# Clinical tools used in young infants born very preterm to predict motor and cognitive delay (not cerebral palsy): a systematic review

REBECCA CAESAR[1,2] | PAUL B COLDITZ[2,3] | GIOVANNI CIONI[4] | ROSLYN N BOYD[2]

**1** Allied Health Women's and Families Service, Sunshine Coast Hospital and Health Service District, Sunshine Coast University Hospital, Sunshine Coast, Queensland; **2** Faculty of Medicine, Child Health Research Centre, Queensland Cerebral Palsy and Rehabilitation Research Centre, The University of Queensland, Brisbane, Queensland; **3** Faculty of Medicine, The University of Queensland Centre for Clinical Research, Royal Brisbane and Women's Hospital, The University of Queensland, Brisbane, Queensland, Australia. **4** Department of Developmental Neuroscience, Stella Maris Scientific Institute, Pisa, Italy.

Correspondence to Rebecca Caesar at Paediatric Allied Health, Sunshine Coast University Hospital, PO Box 5340, Sunshine Coast MC, Qld 4560, Australia.
E-mail: rebecca.caesar@health.qld.gov.au.

**PUBLICATION DATA**
Accepted for publication 21st September 2020.
Published online

**ABBREVIATIONS**

| | |
|---|---|
| AUC | Area under the receiver operating characteristic curve |
| Bayley-III | Bayley Scales of Infant and Toddler Development, Third Edition |
| BSID-II | Bayley Scales of Infant Development, Second Edition |
| DOR | Diagnostic odds ratio |
| GMA | General Movement Assessment |
| HINE | Hammersmith Infant Neurological Examination |
| NAPI | Neurobehavioural Assessment of the Preterm Infant |
| NBAS | Neonatal Behavioural Assessment Scale |
| NPV | Negative predictive value |
| PPV | Positive predictive value |
| SROC | Summary receiver operating characteristic curve |

**AIM** This systematic review evaluates the accuracy of clinical tools used at a corrected age of 6 months or younger to predict motor and cognitive delay (not cerebral palsy) at 24 months' corrected age, in infants born very preterm.
**METHOD** Six databases were searched. Quality was evaluated using the Quality Assessment of Diagnostic Accuracy Studies tool. Predictive analysis included calculation of sensitivity and specificity, inspection of summary receiver operating characteristics curves, and bivariate meta-analysis.
**RESULTS** Six assessments were identified in 10 studies of 992 infants. Overall prevalence of motor delay was 13.8% and cognitive delay was 11.7%. Methodological quality was variable for patient selection, reference standard, flow, and timing. All studies had a low risk of bias for the index test. General Movement Assessment (GMA) predicted motor and cognitive outcomes with good accuracy for mild, moderate, and severe delays (fidgety age: pooled diagnostic odds ratio=12.3 [5.9–29.8]; hierarchical summary receiver operating characteristics curve=0.733). The Hammersmith Infant Neurological Examination (HINE) demonstrated excellent predictive accuracy for severe motor delay (3mo and 6mo; sensitivity 93% [68–100%], specificity 100% [96–100%]) but showed limited ability to predict milder delays.
**INTERPRETATION** In the population of infants born very preterm, few assessment tools used at 6 months or younger corrected age have proven predictive accuracy for cognitive and motor delay at 24 months' corrected age. Only the GMA and HINE demonstrated useful predictive validity.

In a cohort of infants born very preterm (<32wks) and at very low birthweight (<1500g), the likelihood of developmental delay in multiple domains is as high as 47%.[1–4] Cognitive impairments are evident in 25% to 50% of this population with moderate-to-severe deficits in academic achievement, attention problems, and executive functioning strongly correlated to preterm birth.[5,6] Infants born very preterm have an increased risk of mild motor delay not associated with cerebral palsy (CP).[7–9] Motor delays remain evident at school age with studies reporting an incidence of 34% of children born very preterm scoring equal to or less than the 15th centile on the Movement Assessment Battery for Children at 5 years' corrected age.[10]

Developmental delay is an umbrella term used to describe a suboptimal neurodevelopmental outcome that impacts function in one or more domain but does not completely limit a child's participation and is not associated with the characteristic motor types of CP. For assessment purposes, mild delay is most often defined as scores between 1 and 2SD below the mean, moderate delay by scores between 2 and 3SD below the mean, and severe delay as scores more than 3SD below the mean on standardized clinical assessment tools norm-referenced to corrected age.[1]

Clinicians need reliable early biomarkers to accurately predict motor and cognitive delays in infants born preterm.

Parents want to know if their child's development will be impacted by their preterm birth; funding for therapy or services is often dependent on the verification of risk for adverse neurodevelopmental outcomes. Early identification can facilitate an infant's timely entry into early intervention services that may harness brain plasticity and optimize functional outcomes.[11–13] Recently published guidelines for the early detection of CP are helping clinicians transition infants at 'high risk of CP' into intervention services by 6 months of age.[14] Infants at risk for milder deficits in the cognitive or motor domains could also benefit from early detection since early delays do not dissipate, rather they may become more apparent with increasing functional demands in educational and social contexts.[10,15–20]

To date, early prediction of milder motor and cognitive delays has been difficult.[13] Follow-up programmes for infants born very preterm typically provide blanket follow-up services using intermittent monitoring to identify motor and/or cognitive deficits as the child matures. Twenty-four months' corrected age is a standard time point for the identification of adverse outcomes in both clinical practice and research settings and is often recommended in clinical and research guidelines.[12,21,22] If clinical assessment tools used before 6 months' corrected age could robustly predict cognitive and motor deficits at 24 months' corrected age, tertiary services could facilitate earlier referral of infants into funded intervention providers and realize a cost-effective benefit.

