Supplementary Table S2. Current evidence of neurodevelopmental assessment tool for motor function

	Reliability (ICC, Cohen's kappa)	Internal consistency (Cronbach's alpha)	Content validity +, ?, -, 0	Criterion validity (correlation coefficient, MCID, sensitivity/specificity)	Construct validity +, ?, -, 0
GMs	+++	0	+	+++	+
[1] n=20	Intrarater κ=1.00		+ Experts in field		Theory supported by ultrasound studies
[2] n=130				CP or DQ <85 outcome at 24 mo Sensitivity 0.95 (89.49%–100.0%) Specificity 0.96 (90.96%–100.0%)	
[3] n=58				CP or DQ <85 outcome at 24 mo Sensitivity/specificity according to initial assessment of GMs 38–42 wk: 0.94/0.59 43–47 wk: 0.94/0.86 48–56 wk: 0.94/0.83	
[4] n=31				CP or DQ <85 outcome at 24 mo Sensitivity/specificity according to initial assessment of GMs 28–37 wk: 90.6/57.6 38–42 wk: 100/64.5 43–65 wk: 96.2–100/74.2–98.8	
[5] n=30	Interrater Item speed/amplitude/variability in arm pattern κ=0.51–0.84 Item variability in leg pattern, general pattern: 0.36–0.40				
[6] n=19	Interrater κ=0.92			Neurological outcome at 24 mo Yates trend test, p<0.02	
[7] n=27	Interrater κ=0.84			Neurological outcome at 24 mo R=0.561, p<0.005	
[8] n=16	Interrater κ=0.91				
[9] n=37	Interrater ICC=0.80–0.94 Intrarater ICC=0.727-0.792				
[10] n=84	Inter-observer agreement=90.2%			CP at 24–36 mo Sensitivity/specificity according to initial assessment of GMs (%) 28–37 wk: 100/38 38–42 wk: 100/41 43–46 wk: 100/53 47–60 wk: 100/82	
TIMP	+++	+	+	+	+
[1]			+ Expert panel Literature review		Rasch analysis

(Continued to the next page)

Supplementary Table S2. Continued

	Reliability (ICC, Cohen's kappa)	Internal consistency (Cronbach's alpha)	Content validity +, ?, -, 0	Criterion validity (correlation coefficient, MCID, sensitivity/specificity)	Construct validity +, ?, -, 0
[2] n=96				Term and preterm infants 3 mo: AIMS r=0.646 Prediction of AIMS <5%ile Sensitivity/specificity according to age (day) 7 day: 0.45/0.74 30 day: 0.64/0.68 60 day: 0.75/0.78 90 day: 0.92/0.76	
[3] n=35				Delay on BOT Sensitivity 50.0% (15.6%–85.7%) Specificity 100%	
[4] n=61				PDMS DQ <70 Sensitivity/specificity (%) 1 mo: 33 (19–47)/94 (87–100) 2 mo: 50 (35–65)/86 (76–96) 3 mo: 72 (59–83)/91 (83–99)	
[5] n=21	Intrarater ICC=0.98–0.99 Interrater ICC=0.95				
[6] n=655	Interrater α=0.65–0.99 Intrarater α=0.97–0.99 Test-retest r=1.00, ICC=1.00	α=0.71–0.98	+ High concordance among the experts, for language clarity and pertinence, were found	Discriminant validity: to identify groups at-risk for delays (p<0.0001) Predictive validity: capability to predict motor delay (r=0.55–0.89)	
MAI	+	+	?	+- More recent studies report more favorable outcomes	+
[1]			Educational experience, literature review of general observations of infants movement	Cut-off score : suspect 8–13 : high risk >13	No survey; item selections by authors High risk score study (n=35) to establish interrater reliability >0.90
[2] n=160				BSID-motor r=-0.67 (4 mo), -0.68 (8 mo) Cut-off ≥10; sensitivity/specificity 4 mo: 83.3%/78.2% 8 mo: 96.0%/64.5%	
[3] Review	Fair for sum of risk scores Interrater r=0.72 Test-retest r=0.76	Interrater/test-retest Muscle tone κ=0.19–0.61/0.05–1.00 Reflex κ=-0.03–0.67/-0.05–0.87 Automatic κ=0.18–0.70/0.01–0.75 Volitional κ=0.09–0.79/-0.04–0.97		BSID (1–2 yr); -0.32 to -0.42 (poor) PDMS; -0.12 (not correlated) Recommend to use MAI together with norm-referenced test Volitional–Griffiths mental developmental r=-0.63, Reflex–Griffiths locomotor subscale r=-0.61	

