the observation of spontaneous movements to varying degrees, but only the GMs classified the quality of movement patterns. Seven assessments included the traditional neurological components of elicited infant reflexes, antigravity postures, and muscle tone. Neurobehavioural domains were a major feature of most assessments apart from GMs and included responses to auditory and visual stimuli, irritability, and consolability (state). The APIB, NBAS, NNNS, and NMBA assessments all included items to evaluate the autonomic system, including responses (e.g. to colour) and stability of vital signs. The primary purpose of all assessments was discriminative: only four were appropriate for use as predictive tools, and although two measures had reported evaluative validity (TIMP and NAPI), only TIMP is suitable for use as an evaluative tool or outcome measure. All were criterion referenced, with norms reported for Dubowitz (low-risk term infants), NAPI (low-risk preterm infants from 32 to 37 weeks gestation), NNNS (term infants), and TIMP (preterm infants from 34 weeks gestation to 4 months post term).

The age for commencement of assessment varied from the early preterm period to 4 months post term. All assessments could be used from 32 weeks gestation onwards, except the NBAS, which is suitable from 36 weeks postmenstrual age. The GMs has critical periods for assessment of specific movements with the writhing period merging into the fidgety period between 6 and 8 weeks post term. The end point for included measures ranged from term to 4 months post term.

Validity

Evidence for the validity of the included assessments is summarized in Table SI (published online). All assessments had adequate content validity and most were based on extensive literature reviews and observations of preterm and term infant behaviour by experts in the field. The APIB, NAPI, TIMP, and NMBA all had strong construct validity. The NAPI had weak correlations between neurobehavioural constructs and tests of physiological status for preterm infants.[29] All assessments were able to discriminate between preterm infants with neurological problems and typical development and/or those at high or low risk for future CP. The APIB has demonstrated strong concurrent validity with MRI and electroencephalography. Poorer scores on Dubowitz and NNNS were correlated with MRI abnormality. General movement at 1 month and 3 months is strongly correlated with white matter abnormalities on MRI at term, and GMs and the NBAS correlated with neurological examinations. The NAPI demonstrated concurrent validity with the Einstein Neonatal Neurobehavioural Assessment Scale, as did the TIMP, which also had a low correlation with the GMs.

Evidence for predictive validity is available for the Dubowitz, GMs, NBAS, and TIMP (Table II). The GMs has the best combination of sensitivity and specificity, with cramped synchronized movements up to 6 to 9 weeks post term being highly predictive of future bilateral spastic-type CP. Only the TIMP is correctly utilized as an evaluative measure.[30,31] The NAPI has been used to report longitudinal changes in scores with maturation over time, but this as a result of intervention.

Reliability

The evidence for reliability is summarized in Table SII (published online). The TIMP and GMs have strong inter- and in-

Table I: Characteristics of included assessments

Assessment tool	Age range (min weeks PMA: max post term)	Primary purposes	Expanded purpose	Subsystems/components tested	Type of test	Normative sample
APIB	28wks–1mo	Discriminative	Documents spectrum of preterm and term infants' neurobehavioural functioning/competence	Autonomic, motor, state, attention/interaction, self-regulation	Criterion	n/a
Dubowitz	30wks–4mo	Discriminative/predictive	Provides a detailed profile of neurological status and identifies infants with neurological abnormalities	Posture and tone, reflexes, movements, neurobehavioural responses	Criterion/norm	224 low-risk term infants, 37–42wks gestation only (optimality score), from UK
GMs	Preterm–4mo	Discriminative/predictive	Documents spontaneous movements to identify early CNS dysfunction	Movement patterns	Criterion	n/a
NAPI	32wks–term	Discriminative/evaluative	Measures the progression of neurobehavioural performance	Motor development and vigour, scarf sign, popliteal angle, attention/orientation,% asleep, irritability, vigour of cry	Criterion/norm	521 infants from 32–37wks gestation from USA
NBAS	36wks–6wks post term	Discriminative/predictive	Identifies full range of individual neurobehavioural functioning and identifies areas of difficulty	Autonomic, motor and reflexes, state, social/attentional	Criterion	n/a
NMBA	30–36wks	Discriminative	Identifies preterm infants developmentally at risk	Neurological (e.g. tone, reflexes), behavioural, autonomic, motor functions	Criterion	n/a
NNNS	30wks–4mo (46/48wks post term)	Discriminative	Assesses at-risk infants (particularly substance exposed), documenting neurological integrity and broad range of behavioural functioning	Neurological (tone, reflexes), behavioural, stress/abstinence items	Criterion/norm	125 term infants (38–41wks gestation) from USA
TIMP	32wks–4mo	Discriminative/evaluative/predictive	Evaluates motor control and organization of posture and movement for functional activities	Orientation head in space, response to auditory and visual stimuli, body alignment, limb movements	Criterion/norm	990 infants at risk of poor neurological outcome from 11 geographic locations across the USA