Until now, no systematic review has examined the predictive accuracy of all available clinical tools used before 6 months of age to assist in diagnosing motor and/or cognitive delay at 24 months' corrected age. There are four existing systematic reviews that examine the predictive accuracy of the General Movement Assessment (GMA) to identify neurodevelopmental outcomes (outcomes reported between 1–10y).[23–26] There is one review of neurobehavioural tools used during the preterm period (outcomes at 6mo–10y)[27] and one review of neuromotor assessments used in the first year of life to predict cognitive outcomes (outcomes at 3mo–26y).[28] There are two clinimetric reviews that broadly evaluate neuromotor and neurobehavioural clinical assessment tools used before 6 months of age (with outcomes from 4mo–8y).[29,30]

Meaningful comparison of the predictive accuracy of assessment tools is problematic when outcome assessment measures and age at assessment are heterogeneous. Sensitivity and specificity, as well as concurrent validity of assessment tools, vary across the early lifespan and attrition bias inherently impacts studies of longer duration.[21,31–34] In previous reviews where outcomes have ranged from 3 months to 10 years, useful comparison between predictive assessments has been difficult. In this study, use of a single clinically relevant time point at outcome will improve the validity of comparisons between index tests.

The purpose of this review was to systematically evaluate the literature for all clinical assessments and tools (neuromotor, neurological, or neurobehavioural) that are used up to 6 months' corrected age to ascertain their accuracy for the diagnosis of motor and cognitive delay in infants born very preterm and at very low birthweight at 24 months' corrected age.

**What this paper adds**

- General movements have predictive validity for both motor and cognitive dysfunction in infants born very preterm.
- The Hammersmith Infant Neurological Examination showed the highest predictive accuracy for severe motor delay.
- The General Movement Assessment was the best tool to predict mild-to-moderate motor and cognitive delays.

## METHOD

### Search strategy

A limited search of computerized databases was undertaken, including PubMed (1966–January 2020), Cumulated Index to Nursing and Allied Health Literature (1982–January 2020), Scopus (1966–January 2020), Embase (1988–January 2020), PsycINFO (1927–January 2020), and the Cochrane Library (1972–January 2020). Limits applied included 'humans' and 'English language'. A PICO (problem/patient/population, intervention/indicator, comparison, and outcome) search strategy was implemented using MeSH headings and keywords for the target population of 'premature infants' with outcomes of 'motor' or 'cognitive delay'. Additional search terms (listed in Appendix S1, online supporting information) were added to identify studies with predictive index tests and limit the search to reference tests in the target age group (24 months' corrected age).

Subsequent searches were conducted combining key population search terms with each assessment tool identified in the initial search. Similar searches were conducted for standardized assessment tools known to the authors and not identified in the initial database search. Manual searching of all systematic reviews, targeted reference list scanning inclusive of unpublished theses, and citation tracking of key articles were used to minimize the chance of missing key studies.

### Inclusion and exclusion criteria

Studies were eligible for inclusion if they were written in the English language, conducted in humans, and where all the following criteria were met: the participant population was defined as infants born at 32 weeks or earlier and/or weighing 1500g or less; standardized neonatal clinical assessments or clinical tools with strong psychometric properties were administered as index tests before 6 months' corrected age; an applicable, valid, reliable, and standardized reference test identified cognitive and/or motor delays (not CP) at 24 months' corrected age.

Studies were excluded if they were case series, commentaries, randomized controlled trials, cross-sectional or descriptive studies, single case reports, and studies with a sample size <30 individuals. Predictors other than observational clinical tests did not meet the criteria. Predictive tests and variables excluded were perinatal and

physiological variables, neonatal disease scoring systems, and illness severity scores consisting of demographic, physiological, and perinatal data, magnetic resonance imaging and cranial ultrasound, electroencephalography, and other investigations.[35] Studies did not meet the criteria if they primarily assessed the outcomes of CP, genetic conditions, or progressive or neurodegenerative diseases. Studies were also excluded when outcomes were assessed before or after 24 months' corrected age, when participants in the same study were assessed on different outcome measures, and when predictive data could not be extracted.

### Methods of data extraction

The search strategy was jointly devised by RC and RB. RC and RB independently agreed on the selection of eligible studies, achieving consensus on which ones to include (Table S1, online supporting information). Consensus was reached between RC and RB on which data to extract from the studies included.

### Data extraction and quality assessment

The characteristics of all studies included in the review were collected including: type of study design; sample size; clinical assessment tool used as the index test; age at initial assessment; mean gestation; mean birthweight; sex; clinical assessments used as reference tests at 24 months' corrected age; and incidence and diagnostic description of any motor or cognitive delays identified in each study population (Table S2, online supporting information).

Data sought included sensitivity, specificity, negative predictive values (NPVs), and positive predictive values (PPVs) to predict motor and/or cognitive delays at 24 months' corrected age (Tables S3 and S4, online supporting information). Where possible, these were obtained from published data or otherwise extracted and calculated using 2×2 diagnostic contingency tables. Forest plots of sensitivity and specificity of all estimable index tests were generated in RevMan v5.3 (Cochrane, London, UK; Figs S1 and S2, online supporting information). Summary receiver operating characteristic curves (SROCs) were produced to compare aspects of sensitivity and specificity across index tests and age at assessment (Figs S3 and S4, online supporting information). Meta-analysis using the 'mada' package in R Commander (R Foundation for Statistical Computing, Vienna, Austria) was used for a subset of studies where the GMA was the common index test employed to predict developmental outcome. Accuracy of prediction of general movements for motor and/or cognitive delay at 24 months' corrected age was explored at writhing age (term to 8 weeks' corrected age) and at fidgety age (8–20 weeks' corrected age) using pooled diagnostic odds ratios (DORs) and bivariate random effects modelling to generate hierarchical SROCs (Figs S5–S8, online supporting information).