(Continued to the next page)

Supplementary Table S2. Continued

	Reliability (ICC, Cohen's kappa)	Internal consistency (Cronbach's alpha)	Content validity +, ?, -, 0	Criterion validity (correlation coefficient, MCID, sensitivity/specificity)	Construct validity +, ?, -, 0
[4] n=50					Too stringent, not accurately reflect neuromotor behavior → necessary of deleting/revising items
[5] n=134				Cut-off (term; at 4 mo) >13: sensitivity 0.49, specificity 0.77 >9: sensitivity 0.63, specificity 0.53 BSID r=-0.23	
[6] n=75				Cut-off (preterm; at 4 mo) >4: sensitivity 100%, specificity 80% >9: sensitivity 66.6%, specificity 94.1%	
[7] Review				4 mo (n=246), BSID (PDI) (r=-0.63) 4 mo (n=160), BSID-III at 18 mo Sensitivity 69%, specificity 53%, PPV 56%, NPV 67% 4 mo (n=134), BSID at 2 yr–MAI >13 Sensitivity 49%, specificity 77%, PPV 50%, NPV 76%, AUC 0.629 4 mo (n=34), CP–MAI >10 Sensitivity 64%, specificity 91%, PPV 78%, NPV 84% 4–8 mo (n=160), BSID at 18 mo r=-0.58 to -0.44 Sensitivity 83.3%–96.0%, specificity 64.5%–78.2%, PPV 51.9%–59.2%, NPV 84.9%–90.8%	
AIMS	+++	+	+	++	+
[1], [2] Norm n=2,200	(n=506) Interrater ICC 0.9959-0.997 Stability ICC 0.859-0.993		Neuromaturational model of motor development	PDMS GMQ; r=0.90–0.99 BSID motor scale; r=0.84–0.97	Psychometric analysis for measuring small increments of change in the motor development
[3] n=164				10th %ile at 4 mo Sensitivity 77.3%/specificity 81.7% 5th %ile at 8 mo Sensitivity 86.4%/specificity 93.0%	
[4] n=650	Stability (20 yr dataset); r=0.99 (comparison with [1])				
[5]	ICC 0.98–0.99 (even without training, but only with given manual)				
[6] Serbia n=60	Interrater ICC 0.860-0.997 Stability ICC 0.860-0.997	0–3 mo α=0.940 4–7 mo α=0.998 8–14 mo α=0.996			

(Continued to the next page)

Supplementary Table S2. Continued

	Reliability (ICC, Cohen's kappa)	Internal consistency (Cronbach's alpha)	Content validity +, ?, -, 0	Criterion validity (correlation coefficient, MCID, sensitivity/specificity)	Construct validity +, ?, -, 0
HINE	-	-	+	+++	?
[1], [2]			Literature review Optimal score development from 12, 18 mo normal children (n=92) Score comparison of HIE (n=53)	1 yr; ≥67 → independent walk 40–67 → restricted mobility <40 → severely limited	Item selections for detecting abnormal signs; Further domain addendum through experts opinions
[3] Review n=10 Children =3,452				Reported cut-off 50–73; ≤56 (3 mo), ≤65 (12 mo) CP Sensitivity/specificity >90%; 90% predictive accuracy of CP (>5 mo CA)	
[4]				CP, GMFCS level r=-0.82	
[5]	Recommend education to establish inter-rater reliability (not sufficient evidence)				
PDMS-II	+++	+++	+	+++	+
[1] New norms (n=2,003)	Test-retest r=0.82–0.93 Inter-rater r=0.96–0.99 Excellent for all subsets, all ages	24–35 mo α=0.97, 36–47 mo α=0.95, 48–59 mo α=0.97, 60–71 mo α=0.98 Total α=0.9934	Literature review & created by experts and feedback from therapists Hierarchical sequence of items (age-equivalent basal, ceiling level)	Mullen Scales of Early Learning GMQ p=0.86 FMQ p=0.80	Conventional item analysis Modern differential item functioning analysis to new norms Age, subtest raw scores correlation exists
[2] Taiwan n=141	ICC 0.97	α=0.89		MABC-2; p=0.80–0.84 BOT-2; p=0.84–0.88 MCID 8.39; sensitivity 61.65% /specificity 71.34%	Psychometric properties (3–6 yr)
[3] China n=184		α=0.89–0.97		MABC-2 fine motor–FMQ; r=0.750 MABC-2 gross motor–GMQ; r=0.743 MABC-2 total–TMQ; r=0.631	
[4] Iran n=88	ICC 0.982 (GMQ)/0.989 (FMQ) ICC 0.988 (TMQ)	α=0.93 (GMQ)/α=0.90 (FMQ) α=0.92 (TMQ)		BSID-2 motor scale FMQ r=0.91; GMQ 0.93	
[5] New Mexico n=110				BSID-2 motor scale–Reflex; r=0.89 (n=13) Stationary; r=0.93 (n=110) Locomotion; r=0.97 (n=110) Object manipulation; r=0.90 (n=95) Grasping; r=0.85 (n=110) Visual/motor integration; r=0.94 (n=110) GMQ r=0.75 FMQ r=0.67 TMQ 0.76	