APIB, Assessment of Preterm Infants' Behaviour;[22] Dubowitz, the Dubowitz Neurological Assessment of the Preterm and Full-term Newborn Infant;[35] GMs, Prechtl's Assessment of General Movements;[36] NAPI, Neurobehavioural Assessment of the Preterm Infant;[37] NBAS, Brazelton Neonatal Behavioural Assessment Scale;[10] NMBA, Neuromotor Behavioural Assessment;[28] NNNS, Neonatal Intensive Care Network Neurobehavioural Scale;[23] TIMP, Test of Infant Motor Performance;[25] n/a, not available; PMA, postmenstrual age; CNS, central nervous system.

Table II: Evidence of predictive validity

Assessment tool	Outcome assessment	Sample characteristics	Age at outcome	Age at initial assessment	Sensitivity (%)	Specificity (%)	Correlations
Dubowitz	MRI (WMA/GMA)[58]	VLBW preterm infants (*n*=66)	4mo post term	Term	88 (NPV 92)	46 (PPV 34)	n/a
	INFANIB[62]	Preterm infants BW<1250g (*n*=47)	12mo corrected age	Term	n/a	n/a	Total factor scores F92.42=8.7; $p<0.001$; $R^2=0.71$
	Amiel-Tison Milani-Comparetti Examination of Motor Development BSITD II[59]	Low-risk infants BW >1000g (*n*=100)	12mo corrected age	30–33wks gestation 34–36wks gestation 37–40wks gestation	2 + deviant signs; 91 (NPV 92) 4 + deviant signs; 55 (NPV 74)	2 + deviant signs; 79 (PPV 76) 4=deviant signs; 97 (PPV 92)	$r=0.7602$ ($p<0.001$) with neurodevelopmental outcome
GMs	CP (*n*=19) or learning disability* (*n*=2)	Preterm infants (*n*=29)	12–36mo	26–62wks PMA	100	59.1	n/a
	CP or DQ <85 (*n*=60)	Preterm and term infants (*n*=130)	24mo	46–60wks PMA	95	96	n/a
	CP or DQ <85 (*n*=18)	Term infants (*n*=58)	24mo	38–42wks PMA 43–47wks PMA 48–56wks PMA	94	59 86 83	n/a
	CP or DQ <85 (*n*=31)	Preterm infants (*n*=65)	24mo	28–37wks PMA 38–42wks PMA 43–65wks PMA	90.6 100 96.2–100	57.6 64.5 74.2–98.8	n/a
	CP or DQ <85 (*n*=11)[7]	IUGR and term control (*n*=62)	24mo	Term 49–51wks PMA 54–56wks PMA	83.33 100[a] 100[a]	80 100[a] 93.0[a]	n/a
	AIMS, NSMDA, CP (*n*=5)[6]	Preterm infants (*n*=86)	12mo	1mo	82.8 AIMS (NPV=84.4) 86.7 NSMDA (NPV=93.8) 100.0 CP (NPV=100)	48.2 AIMS (PPV=45.3) 42.9 NSMDA (PPV=24.5) 40.0 CP (PPV=9.4)	0.31 0.24 0.2
				3mo	40.0 AIMS (NPV=72.3) 62.5 NSMDA (NPV=90.8) 100.0 CP (NPV=100.0)	85.5 AIMS (PPV=60.0) 85.5 NSMDA (PPV=50.0) 81.3 CP (PPV=0.3)	0.27 0.42 0.47
NBAS	Neuro exam MRI/CT EEG McCarthy Scales of Intelligence Behavioural Problems[63]	Preterm infants (*n*=209)	5y	36–38wks PMA 40–42wks PMA 44–46wks PMA	50–78 correctly classified as mild disability; 71–85 correctly classifies as severe disability	94–97	n/a
	Paediatrician classification of typical/atypical	Neurologically suspect term infants (*n*=53)	7y	3d post term	80	78.90	n/a
	Griffiths Developmental Scales[12]	Term infants (*n*=40)	4mo, 8mo, 12mo	3d post term 14d post term			State control items related to motor development at 8mo and 12mo