The overall quality of the studies included in the review was appraised using the revised Quality Assessment of Diagnostic Accuracy Studies tool.[36] The tool was applied to all studies to assess the risk of bias and applicability for patient selection, conduct and interpretation of the index and reference tests, and explore the risk of bias in terms of patient flow (Figs S9 and S10, online supporting information). The applicability of each study was assessed according to the extent that patients and study settings, the index test, its conduct and interpretation, and the target condition (as defined by the reference standard) matched the review question. Studies were rated according to the number of concerns ('unclear' and 'high' risk) identified across all domains (Table S2).

## RESULTS

A total of 859 records were retrieved based on the searches of six databases (Fig. S11, online supporting information).

Ten studies met all aspects of the inclusion criteria, comprising 992 infants born preterm at less than 32 weeks and/or weighing less than 1500g at birth.[21,37–45] Six clinical assessment tools were identified as eligible index tests including the GMA, the Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III), the Prechtl Neurological Examination, the Neonatal Behavioural Assessment Scale (NBAS), the Neurobehavioural Assessment of the Preterm Infant (NAPI), and the Hammersmith Infant Neurological Examination (HINE).

The characteristics of the studies included in the review and their population demographics are shown in Table S2. The population had gestational ages at birth varying from 25 to 32 weeks and birthweight ranging from 520g to 2035g. Gestational age could not be obtained in one study in a subgroup of infants born with a birthweight less than 1000g.[40] Sex was reported in 7 of the 10 studies with a median value for male sex of 50% (range 33–65%).[21,37,41–45] All eligible studies evaluated infants with a higher risk for motor and cognitive delay, having biased population selection for risk of adverse neurodevelopmental outcome brought about by very preterm birth and/or very low birthweight.

The timing of the initial assessment was variable. Index tests were completed between day 2 of life and 6 months' corrected age. All reference tests were completed at a mean age of 24 months' corrected age. Ninety per cent of infants were recruited in neonatal intensive care, 90% of index tests were completed within hospital inpatient or outpatient settings, and 70% of reference tests were administered in a hospital setting. One study was conducted in a community setting and two studies completed reference testing in the home environment.

In the studies that identified motor delay at 24 months' corrected age ($n$=8), the overall prevalence was 13.8% (137 children). Three studies used a cut-off point greater than or equal to 1SD below the mean of the reference test to identify 46 (4.6%) children with mild-to-severe delays.[21,42,43] Four studies used a cut-off point greater than or equal to 2SD below the mean to identify 56 (5.6%) children with moderate-to-severe motor delay.[38,39,42,44,45] Eye–hand coordination and fine motor delay was reported

in two studies in 22 children (2.2%).[21,37] One study used a grading classification to identify 15 (1.5%) non-ambulatory children.[44]

Cognitive delays were identified in eight studies with an overall reported incidence of 11.7% (116 children).[21,37,38,40–45] Three studies identified 32 (3.2%) children with mild-to-severe delays (cut-off point of 1SD) and three studies identified 46 (4.6%) children with moderate-to-severe cognitive delay (cut-off point of 2SD).[21,38,42,43,45]

## Methodological quality of the studies included in the review

### *Participant selection*

Four studies were considered as at low risk of bias for patient selection. Three were identified as at high risk of bias.[40,41,43] Two of the three studies deemed at high risk did not employ a random or consecutive recruitment design and two studies applied inappropriate participant exclusions, excluding infants with grade IV intraventricular haemorrhage or severe ventricular dilation, perinatal asphyxia, infections, and infants still on continuous positive airway pressure at term age.[40,41,43] The remaining three studies had an unclear risk of bias for patient selection with large numbers of infants incidentally excluded in the recruitment process and incomplete reporting of recruitment processes.[21,37,38]

There was high concern about the applicability of two studies where study design, including convenience sampling, matched groups, and patient exclusion criteria, indicated that patients may not have matched the review question.[40,41]

### *Index test*

All studies ($n$=10) were rated as at low risk of bias in the index test domain (six index tests). All studies made initial clinical judgements blinded to the reference standard applied at 24 months' corrected age. There were no concerns with the reporting of assessment thresholds.

Applicability was deemed as a high concern in the application of the index test in one study.[38] In this study, participants received the GMA at 1 and/or 3 months ($n$=121 at 1mo, $n$=164 at 3mo). Where the authors had missing data at fidgety age (3mo), they used data from the writhing assessment (1mo) to predict the 2-year outcome ($n$=26). This made comparison with the other studies difficult and the interpretation of accuracy for the prediction of outcome unclear.