(Continued to the next page)

Supplementary Table S2. Continued

	Reliability (ICC, Cohen's kappa)	Internal consistency (Cronbach's alpha)	Content validity +, ?, -, 0	Criterion validity (correlation coefficient, MCID, sensitivity/specificity)	Construct validity +, ?, -, 0
[6] Korea n=50	ICC Stationary Korean version 0.985 Locomotion Korean version 0.994 Object manipulation Korean version 0.972 Grasping Korean version 0.941 Visual/motor integration Korean version 0.988		Review committee Lynn score for content validity 3.34	BOT-2 (Korean) Stationary r=0.756–0.904 Locomotion r=0.657–0.769 Object manipulation r=0.651–0.846 Grasping r=0.710–0.884 Visual-motor integration r=0.785–0.951	
NSMDA	+	+	+	+	+
[1]	Interrater r=0.80	Cross-correlation matrix item scoring (12+24 mo) 0.73 p<0.001 Functional grade (12+24 mo) 0.87 p<0.001	Literature review Developed by an experienced pediatric physiotherapist		Factor analysis (up to 2 yr of age) Stability of test results over time (up to 2 yr)
[2] n=148				Low-birthweight infants 24 mo: no significant difference between NSMDA and paediatrician's classification χ^2=0.08 Prediction of development at 24 mo, according to age at testing (%) 1 mo: sensitivity 68.8, specificity 72.6 4 mo: sensitivity 80.0, specificity 56.9 8 mo: sensitivity 82.4, specificity 83.7 12 mo: sensitivity 58.8, specificity 93.3	Factor analysis, discriminates between normal and abnormal development
[3] n=160				BSID-III motor <85 1 mo: r=-0.79 , p<0.001 4 mo: r=-0.79, p<0.001 8 mo: r=-0.93, p<0.001 12 mo: r=-0.88, p<0.001	
[4] n=132				NSMDA at 1 yr - McCarthy Scales of Children's Abilities at 4 yr: sensitivity 30%–57%, specificity 70%–98%, PPV 78%, NPV 87%	

ICC, intraclass correlation coefficient; MCID, minimal clinically important difference; GMs, general movements; CP, cerebral palsy; DQ, developmental quotient; TIMP, Test of Infant Motor Performance; AIMS, Alberta Infant Motor Scale; BOT, Bruininks-Oseretsky Test of Motor Proficiency; PDMS, Peabody Developmental Motor Scales; MAI, Movement Assessment of Infants; BSID, Bayley Scales of Infant Development; PDI, psychomotor developmental index; PPV, positive predictive value; NPV, negative predictive value; AUC, area under curve; GMQ, gross motor quotient; HINE, Hammersmith Infant Neurologic Examination; HIE, hypoxic ischemic encephalopathy; CA, corrected age; GMFCS, gross motor function classification system; FMQ, fine motor quotient; MABC-2, movement assessment battery for children-2; TMQ, total motor quotient; NSMDA, Neuro-sensory Motor Developmental Assessment.

SUPPLEMENTARY REFERENCES

[GMs]

[1] Einspieler C, Prechtl H, Bos A, Ferrari F, Cioni G. Prechtl's method on the qualitative assessment of general movements in preterm, term and young infants. London: Mac Keith Press; 2008.

[2] Prechtl HF, Einspieler C, Cioni G, Bos AF, Ferrari F, Sontheimer D. An early marker for neurological deficits after perinatal brain lesions. Lancet 1997;349:1361-3.