Table II: continued

Assessment tool	Outcome assessment	Sample characteristics	Age at outcome	Age at initial assessment	Sensitivity (%)	Specificity (%)	Correlations
TIMP	AIMS scores <5th centile (*n*=16) at 6mo	Preterm and term infants (*n*=96)	6mo	32wks PMA – 4mo CA (*z*-score-0.5 SD)[b]	62.5	77.4	0.37–0.67
	14 at 9mo		9mo		91.7	75.7	0.20–0.56
	12 at 12mo		12mo		45–92	68–78	0.32–0.55
	BOTMP (motor delay) (*n*=8)	Preterm and term infants (*n*=35)	4y 9mo	32wks PMA – 4mo CA (*z*-score 1.6SD)[b]	50	100	0.36 (BOTMP score)
	Gross Motor Delay (PDMS DQ >70) (*n*=12)[7]	Preterm and term infants (*n*=61)	4–5y	1mo (*z*-score 0.5 SD)[b]	33	94	0.43 (PDMS GMQ)
				2mo (*z*-score 0.5 SD)[b]	50	86	0.42 (PDMS GMQ)
				3mo (*z*-score 0.5 SD)[b]	72	91	0.65 (PDMS GMQ)

[a]Abnormal and absent fidgety movements were combined. [b]Cut-off score for typical versus atypical motor development. Dubowitz, The Neurological Assessment of the Preterm and Full-term Newborn Infant; GMs, Prechtl's Assessment of General Movements; NBAS, Brazelton Neonatal Behavioural Assessment Scale; TIMP, Test of Infant Motor Performance; CP, cerebral palsy; IUGR, intrauterine growth retardation; WMA, white matter abnormality; GMA, grey matter abnormality; INFANIB, Infant Neurological International Battery; BSITD-II, Bayley Scales of Infant and Toddler Development – Version II; AIMS, Alberta Infant Motor Scale; NSMDA, Neuro Sensory Motor Developmental Assessment; PDMS, Peabody Developmental Motor Scale; GMQ, Gross Motor Quotient; BOTMP, Bruininks-Oseretsky Test of Motor Proficiency; DQ, Developmental Quotient; VLBW, very low birthweight; BW, birthweight; PMA, postmenstrual age; CA, corrected age; NPV, negative predictive value; PPV, positive predictive value; n/a, not available; gest, gestation; MRI, magnetic resonance imaging; CT, computed tomography; EEG, electrocardiogram.
*North American usage: mental retardation.

trarater reliability utilizing the appropriate statistics, with adequate test–retest reliability. The NAPI has good interrater reliability and evidence for stability of the domains over time. The NBAS, Dubowitz, and APIB report percentage agreement for interrater reliability. These results should be interpreted with caution as percentage agreement and Pearson's correlations can overestimate the true reliability. Strong internal consistency has been reported for the TIMP using Rasch analysis on 990 infants. The NNNS has weak test–retest reliability.

Clinical utility

Clinical utility for the included measures is summarized in Table III. The time taken to administer the assessments varied from 10 minutes (Dubowitz) to 1 hour (NNNS, APIB). The APIB and NNNS require extensive formal training with accreditation on the NBAS as a prerequisite. The GMs requires formal training that is readily available but costly. The advantage of the GMs is that minimal handling is required. All other assessments require the infant to be handled for elicited items. The GMs, NBAS, Dubowitz, TIMP, and NAPI have commercially available manuals and are easily purchased. The NNNS, APIB, and NMBA have published procedures. The Dubowitz, NAPI, NMBA, and TIMP do not require formal training but it is recommended that examiners be very familiar with handling preterm infants and watch instructional DVDs where available.

DISCUSSION

In this systematic review, eight assessments were identified that were suitable for preterm infants up to 4 months corrected age. The primary purpose of these assessments is to discriminate or identify at-risk preterm infants by documenting the full spectrum of neurobehavioural and/or neuromotor functioning. The assessments measure multiple domains including observation of antigravity postures and/or quality of spontaneous movements and elicite items such as infant motor patterns or reflexes and muscle tone. Components of neurobehavioural performance include attention/orientation (visual/auditory responses), autonomic functioning (colour, vital signs), and state (irritability, consolability).

All the neonatal assessments included were based on the early work of Brazelton and Prechtl, with respect to neurobehaviour and spontaneous movements, as well as on the work of Dubowitz, Saint-Anne Dargassies, and Amiel-Tison regarding more traditional neurological domains. The more recent GMs assessment reflects a unique construct based on complex movement patterns endogenously produced by central pattern generators.[26,32]

The assessments with the strongest psychometric properties were the GMs and the TIMP. The data produced from these assessments are required for more accurately predicting outcome and for establishing the true effect of interventions. Interpretation of results can be difficult without preterm norms, with several of the assessments reporting norms based only on term-born infants. The NAPI and the TIMP are the exceptions with published preterm norms.