### *Reference standard*

Seven studies were considered to have low risk of bias for the reference standard.[37,40–45] Each study judged as low risk used standardized, valid, and reliable clinical tests at 24 months' corrected age and reported that examiners had no knowledge of the results of the index tests. Three studies were judged as at unclear risk of bias where blinding of the reference standard to the index test was not clearly stated.[21,38,39]

Applicability of the reference standard to the review question was considered of high concern in one study where the reference standard broadly assessed motor outcomes of walking/not walking but could not detect mild or moderate motor delays.[44] Another study was deemed of high concern for applicability since motor and cognitive outcomes were compared in terms of lower mean scores when mean scores in all subgroups were mostly within the normal range.[37]

### *Flow and timing*

In terms of flow and timing, one study was considered as having a high risk of bias. Not all participants in this study received both motor and cognitive subtests on the reference test and 30% of patients who completed the index test did not receive the reference standard.[38] Three studies were judged to have an unclear risk of bias where smaller numbers of patients did not complete the reference standard or the number of patients who received the reference standard was not reported.[37,40,41] The remaining six studies were considered to have a low risk of bias.[21,39,42–45]

A visual summary of the methodological quality of all studies included in the review ($n$=10) is provided in Figures S9 and S10.

## Accuracy of prediction of cognitive and motor delay

### *GMA*

General movements are assessed from preterm until 20 weeks' corrected age within three distinct time periods: the preterm period (any gestational age until term age); writhing age (term age until 6–8wks); and fidgety age (6–9wks post-term continuing until 20wks).[46] Assessors employ either the Prechtl or Hadders-Algra assessment methods when defining and categorizing spontaneous movement.[47] Both methods are global, gestalt-based, dichotomous tests that assess normal versus abnormal movement. Although there is some variation in the terminology of subclassification between the two methods, the presence or absence of normal/abnormal movement is the overall predictor of outcome.[47] Sensitivity and specificity, NPV and PPV in this review, were calculated using normal versus abnormal general movements in each study, allowing comparison between methods.

Five studies, with a total of 347 infants born preterm, evaluated general movements for diagnostic accuracy of prediction of motor outcome at 24 months' corrected age.[37,39,42,43,45] Findings for sensitivity, specificity, PPV, and NPV of general movements at preterm, writhing, and fidgety ages to predict motor and cognitive outcomes at 24 months' corrected age are presented in Table S3 (online supporting information). A side-by-side forest plot of general movements at all time points was used to display point estimates and confidence intervals (CIs) of sensitivity and specificity for both motor and cognitive outcomes (Fig. S1). SROCs were generated from sensitivity and specificity for all studies that used general movements as

an index test at preterm, writhing, or fidgety age (Fig. S3).[37–39,42,43,45]

One study described the predictive accuracy of general movements in the preterm period to predict minor neurological outcomes (motor) with sensitivity of 80%, specificity of 50%, PPV of 21%, and NPV of 94%.[39] CIs for sensitivity were very wide (0.28–0.99) with inspection of the SROC suggesting that predictive accuracy at this time point was sufficient at best (Table S3 and Fig. S3).

An important finding was the similarity seen between summary curves for general movements to predict motor and cognitive delays at writhing and fidgety ages, suggesting comparable predictive validity across motor and cognitive domains. Estimating the area under the receiver operating characteristic curve (AUC) using a bivariate random effect model showed that accuracy for prediction of motor versus cognitive delays was non-discriminating to sufficient at writhing age (AUC motor=0.48, AUC cognitive=0.62) and good to very good at fidgety age (AUC cognitive=0.78, AUC motor=0.88).[48]

In one study that used mixed assessment time points for prediction, inspecting the AUC indicated that prediction of 'atypical outcome' using either writhing or fidgety age was sufficient (0.6–0.7).[38] However, the methodological quality of this study was judged as of high concern for applicability of the index test, rendering comparison with other studies inconclusive (Fig. S10).

Sensitivity and specificity for motor and cognitive outcomes was pooled across four studies to determine the overall DOR for general movements to predict motor and cognitive delay at 24 months' corrected age.[37,42,43,45] Meta-analysis of general movements at writhing age for 179 patients yielded a DOR of 1.53 (0.82–2.84) with a corresponding log(DOR) of 0.42 (−0.83 to 2.84) for the prediction of developmental delay in motor or cognitive domains at 24 months' corrected age (Fig. S5).[42,45] A hierarchical bivariate model was then used to calculate the hierarchical SROC, yielding an AUC of 0.51 (95% CIs plotted; Fig. S7). The log(DOR) was below 1 and the AUC non-discriminatory, suggesting that general movements at this age would be more useful to predict 'no delay' or 'ruling out' the risk for outcome.

At fidgety age, a meta-analysis for 222 patients who received general movements produced an DOR of 12.3 (5.87–25.78) with corresponding log(DOR) of 2.51 (1.77–3.25; Fig. S6).[37,43,45] The hierarchical SROC calculation yielded an AUC of 0.733 (95% CIs plotted; Fig. S8). The results of a meta-analysis at fidgeting age indicating good predictive accuracy of general movements for the identification of motor or cognitive dysfunction at 24 months' corrected age.

### HINE

One study of 189 infants born very preterm met the full inclusion criteria for the evaluation of the HINE as a predictor of motor and cognitive outcome at 24 months' corrected age.[44] A strong correlation was found between the HINE at 3 and 6 months and outcome on the visual–motor problem-solving subscale of the Cognitive Adaptive Test/Clinical Linguistic and Auditory Milestone Scale at 24 months' corrected age ($r$=0.75 with $p$<0.001 at 3mo; $r$=0.78 with $p$<0.001 at 6mo). Prediction of independent walking at 24 months' corrected age was also high with sensitivity of 93% (63–100%) at 3 and 6 months, specificity of 100% (96–100%) at 3 and 6 months, and SROC AUC indicating excellent predictive accuracy for ambulation (0.098 at 3mo, 0.983 at 6mo; Fig. S4).[44]

### Bayley-III

One study examined 30 high-risk infants born preterm to determine the effectiveness of the Bayley-III assessment to track development over the first 2 years of life.[21] Cognitive, fine motor, and gross motor subscales were administered at 3, 4, and 6 months and repeated at 24 months' corrected age. Predictive data were approximated and derived from line charts presented in the results, as well as a summary table showing the prevalence of delay at outcome. All subscales of the Bayley-III resulted in highly unstable delay classifications, low sensitivities, and poor PPVs across time (Table S4, online supporting information).