[3] Cioni G, Prechtl HF, Ferrari F, Paolicelli PB, Einspieler C, Roversi MF. Which better predicts later outcome in full-term infants: quality of general movements or neurological examination? Early Hum Dev 1997;50:71-85.

[4] Cioni G, Ferrari F, Einspieler C, Paolicelli PB, Barbani MT, Prechtl HF. Comparison between observation of spontaneous movements and neurologic examination in preterm infants. J Pediatr 1997;130:704-11.

[5] van Kranen-Mastenbroek V, van Oostenbrugge R, Palmans L, Stevens A, Kingma H, Blanco C, et al. Inter- and intra-observer agreement in the assessment of the quality of spontaneous movements in the newborn. Brain Dev 1992;14:289-93.

[6] Bos AF, van Loon AJ, Hadders-Algra M, Martijn A, Okken A, Prechtl HF. Spontaneous motility in preterm, small-for-gestational age infants. II. Qualitative aspects. Early Hum Dev 1997;50:131-47.

[7] Bos AF, Martijn A, Okken A, Prechtl HF. Quality of general movements in preterm infants with transient periventricular echodensities. Acta Paediatr 1998;87:328-35.

[8] Cioni G, Bos AF, Einspieler C, Ferrari F, Martijn A, Paolicelli PB, et al. Early neurological signs in preterm infants with unilateral intraparenchymal echodensity. Neuropediatrics 2000;31:240-51.

[9] Yeh KK, Liu WY, Wong AM, Chung CY, Lien R, Chuang YF. Intra-observer reliability of Prechtl's method for the qualitative assessment of general movements in Taiwanese infants. J Phys Ther Sci 2016;28:1588-94.

[10] Ferrari F, Cioni G, Einspieler C, Roversi MF, Bos AF, Paolicelli PB, et al. Cramped synchronized general movements in preterm infants as an early marker for cerebral palsy. Arch Pediatr Adolesc Med 2002;156:460-7.

[TIMP]

[1] Campbell SK. The Test of Infant Motor Performance. Test user's manual version. 2.0. Chicago: Infant Motor Performance Scales; 2005.

[2] Campbell SK, Kolobe TH, Wright BD, Linacre JM. Validity of the Test of Infant Motor Performance for prediction of 6-, 9- and 12-month scores on the Alberta Infant Motor Scale. Dev Med Child Neurol 2002;44:263-72.

[3] Flegel J, Kolobe TH. Predictive validity of the test of infant motor performance as measured by the Bruininks-Oseretsky test of motor proficiency at school age. Phys Ther 2002;82:762-71.

[4] 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;84:1144-56.

[5] Lekskulchai R, Cole J. Effect of a developmental program on motor performance in infants born preterm. Aust J Physiother 2001;47:169-76.

[6] Chiquetti EMDS, Valentini NC, Saccani R. Validation and reliability of the Test of Infant Motor Performance for Brazilian infants. Phys Occup Ther Pediatr 2020;40:470-85.

[MAI]

[1] Chandler LS, Andrews MS, Swanson MW. Movement assessment of infants: a manual. Rolling Bay: Moving Assessment of Infants; 1980.

[2] Swanson MW, Bennett FC, Shy KK, Whitfield MF. Identification of neurodevelopmental abnormality at four and eight months by the movement assessment of infants. Dev Med Child Neurol 1992;34:321-37.

[3] Palisano RJ. Review of research on reliability and validity of the movement assessment of infants. Pediatr Phys Ther 1989;1:167-72.

[4] Schneider JW, Lee W, Chasnoff IJ. Field testing of the movement assessment of infants. Phys Ther 1988;68:321-7.

[5] Rose-Jacobs R, Cabral H, Beeghly M, Brown ER, Frank DA. The Movement Assessment of Infants (MAI) as a predictor of two-year neurodevelopmental outcome for infants born at term who are at social risk. Pediatr Phys Ther 2004;16:212-21.

[6] Piper MC, Pinnell LE, Darrah J, Byrne PJ, Watt MJ. Early developmental screening: sensitivity and specificity of chronological and adjusted scores. J Dev Behav Pediatr 1992;13:95-101.

[7] Goo M, Tucker K, Johnston LM. Muscle tone assessments for children aged 0 to 12 years: a systematic review. Dev Med Child Neurol 2018;60:660-71.

[AIMS]

[1] Piper MC, Pinnell LE, Darrah J, Maguire T, Byrne PJ. Construction and validation of the Alberta Infant Motor Scale (AIMS). Can J Public Health 1992;83 Suppl 2:S46-50.