Table III: Clinical utility of included assessments

Assessment tool	Time to administer (minutes)	Test procedure	Manual/score sheet available	Formal training required?	Scoring	Interpretation of scores
APIB	30–60	Naturalistic observation, elicited items administered according to standardized instructions – graded sequence of six packages of increasingly vigorous environmental inputs	Als et al., Theory and Research in Behavioural Pediatrics (1982)	Yes – can take 1–2y – training centres in the USA	Yields six main system variables (means of 81 scores of six sets of systems scores), ranging from 1 to 9, and 26 additional variables. Three separate status scores: baseline, response and post assigned for four systems	Low scores (1–3) – well-modulated and well-organized behavioural regulation. High scores (7–9) – easily disorganised, poorly modulated behavioural regulation, low thresholds for stress
Dubowitz	10–15	Naturalistic observation, elicited items according to standard sequence	Manual available via Mac Keith Press, approximately Aus$95.00	No	34 items scored according to described criteria. Optimality score for term infants	Use of cut-off points of 10th and 5th centiles – scores of 1.0, 0.5, or 0 – sum optimality scores from individual items
GMs	10–30	Infants' spontaneous movements filmed with no interference	Manual available via Mac Keith Press US$55.00	Yes – 4–5d training with GMs trust	Require three general movement sequences for pattern recognition – classify normal versus abnormal (cramped synchronized, chaotic, poor repertoire) preterm to 6–9wks post term. Fidgety period 9–20wks classify normal, absent, abnormal	Individual developmental trajectory
NAPI	30	Naturalistic observation, elicited items and handling required – strict invariant sequence of item presentation	Manual and score sheets available via Child Development Media website $US80	Demonstration video and assessment of reliability, practise on at least 20 infants	Item scores (numeric range 1–6) for 19 items, and behavioural state ratings (rate 14 times) and summary ratings[30]	Transfer item scores and summary ratings to appropriate converted (centile) scores – compare cluster scores to norms
NBAS	20–30	Naturalistic observation, elicited items and handling required administered according to infant's state	Manual available via Mac Keith Press US$84.95	Yes – training sessions, complete 20 exams, must achieve agreement of 90% with scoring	Six main clusters/packages scored 1–9 (28 items), seven supplementary items scored 1–9, and 18 reflex/motor items scored 0–3	Data often reduced to summary scores conceptually or empirically derived, behavioural and reflex items can be reduced into seven clusters
NMBA	10–15	Naturalistic observation and elicited items according to standard sequence	Burns et al. (1997), instructions and score sheets available from Mater Mothers Hospital, Brisbane	No, but infant handling experience recommended	Four major domains, 18 items rated on scale ranging from 0 to 4	Scores combined and categorized into normal, suspect, or abnormal
NNNS	30	Naturalistic observation and elicited items in structured sequence and according to infant's state	Download from *Pediatrics* 2004; **113** (3)	Yes – training programmes available, practice until ready for certification	12 packages with 45 items (scores range 1–11) and 7 stress scale packages with 70 items (yes/no scores)	13 summary scores – higher scores for attention, quality of movement, and regulation=better performance; higher scores for non-optimal reflexes, excitability, and lethargy subscales=poorer performance

Table III: Continued

Assessment tool	Time to administer (minutes)	Test procedure	Manual/score sheet available	Formal training required?	Scoring	Interpretation of scores
TIMP	20–40	Observation of spontaneous items, elicited items administered according to standardized instructions	Manual available via http://www.thetimp.com website $US35, score sheets $62 for 25	No – instructional DVD and infant handling experience recommended	42 items: 13 dichotomous observed and 29 elicited – 4–7 level rating scale	Raw scores to age standards – average, low average, below average, far below average; centile ranks

APIB, Assessment of Preterm Infants Behaviour; Dubowitz, the Dubowitz Neurological Assessment of the Preterm and Full-term Newborn Infant; GMs, Prechtl's Assessment of General Movements; NAPI, Neurobehavioural Assessment of the Preterm Infant; NBAS, Brazelton Neonatal Behavioural Assessment Scale; NMBA, Neuromotor Behavioural Assessment; NNNS, Neonatal Intensive Care Network Neurobehavioural Scale; TIMP, Test of Infant Motor Performance.

Concurrent validity is inadequate for many neonatal assessments as there is no criterion standard/consensus in the literature. Traditional neurobehavioural assessments are limited as predictive tools because most report high sensitivity for detecting a problem or condition, but demonstrate low specificity and, therefore, a high rate of false positives. The risk lies in over-servicing some infants who may have a normal motor outcome. There is a need for assessments with high specificity to correctly identify those infants who are low risk and probably do not require early intervention.[8] A combination of the GMs and an MRI performed at term and at 3 months is recommended.[6]

Evaluative validity or responsiveness was poorly reported. It remains difficult to determine whether change is due to natural history and the variable performance of preterm infants. Furthermore, large numbers are required for studies to determine an effect size. There are inherent challenges in implementing intervention studies in neonatal intensive care units because many studies are underpowered with respect to detecting true differences beyond maturation. An evaluative assessment must determine whether scores that detect a clinically important change are due to external intervention beyond measurement error.[21] Only the TIMP has evidence to support evaluative validity from two randomized controlled trials.[25]

In contrast, the reliability data for many assessments were inadequate. This may reflect the inherent difficulties faced by suitably trained health professionals collecting data on fragile neonates with varying performance in the early preterm period. Unlike several other assessments used in the NICU, the GMs had excellent reliability data with the appropriate statistics, as scoring is taken from video footage of spontaneous voluntary movement and does not rely on handling the infant.