Subtests assessed from 3 to 6 months were generally strong at identifying typical development (high specificities and NPVs) but poor at identifying atypical development (low sensitivities and PPVs). The AUC confirmed the non-discriminative predictive accuracy for gross motor and cognitive delay at 24 months' corrected age (Fig. S4). The AUC for the prediction of fine motor outcome suggested good predictive accuracy but should be interpreted with caution since the CIs for sensitivity were very wide, rendering the findings almost meaningless.

### Prechtl Neurological Examination

The Prechtl Neurological Examination was applied as an index test at term-corrected age in a single study of 100 infants.[42] This is a standard neurological examination of posture, tone, and reflexes and is a distinctly different clinical tool to Prechtl's Method of Qualitative Assessment of General Movements.[42,46]

Two reference standards were applied at 2 months' corrected age: the optimality score of the Touwen Infant Neurological Examination and the Bayley Scales of Infant Development, Second Edition (BSID-II). The sensitivity, specificity, PPV, and NPV of the Prechtl Examination to predict suspect or abnormal classification of tone, posture, and motor function and outcome on the Psychomotor Developmental Index and Mental Developmental Index subscales is presented in Table S4. Paired forest plots highlighted low sensitivity and high specificity for prediction in both motor and cognitive domains (Fig. S2). Accuracy of prediction on the BSID-II at 24 months' corrected age, as indicated by the AUC, was fair for motor delay and non-discriminative for cognitive delay (Fig. S4).

### *Other index tests*

In a study of 170 infants born at fewer than 32 weeks, Harijan et al.[41] explored the interrater reliability, discriminative, construct, and predictive validity of the NAPI. Cognitive and motor development was assessed at 24 months' corrected age using the Mental Developmental Index and Psychomotor Developmental Index subscales of the BSID-II. Of the three NAPI subscales of motor development and vigour, alertness and orientation, and irritability, only irritability was weakly correlated with cognitive development on the BSID-II Mental Developmental Index scores ($r=-0.16$, $p<0.040$). No correlation was found with the BSID-II Psychomotor Developmental Index scores in the motor domain. Patient selection methods were judged as at high risk of bias and high concern for applicability in this study, further diluting any potential findings (Fig. S10). Overall, the study failed to find evidence of predictive validity of the NAPI for developmental outcomes at 24 months' corrected age.

The NBAS was applied as an index test at 37 weeks' corrected age by a trained neonatologist in one study of 46 small-for-gestational age infants (<1000g).[40] At 24 months, multivariate analysis of variance showed that motor outcome assessed by Psychomotor Developmental Index scores on the BSID-II was predicted by small-for-gestational age status, interaction of small for gestational age with birthweight, NBAS motor maturity subscale, and 12-month Psychomotor Developmental Index scores (32% of variance explained). No other data were available to further analyse the predictive accuracy of the NBAS. High risk of bias and high concerns for applicability of patient selection methods were also evident in this study (Fig. S10).

## DISCUSSION

Of the six index tests that met full inclusion criteria, the HINE demonstrated the highest accuracy for the prediction of severe motor outcome in terms of sensitivity and specificity (sensitivity=93%, specificity=100%). The reference test used in this study of 103 infants born very preterm was a dichotomous categorization of independent walking versus non-ambulatory function. This allowed identification of severe outcomes but was non-discriminative for mild or moderate motor delay. The applicability of the reference test was rated as of high concern since it is probable that such severe motor delay was associated with a concurrent or later diagnosis of CP. Consequently, the HINE, with its current cut-off points, is a highly predictive tool for severe motor delay (with or without a diagnosis of CP) but has limited clinical utility for the detection of milder motor outcomes.[14] In terms of prediction for cognitive outcomes, the HINE showed strong correlation with the Cognitive Adaptive Test of the Cognitive Adaptive Test/Clinical Linguistic and Auditory Milestone Scale suggesting that it may have utility for the prediction of cognitive delay. Further research to validate predictive accuracy for cognitive outcome is indicated.

All studies included in this review assessed the predictive validity of the GMA in terms of individual time points rather than the combined results of general movement trajectories. General movements at fidgety age demonstrated good-to-very good predictive accuracy for motor and cognitive delays at 24 months' corrected age. General movements at writhing age were non-discriminating to sufficient. General movements in the preterm period showed sufficient accuracy for the prediction of motor delay at 24 months' corrected age. No studies assessed the ability of general movements at preterm age to predict cognitive delay.