[2] Piper MC, Darrah J. Motor assessment of the developing infant. Philadelphia: Saunders; 1994.

[3] Darrah J, Piper M, Watt MJ. Assessment of gross motor skills of at-risk infants: predictive validity of the Alberta Infant Motor Scale. Dev

Med Child Neurol 1998;40:485-91.
[4] Darrah J, Bartlett D, Maguire TO, Avison WR, Lacaze-Masmonteil T. Have infant gross motor abilities changed in 20 years? A re-evaluation of the Alberta Infant Motor Scale normative values. Dev Med Child Neurol 2014;56:877-81.
[5] Blanchard Y, Neilan E, Busanich J, Garavuso L, Klimas D. Interrater reliability of early intervention providers scoring the Alberta Infant Motor Scale. Pediatr Phys Ther 2004;16:13-8.
[6] Lackovic M, Nikolic D, Filimonovic D, Petronic I, Mihajlovic S, Golubovic Z, et al. Reliability, consistency and temporal stability of Alberta Infant Motor Scale in Serbian infants. Children (Basel) 2020;7:16.

[HINE]

[1] Dubowitz L, Dubowitz V, Mercuri E. Neurological assessment of the preterm and fullterm newborn infant. 2nd ed. London: Mac Keith Press; 2007.
[2] Cioni G, Mercuri E. Neurological assessment in the first two years of life. London: Mac Keith Press; 2008.
[3] Romeo DM, Ricci D, Brogna C, Mercuri E. Use of the Hammersmith Infant Neurological Examination in infants with cerebral palsy: a critical review of the literature. Dev Med Child Neurol 2016;58:240-5.
[4] Romeo DM, Cioni M, Scoto M, Mazzone L, Palermo F, Romeo MG. Neuromotor development in infants with cerebral palsy investigated by the Hammersmith Infant Neurological Examination during the first year of age. Eur J Paediatr Neurol 2008;12:24-31.
[5] Maitre NL, Chorna O, Romeo DM, Guzzetta A. Implementation of the Hammersmith Infant Neurological Examination in a high-risk infant follow-up program. Pediatr Neurol 2016;65:31-8.

[PDMS-II]

[1] Folio M, Fewell R. (PDMS-2) Peabody developmental motor scales. 2nd ed. Austin: Pro-Ed; 2000.
[2] Wuang YP, Su CY, Huang MH. Psychometric comparisons of three measures for assessing motor functions in preschoolers with intellectual disabilities. J Intellect Disabil Res 2012;56:567-78.
[3] Hua J, Gu G, Meng W, Wu Z. Age band 1 of the Movement Assessment Battery for Children-Second Edition: exploring its usefulness in mainland China. Res Dev Disabil 2013;34:801-8.
[4] Tavasoli A, Azimi P, Montazari A. Reliability and validity of the Peabody Developmental Motor Scales-second edition for assessing motor development of low birth weight preterm infants. Pediatr Neurol 2014;51:522-6.
[5] Provost B, Heimerl S, McClain C, Kim NH, Lopez BR, Kodituwakku P. Concurrent validity of the Bayley Scales of Infant Development II Motor Scale and the Peabody Developmental Motor Scales-2 in children with developmental delays. Pediatr Phys Ther 2004;16:149-56.
[6] Lee JH, Kim KM, Chang M, Hong E. Study of validity and interrater reliability of Korean version of the Peabody Developmental Motor Scale 2. J Korean Acad Sens Integr 2019;17:14-25.

[NSMDA]

[1] Burns YR, Ensbey RM, Norrie MA. The Neuro-sensory Motor Developmental Assessment part 1: development and administration of the test. Aust J Physiother 1989;35:141-9.
[2] Burns YR, Ensbey RM, Norrie MA. The Neuro-Sensory Motor Developmental Assessment part II: predictive and concurrent validity. Aust J Physiother 1989;35:151-7.
[3] Kaya-Kara Ö, Kerem-Günel M, Yiğit Ş. Correlation of the Bayley scales of infant-toddler development-3rd edition and neuro-sensory motor assessment in preterm infants during the first year of life. Turk J Pediatr 2019;61:399-406.
[4] Burns Y, O'Callaghan M, McDonell B, Rogers Y. Movement and motor development in ELBW infants at 1 year is related to cognitive and motor abilities at 4 years. Early Hum Dev 2004;80:19-29.