There are several clinical implications resulting from this review. Firstly, training may be prohibitive in terms of cost and time but is reflective of the need for skill-based learning in the unique environment of the NICU. Secondly, the potential burden on the physiological instability of the preterm infant and the amount of expert handling required for assessment needs to be considered.[22,23] Thirdly, many items of the TIMP reflect care-giving practices, so it is highly relevant for clinicians and carers.[25] Serial assessments to document change and build up a developmental trajectory are also very useful.

There are a few studies that have looked at the relationship between structural MRI and neurodevelopmental outcome using standardized neonatal assessments.[1,6,33] The findings of Brown et al.[33] strengthen the hypothesis that preterm infant behavioural functioning at term is related to cerebral development. Other studies have determined the ability of early assessments including GMs and the Dubowitz assessment to predict later neurodevelopmental outcome.[1,6,34]

The relationship of brain structure on MRI and neuromotor function in the early preterm period (30 weeks gestation) with early neuromotor or neurobehavioural assessments has not yet been determined. With the advent of magnetic resonance-compatible incubators, early imaging from the preterm period may provide the opportunity for earlier links between brain structure and function. A major focus of devel-

Table IV: Quality assessment summary

Assessment tool	Clinical utility	Scale construction	Standardization	Reliability	Validity				Overall utility
					Content	Construct	Criterion	Responsiveness	
APIB	A	E	E	A	A	E	E	NE	A
Dubowitz	E	E	A	P	A	NE	A	NE	A
GMs	E	n/a	E	E	A	E	A	n/a	E
NAPI	E	E	E	A	E	A	A	A	A
NBAS	A	E	E	P	A	E	E	A	A
NMBA	A	A	P	NE	A	A	NE	NE	P
NNNS	A	E	E	A	A	NE	NE	A	A
TIMP	E	E	E	E	E	E	E	A	E

APIB, Assessment of Preterm Infants Behaviour; Dubowitz, the Dubowitz Neurological Assessment of the Preterm and Full-term Newborn Infant; GMs, Prechtl's Assessment of General Movements; NAPI, Neurobehavioural Assessment of the Preterm Infant; NBAS, Brazelton Neonatal Behavioural Assessment Scale; NMBA, Neuromotor Behavioural Assessment; NNNS, Neonatal Intensive Care Network Neurobehavioural Scale; TIMP, Test of Infant Motor Performance; E, excellent; A, adequate; P, poor; n/a, not applicable; NE, no evidence available.

opmental follow-up of preterm infants in the last 20 years has been the detection and prediction of CP. Researchers and clinicians also require accurate and sensitive measures to detect and predict learning difficulties or pervasive developmental delay such as autistic spectrum disorder. As a result, early neonatal assessments with excellent psychometric properties will be required with either a combination of assessments that measure different constructs or assessments combined with advanced imaging procedures. Of equal importance will be the ability of assessments to detect or discriminate typically developing preterm infants. This ensures that resources are targeted more efficiently.

There are many aspects to take into account when choosing the most appropriate assessment to use for preterm infants. Clinicians wanting a valid and reliable neonatal assessment that predicts the likelihood of CP should choose the GMs, which, along with MRI at term, have the best combination of sensitivity and specificity in terms of prediction of outcome at 12 months of age. However, if a shorter discriminative and predictive assessment is required, then the Dubowitz may be the assessment of choice. Researchers looking at multiple neonatal neurobehavioural variables may prefer the APIB or the NNNS,[33] but need to recognize that the training required is extensive and may be inaccessible. A good overall assessment recommended for clinicians that is discriminative, predictive, and evaluative with strong psychometric properties is the TIMP. The primary purpose of the assessment chosen is important. The neurobehavioural components of the NBAS or NAPI,[10] for example, can assist parents and/or caregivers to understand their infant's cues, enhance bonding, and promote maternal and infant emotional well-being.

For clinical purposes the GMs, Dubowitz, NAPI, NBAS, and TIMP would all be suitable. However, for research purposes the APIB and NNNS would be the measures of choice. The GMs and TIMP can be utilized across both settings, having both excellent utility and psychometric properties. A quality summary (linked to the Modified Outcome Measures Rating Form) of each assessment has been summarized in Table IV.