In clinical practice, longitudinal trajectories of general movements rather than discrete time points are used to track progress and predict outcome.[13] Criterion standard recommendations encourage clinicians to assess two or three videos in the preterm period, one video at writhing age, and two videos at fidgety age.[13,49] It is often a combination of clinical findings across different time points that allows improved interpretation of risk for outcomes such as CP.[25,50,51] This may also be important for the prediction of motor and cognitive delays that are non-CP related. Different combinations of normal and abnormal general movements across the preterm and writhing periods combined with movement assessment at fidgety age may improve accuracy for the prediction of milder cognitive or motor outcomes.[13] Alternatively, the strong NPVs consistently seen across preterm (NPV=94%), writhing (NPV=60–100%), and fidgety ages (NPV=89–96%) may indicate that where normal movement is consistently present across a trajectory, there is a high chance of a typical outcome in both motor and cognitive domains.[13,23,39]

The GMA is a criterion standard clinical tool for the prediction of neuromotor deficits including CP.[13,14,50] Its association with cognitive outcome is less well defined but there is increasing evidence to support its utility for the prediction of subsequent cognitive dysfunction.[23] An important finding of this systematic review is that general movements at fidgety age demonstrated predictive accuracy for cognitive delay at 24 months' corrected age, adding considerable support for the use of the GMA as a predictive tool for cognitive outcome.

Standardized testing is often embedded in preterm follow-up programmes to track progress and predict neurodevelopmental outcome. The Bayley-III is a widely used measure that assesses developmental function in the cognitive, language, and motor domains. The use of the Bayley-III below 6 months of age to predict gross motor and cognitive outcome was not supported by this review, which found that the assessment at 3, 4, and 6 months of age was non-discriminative for prediction at 24 months' corrected age.

The Prechtl Neurological Examination is a less known bedside neurological examination. When used at term age, accuracy of prediction was only fair for motor delay and non-discriminative for cognitive delay. Adding this assessment to clinical practice would be of little benefit if the prediction of outcome was the primary aim. Likewise, the NAPI showed no evidence of predictive validity and cannot be recommended as a predictive tool. The motor maturity

subscale of the NBAS showed modest predictive association with motor outcomes at 24 months' corrected age when used in combination with other assessment tools and perinatal risk factors. High risk of bias and applicability concerns in study design suggest cautious interpretation of the relevance of this finding. The validity of the NBAS as a stand-alone tool to predict motor outcome at 24 months' corrected age is unproven. Had data for individual patient meta-analysis been available, more precise estimates of associations would have been possible.

### Clinical implications

Evidence from this review highlights the limited number of clinical assessments that clinicians currently have in their toolbox to assist with early prediction of motor and cognitive delays in very preterm populations. The GMA remains the best choice for the prediction of motor and cognitive outcomes in this cohort. General movements are an essential tool for the early detection of CP with the added benefit of good predictive accuracy for milder outcomes in both motor and cognitive domains. As a purely observational assessment, it has excellent clinical utility with prerequisite training requirements that ensure good-to-excellent interrater reliability. Used in combination with other norm-referenced and standardized clinical tools, magnetic resonance imaging, and other investigative modalities, the GMA can provide parents with increased clarity in their understanding of their child's risk for adverse outcomes and assist clinicians with prioritization of service delivery and early intervention planning.

### Implications for research

Further research is indicated to explore the predictive accuracy of general movement trajectories to predict normal outcomes across developmental domains. Since prediction for milder delays in cognitive and motor domains is currently 'good' at best, it may be a more effective strategy to rule out risk for developmental delay. Prediction of normal outcome has the advantage of providing welcome reassurance to parents while facilitating flow of resources towards the smaller group of 'at-risk' infants.

Due to the complex nature of intertwined biological and environmental variables contributing to adverse outcomes, improving the accuracy of prediction of mild-to-moderate motor and cognitive delays will likely require a combination of clinical biomarkers and clinical tools used across developmental trajectories.

We recommend that future research incorporate a collection of the clinical assessments identified in this review as having the strongest associations with outcome at 24 months' corrected age, with research protocols designed to enable individual patient meta-analysis.

### Limitations

This review assessed index tests for diagnostic accuracy at 24 months' corrected age. There may be evidence for other clinical assessment tools used in the first 6 months of life with predictive validity for longer term outcomes.

The outcome assessments included in this review differ in their predictive accuracy for cognitive and motor deficits at preschool and school age.[52–54] Clinicians should be mindful of this limitation when considering the clinical implications of this review. Not all aspects of intelligence and motor outcome are predicted at 24 months' corrected age.[52,53]

Since studies included in this review were limited to high-risk populations of infants born very preterm and infants born at very low birthweight, applicability to term-aged infants or infants born late preterm is unknown.

Finally, the authors acknowledge possible limitations inherent in this review, given that an a priori protocol was not registered.

## CONCLUSION

Clinicians working with infants born very preterm require accurate biomarkers to assist in early diagnosis and facilitate timely transition to funded early intervention services. While international best practice guidelines are in place for the detection of CP, little is known about the relative predictive accuracy of assessments for motor and cognitive delay at 24 months' corrected age. The results of this review indicate that very few assessments before 6 months' corrected age have the accuracy to predict milder motor and cognitive delays at 24 months' corrected age. While the HINE assessment has excellent predictive ability for severe motor delay, it is non-discriminative for milder deficits. The GMA has the best predictive ability for mild and moderate motor and cognitive delay. Further research is warranted to investigate the efficacy of general movement trajectories to improve prediction of both mild-to-moderate delay and typical outcomes. Researchers could also explore the accuracy of other neuromotor, neurobehavioural, neurological, and neurosensory assessments or combinations of clinical assessments to predict outcomes at 24 months' corrected age.

## ACKNOWLEDGEMENTS

RC is partially funded by research grants from the Health Practitioner Research Scheme, Queensland Health, and by Wishlist (Sunshine Coast Health Foundation). RB is supported by a National Health and Medical Research Council Fellowship and a National Health and Medical Research Council Centre of Research Excellence (Australian Cerebral Palsy Clinical Trials network). The authors have no conflict of interest to declare regarding their own research interests or any of the studies included in this review. The declared funding source of each study included in the review is outlined in Table S5 (online supporting information).

### Data availability statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

## SUPPORTING INFORMATION

The following additional material may be found online:

**Appendix S1:** Search terms used in the review (MeSH or headings).

**Table S1:** List of studies excluded after full-text review

**Table S2:** Study characteristics: population demographics

**Table S3:** Sensitivity, specificity, PPV, and NPV of general movements at preterm, writhing, and fidgety age to predict motor and cognitive delay at 24 months' corrected age in infants born very preterm

**Table S4:** Sensitivity, specificity, PPV, and NPV of other clinical tests (not general movements) used at 6 months or earlier to predict motor and cognitive delay at 24 months' corrected age in infants born very preterm

**Table S5:** Funding source declared by all studies included in the review

**Figure S1:** Forest plot of sensitivity and specificity of general movements at preterm, writhing, and fidgety age to predict cognitive or motor delay at 24 months corrected age in very preterm infants. Key: GMs = General Movements; TP = True Positive; FP = False Positive; FN = False Negative; TN = True Negative.

**Figure S2:** Forest plot of sensitivity and specificity of HINE, Bayley III and Prechtl neurological examination used ≤ 6 months corrected age to predict cognitive or motor delay at 24 months corrected age in very preterm infants. Key: Bayley III = Bayley Scales of Infant and Toddler Development Third Edition; Prechtl = Prechtl Neurological Examination; HINE = Hammersmith Infant Neurological Examination; TP = True Positive; FP = False Positive; FN = False Negative; TN = True Negative.

**Figure S3:** SROC plot comparing the accuracy of general movements (GMs) at preterm, writhing, and fidgety age to predict motor and cognitive outcomes at 24 months corrected age in very preterm infants. Key: GMs = General Movements.

**Figure S4:** SROC plot comparing the accuracy of other clinical tests used ≤ 6 months to predict motor and cognitive outcomes at 24 months corrected age in very preterm infants. Key: Bayley III = Bayley Scales of Infant and Toddler Development Third Edition; Prechtl = Prechtl Neurological Examination; HINE = Hammersmith Infant Neurological Examination.

**Figure S5:** Forest plot of diagnostic odds ratio of General Movements at writhing age to predict motor and cognitive delay at 24 months corrected age in very preterm infants.

**Figure S6:** Forest plot of diagnostic odds ratio of General Movements at fidgety age to predict motor and cognitive delay at 24 months corrected age in very preterm infants.

**Figure S7:** HSROC curve, including confidence region, for general movements at writhing age to predict motor and cognitive outcomes at 24 months corrected age in very preterm infants. Key: o = summary estimate; Δ = data point; ___ = SROC curve; ___ = confidence region; GMs = General Movements.

**Figure S8:** HSROC curve, including confidence region, for general movements at fidgety age to predict motor and cognitive outcomes at 24 months corrected age in very preterm infants. Key: o = summary estimate; Δ = data point; ___ = SROC curve; ___ = confidence region; GMs = General Movements.

**Figure S9:** Summary of risk of bias and applicability concerns graph: qualitative judgements about each domain presented as percentages across included studies. All index tests included for comparison. Key: GMs = General Movements; NBAS = Neurobehavioural Assessment Scale; NAPI = Neonatal Assessment of the Preterm Infant; Bayley III = Bayley Scales of Infant and Toddler Development Third Edition; Prechtl Neurological = Prechtl Neurological Examination; HINE = Hammersmith Infant Neurological Examination.

**Figure S10:** Summary of risk of bias and applicability concerns: qualitative judgements about each domain for each included study. All index tests included for comparison. Key: GMs = General Movements; NBAS = Neurobehavioural Assessment Scale; NAPI = Neonatal Assessment of the Preterm Infant; Bayley III = Bayley Scales of Infant and Toddler Development Third Edition; Prechtl Neurological = Prechtl Neurological Examination; HINE = Hammersmith Infant Neurological Examination.

**Figure S11:** Flow chart of search strategy and selection process for publications reporting on clinical tests performed at ≤ 6 months CA to predict mild, moderate or severe motor and/or cognitive delay at 24 months CA. Key: AIMS = Alberta Infant Movement Scale; ASQ = Ages and Stages Questionnaire; BSID II = Bayley Scales of Infant Development II; Bayley III = Bayley Scales of Infant and Toddler Development Third Edition; DAYC/DAYC-2 = Developmental Assessment of Young Children Version First or Second Edition; DDST = Denver Developmental Screening Test; Dubowitz (HNNE) = Dubowitz Neurological Assessment/Hammersmith Neonatal, Neurological Examination; FTII = Fagan Test of Infant Intelligence; GMA = General Movement Assessment; HINE = Hammersmith Infant Neurological Examination; NAPI = Neonatal Assessment of the Preterm Infant; NBAS = Neurobehavioural Assessment Scale; NBRS = Nursery Neurobiologic Risk Score; NSMDA = Neuro Sensory Developmental Assessment; NNNS = NICU Network Neurobehavioural Scale (NNNS); NOMAS = Neonatal Oral-Motor Assessment Scale; Prechtl Neurological = Prechtl Neurological Examination; PNE = The Premie-Neuro; SNAP-II = Score for Neonatal Acute Physiology – II; TIMP = Test of Infant Motor Performance; Uzgiris- Hunt Scale = Uzgiris and Hunt Scales.