CONCLUSION

In the absence of consensus on a criterion standard for neonatal assessment for at-risk preterm infants and with few preterm norms available, all assessments are criterion referenced. When considering the most appropriate measure, it is important to determine both its construct and its primary purpose. Longitudinal assessment allows tracking of developmental maturation in the transition from the preterm period to post-term life and is essential because of the variability of preterm infant responses. The GMs, TIMP, and the NAPI have the strongest psychometric properties with good clinical utility. The best predictive assessment is the GMs, but the TIMP has the best evaluative validity. The APIB and the NNNS have more utility for the research setting. There is good validity but poor reliability reported for the NNNS. Further research focusing on the best combination of neurobehavioural assessments for preterm infants combined with MRI will improve accurate prediction of future disability.

ACKNOWLEDGEMENTS

The authors acknowledge the NHMRC Career Development Award (473840) (RB), the Smart State Fellowship from the Queensland Government (RB), the Royal Children's Hospital Foundation, Queensland, and the UQ Foundation and the Mayne Bequest at The University of Queensland School of Medicine.

ONLINE MATERIAL

Additional material and supporting information may be found in the online version of this article.

REFERENCES

1. Woodward LJ, Anderson PJ, Austin NC, Howard K, Inder TE. Neonatal MRI to predict neurodevelopmental outcomes in preterm infants. *N Engl J Med* 2006; **355:** 685–94.
2. Anderson P, Doyle LW. Neurobehavioral outcomes of school-age children born extremely low birth weight or very preterm in the 1990s. *JAMA* 2003; **289:** 3264–72.
3. Adams-Chapman I. Insults to the developing brain and impact on neurodevelopmental outcome. *J Commun Disord* 2009; **42:** 256–62.

4. Stephens BE, Vohr BR. Neurodevelopmental outcome of the premature infant. *Pediatr Clin North Am* 2009; **56:** 631–46.
5. Mathur A, Inder T. Magnetic resonance imaging-insights into brain injury and outcomes in premature infants. *J Commun Disord* 2009; **42:** 248–55.
6. Spittle AJ, Boyd RN, Inder TE, Doyle LW. Predicting motor development in very preterm infants at 12 months' corrected age: the role of qualitative magnetic resonance imaging and general movements assessments. *Pediatrics* 2009; **123:** 512–7.
7. Spittle AJ, Doyle LW, Boyd RN. A systematic review of the clinimetric properties of neuromotor assessments for preterm infants during the first year of life. *Dev Med Child Neurol* 2008; **50:** 254–66 (erratum in *Dev Med Child Neurol* 2008; **50:** 800).
8. Snider LM, Majnemer A, Mazer B, Campbell S, Bos AF. A comparison of the general movements assessment with traditional approaches to newborn and infant assessment: concurrent validity. *Early Hum Dev* 2008; **84:** 297–303.
9. Dubowitz L, Ricciw D, Mercuri E. The Dubowitz neurological examination of the full-term newborn. *Ment Retard Dev Disabil Res Rev* 2005; **11:** 52–60.
10. Brazelton TB, Nugent J. Neonatal Behavioral Assessment Scale, 3rd edition. Clinics in Developmental Medicine No. 137. London: Mac Keith Press, 1995.
11. Barbosa VM, Campbell SK, Sheftel D, Singh J, Beligere N. Longitudinal performance of infants with cerebral palsy on the Test of Infant Motor Performance and on the Alberta Infant Motor Scale. *Phys Occup Ther Pediatr* 2003; **23:** 7–29.
12. Majnemer A, Mazer B. Neurologic evaluation of the newborn infant: definition and psychometric properties. *Dev Med Child Neurol* 1998; **40:** 708–15.
13. Majnemer A, Snider L. A comparison of developmental assessments of the newborn and young infant. *Ment Retard Dev Disabil Res Rev* 2005; **11:** 68–73.
14. Santos RS, Araujo APQC, Porto MAS. Early diagnosis of abnormal development of preterm newborns: assessment instruments. *J Pediatri* 2008; **84:** 289–99.
15. Law M. Outcome Measures Rating Form 2004. http://canchild.icreate3.esolutionsgroup.ca/en/canchildresources/resources/measrate.pdf (accessed 5 January 2011).
16. Cieza A, Geyh S, Chatterji S, Kostanjsek N, Ustun B, Stucki G. ICF linking rules: an update based on lessons learned. *J Rehabil Med* 2005; **37:** 212–8.
17. McDowell I. Measuring Health: A Guide to Rating Scales and Questionnaires, 2nd edition. New York: OUP, 1996.
18. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis* 1985; **38:** 27–36.
19. Tieman BL, Palisano RJ, Sutlive AC. Assessment of motor development and function in preschool children. *Ment Retard Dev Disabil Res Rev* 2005; **11:** 189–96.
20. Gilmore R, Sakzewski L, Boyd R. Upper limb activity measures for 5- to 16-year-old children with congenital hemiplegia: a systematic review. *Dev Med Child Neurol* 2010; **52:** 14–21.
21. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res* 2010; **19:** 539–49.
22. Als H, Butler S, Kosta S, McAnulty G. The Assessment of Preterm Infants' Behavior (APIB): furthering the understanding and measurement of neurodevelopmental competence in preterm and full-term infants. *Ment Retard Dev Disabil Res Rev* 2005; **11:** 94–102.
23. Lester BM, Tronick EZ, Brazelton TB. The Neonatal Intensive Care Unit Network Neurobehavioral Scale procedures. *Pediatrics* 2004; **113:** 641–67.
24. Lester BM, Tronick EZ. History and description of the Neonatal Intensive Care Unit Network Neurobehavioral Scale. *Pediatrics* 2004; **113:** 634–40.
25. Campbell SK. The Test of Infant Motor Performance. Test User's Manual Version 2.0. Chicago: Infant Motor Performance Scales, LLC, 2005.
26. Einspieler C, Prechtl HFR. Prechtl's assessment of general movements: a diagnostic tool for the functional assessment of the young nervous system. *Ment Retard Dev Disabil Res Rev* 2005; **11:** 61–7.
27. Korner AF, Constantinou JC. The Neurobehavioral Assessment of the Preterm Infant: Reliability and Developmental and Clinical Validity. New York, NY: Guilford Press, 2001.
28. Carmichael K, Burns Y, Gray P, O'Callaghan M. Neuromotor behavioural assessment of preterm infants at risk for impaired development. *Aust J Physiother* 1997; **43:** 101–7.
29. Snider L, Tremblay S, Limperopoulos C, Majnemer A, Filion F, Johnston C. Construct validity of the Neurobehavioral Assessment of Preterm Infants. *Phys Occup Ther Pediatr* 2005; **25:** 81–95.
30. Girolami GL, Campbell SK. Efficacy of a neuro-developmental treatment program to improve motor control in infants born prematurely. *Pediatrics* 1994; **6:** 175–84.
31. Lekskulchai R, Cole J. Effect of a developmental program on motor performance in infants born preterm. *Aust J Physiother* 2001; **47:** 169–76.
32. Hadders-Algra M. General movements: a window for early identification of children at high risk for developmental disorders. *J Pediatr* 2004; **145**(Suppl. 2): S12–8.
33. Brown NC, Inder TE, Bear MJ, Hunt RW, Anderson PJ, Doyle LW. Neurobehavior at term and white and gray matter abnormalities in very preterm infants. *J Pediatr* 2009; **155:** 32–8.
34. Spittle AJ, Brown NC, Doyle LW, et al. Quality of general movements is related to white matter pathology in very preterm infants. *Pediatrics* 2008; **121:** e1184–9.
35. Dubowitz L, Dubowitz V, Mercuri E. The Neurological Assessment of the Preterm and Full Term Infant. Clinics in Developmental Medicine No. 148. London: Mac Keith Press, 1999.
36. Einspieler CPH, Bos AF, Ferrari F, Cioni G. Prechtl's Method on the Qualitative Assessment of General Movements in Preterm, Term and Young Infants. Clinics in Developmental Medicine No. 167. London: Mac Keith Press, 2004.
37. Korner AF, Brown JV, Thom VA, Constantinou JC. The Neurobehavioral Assessment of the Preterm Infant, 2nd edition. California: Child Development Media, Inc., 2000.
38. DeGangi G. Critique of the Infanib: infant Neurological International Battery. *Phys Occup Ther Pediatr* 1995; **14:** 109–20.
39. Hammarlund K, Persson K, Sedin G, Stromberg B. A protocol for structured observation of motor performance in preterm and term infants. Interobserver agreement and intraobserver consistency. *Ups J Med Sci* 1993; **98:** 77–82.
40. Amiel-Tison C. Update of the Amiel-Tison neurologic assessment for the term neonate at 40 weeks corrected age. *Pediatr Neurol* 2002; **27:** 196–212.
41. Lacey JL, Rudge S, Rieger I, Osborn DA. Assessment of neurological status in preterm infants in neonatal intensive care and prediction of cerebral palsy. *Aust J Physiother* 2004; **50:** 137–44.
42. Allen MC, Capute AJ. Neonatal neurodevelopmental examination as a predictor of neuromotor outcome in premature infants. *Pediatrics* 1989; **83:** 498–506.
43. Bao XL, Yu RJ, Li ZS, Zhang BL. Twenty-item behavioral assessment for normal newborns in 12 cities in China. *Chin Med J* 1991; **104:** 742–6.
44. Gorga DF, Stern FM, Ross G, Nagler W. Neuromotor development of preterm and full-term infants. *Early Hum Dev* 1988; **18:** 137–49.
45. Lipkin PH, Altshuler LA. Early outcome determination of low-birth-weight infants using the neurodevelopmental risk examination. *Clin Pediatr* 1994; **33:** 398–403.
46. Morgan AM, Koch V, Lee V, Aldag J. Neonatal neurobehavioral examination. A new instrument for quantitative analysis of neonatal neurological status. *Phys Ther* 1988; **68:** 1352–8.
47. Wolf MJ, Beunen G, Casaer P, Wolf B. Neonatal neurological examination as a predictor of neuromotor outcome at 4 months in term low-Apgar-score babies in Zimbabwe. *Early Hum Dev* 1998; **51:** 179–86.
48. Daily DK, Ellison PH. The premie-neuro: a clinical neurologic examination of premature infants. *Neonatal Netw* 2005; **24:** 15–22.
49. Campbell SK, Swanlund A, Smith E, Liao P, Zawacki L. Validity of the TIMPSI for estimating concurrent performance on the Test of Infant Motor Performance. *Pediatrics* 2008; **20:** 3–10.
50. DeGangi GA, Berk RA, Valvano J. Test of motor and neurological functions in high-risk infants: preliminary findings. *J Dev Behav Pediatr* 1983; **4:** 182–9.
51. Molteno CD, Thompson MC, Buccimazza SS, Magasiner V, Hann FM. Evaluation of the infant at risk for neurodevelopmental disability. *S Afr Med J* 1999; **89:** 1084–7.
52. Cheng CM, Chapman JS. Assessment of reliability and validity of the behavioral observation record for developmental care. *Nurs Res* 1997; **46:** 40–5.
53. Stuberg WA, White PJ, Miedaner JA, Dehne PR. Item reliability of the Milani-Comparetti Motor Development Screening Test. *Phys Ther* 1989; **69:** 328–35.
54. Bottos M, Dalla Barba B, D'Este A, Tronick EZ. The Neurobehavioral Assessment Scale as an instrument for early long-term prognosis and intervention in major disability in high-risk infants. *J Pediatr Psychol* 1996; **21:** 755–69.
55. Heineman KR, Bos AF, Hadders-Algra M. The Infant Motor Profile: a standardized and qualitative method to assess motor behaviour in infancy. *Dev Med Child Neurol* 2008; **50:** 275–82.
56. Sell EJ, Figueredo AJ, Wilcox TG. Assessment of preterm infants' behavior (APIB): confirmatory factor analysis of behavioral constructs. *Infant Behav Dev* 1995; **18:** 447–57.
57. Als H, Duffy FH, McAnulty GB, et al. Early experience alters brain structure and function. *Pediatrics* 2004; **113:** 846–57.
58. Woodward LJ, Mogridge N, Wells SW, Inder TE. Can neurobehavioral examination predict the presence of cerebral

injury in the very low birth weight infant? *J Dev Behav Pediatr* 2004; **25:** 326–34.

59. Molteno CD, Grosz P, Wallace P, Jones M. Neurological examination of the preterm and full-term infant at risk for developmental disabilities using the Dubowitz Neurological Assessment. *Early Hum Dev* 1995; **41:** 167–76.

60. Hyman C, Snider LM, Majnemer A, Mazer B. Concurrent validity of the Neurobehavioural Assessment for Preterm Infants (NAPI) at term age. *Pediatr Rehabil* 2005; **8:** 225–34.

61. Rose RU, Westcott SL. Responsiveness of the Test of Infant Motor Performance (TIMP) in infants born preterm. *Pediatrics* 2005; **17:** 219–24.

62. Bozynski ME, DiPietro MA, Meisels SJ, Plunkett JW, Burpee B, Claflin CJ. Cranial sonography and neurological examination at term and motor performance through 19 months of age. *J Dev Behav Pediatr* 1993; **14:** 112–6.

63. Ohgi S, Arisawa K, Takahashi T, et al. Neonatal behavioral assessment scale as a predictor of later developmental disabilities of low birth-weight and/or premature infants. *Brain Dev* 2003; **25:** 313–21.

64. Als H, Duffy FH, McAnulty GB. The APIB, an assessment of functional competence in preterm and full-term newborns regardless of gestational age at birth: II. *Infant Behav Dev* 1988; **11:** 319–31.

65. Mutlu A, Einspieler C, Marschik PB, Livanelioglu A. Intra-individual consistency in the quality of neonatal general movements. *Neonatology* 2008; **93:** 213–6.

66. Costas Moragas C, Fornieles Deu A, Botet Mussons F, Boatella Costa E, de Caceres Zurita ML. Psychometric evaluation of the Brazelton Scale in a sample of Spanish newborns. *Psicothema* 2007; **19:** 140–9.

67. Campbell SK. Test-retest reliability of the Test of Infant Motor Performance. *Pediatrics* 1999; **11:** 60–6.