## REFERENCES


1. Doyle L, Roberts G, Anderson PJ, et al. Outcomes at age 2 years of infants <28 weeks' gestational age born in Victoria in 2005. *J Pediatr* 2010; **156**: 49–53.e1.
2. Doyle LW. Evaluation of neonatal intensive care for extremely-low-birth-weight infants. *Semin Fetal Neonatal Med* 2006; **11**: 139–45.
3. Orcesi S, Olivieri I, Longo S, et al. Neurodevelopmental outcome of preterm very low birth weight infants born from 2005 to 2007. *Eur J Paediatr Neurol* 2012; **16**: 716–23.
4. Schlapbach L, Adams M, Proietti E, et al. Outcome at two years of age in a Swiss national cohort of extremely preterm infants born between 2000 and 2008. *BMC Pediatr* 2012; **12**: 198.
5. Aarnoudse-Moens CS, Weisglas-Kuperus N, van Goudoever JB, Oosterlaan J. Meta-analysis of neurobehavioral outcomes in very preterm and/or very low birth weight children. *Pediatrics* 2009; **124**: 717–28.
6. Bhutta AT, Cleves MA, Casey PH, Cradock MM, Anand KJS. Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis. *JAMA* 2002; **288**: 728–37.
7. Spittle AJ, Cheong J, Doyle LW, et al. Neonatal white matter abnormality predicts childhood motor

impairment in very preterm children. *Dev Med Child Neurol* 2011; **53**: 1000–6.

8. Greene MM, Patra K, Nelson MN, Silvestri JM. Evaluating preterm infants with the Bayley-III: patterns and correlates of development. *Res Dev Disabil* 2012; **33**: 1948–56.
9. Månsson J, Stjernqvist K. Children born extremely preterm show significant lower cognitive, language and motor function levels compared with children born at term, as measured by the Bayley-III at 2.5 years. *Acta Paediatr* 2014; **103**: 504–11.
10. Van Hus J, Potharst ES, Jeukens-Visser M, Kok JH, Van Wassenaer-Leemhuis AG. Motor impairments in very preterm-born children: links with other developmental deficits at 5 years of age. *Dev Med Child Neurol* 2014; **56**: 587–94.
11. Moore T, Hennessy EM, Myles J, et al. Neurological and developmental outcome in extremely preterm children born in England in 1995 and 2006: the EPICure studies. *BMJ* 2012; **345**: e7961.
12. Caesar R, Boyd RN, Colditz P, et al. Early prediction of typical outcome and mild developmental delay for prioritisation of service delivery for very preterm and very low birthweight infants: a study protocol. *BMJ Open* 2016; **6**: e010726.
13. Peyton C, Einspieler C. General movements: a behavioral biomarker of later motor and cognitive dysfunction in NICU graduates. *Pediatr Ann* 2018; **47**: e159–e164.
14. Novak I, Morgan C, Adde L, et al. Early, accurate diagnosis and early intervention in cerebral palsy: advances in diagnosis and treatment. *JAMA Pediatr* 2017; **171**: 897–907.
15. Bhutta AT, Cleves MA, Casey PH, Cradock MM, Anand KJS. Cognitive and behavioral outcomes of school-aged children who were born preterm: a meta-analysis. *JAMA* 2002; **288**: 728–37.
16. de Kieviet JF, Piek JP, Aarnoudse-Moens CS, Oosterlaan J. Motor development in very preterm and very low-birth-weight children from birth to adolescence: a meta-analysis. *JAMA* 2009; **302**: 2235–42.
17. Edwards J, Berube M, Erlandson K, et al. Developmental coordination disorder in school-aged children born very preterm and/or at very low birth weight: a systematic review. *J Dev Behav Pediatr* 2011; **32**: 678–87.
18. Kerstjens JM, de Winter AF, Bocca-Tjeertes IF, Bos AF, Reijneveld SA. Risk of developmental delay increases exponentially as gestational age of preterm infants decreases: a cohort study at age 4 years. *Dev Med Child Neurol* 2012; **54**: 1096–101.
19. Rogers CE, Anderson PJ, Thompson DK, et al. Regional cerebral development at term relates to school-age social–emotional development in very preterm children. *J Am Acad Child Adolesc Psychiatry* 2012; **51**: 181–91.
20. Wild KT, Betancourt LM, Brodsky NL, Hurt H. The effect of socioeconomic status on the language outcome of preterm infants at toddler age. *Early Hum Dev* 2013; **89**: 743–6.
21. Lobo MA, Paul DA, Mackley A, Maher J, Galloway JC. Instability of delay classification and determination of early intervention eligibility in the first two years of life. *Res Dev Disabil* 2014; **35**: 117–26.
22. National Institute for Health and Care Excellence. Developmental Follow-up of Children and Young People Born Preterm [Internet]. NICE: London, 2017.
23. Einspieler C, Bos AF, Libertus ME, Marschik PB. The General Movement Assessment helps us to identify preterm infants at risk for cognitive dysfunction. *Front Psychol* 2016; **7**: 406.
24. Darsaklis V, Snider LM, Majnemer A, Mazer B. Predictive validity of Prechtl's method on the qualitative assessment of general movements: a systematic review of the evidence. *Dev Med Child Neurol* 2011; **53**: 896–906.
25. Burger M, Louw QA. The predictive validity of general movements: a systematic review. *Eur J Paediatr Neurol* 2009; **13**: 408–20.
26. Einspieler C, Prechtl HF, Ferrari F, Cioni G, Bos AF. The qualitative assessment of general movements in preterm, term and young infants: review of the methodology. *Early Hum Dev* 1997; **50**: 47–60.
27. Craciunoiu O, Holsti L. A systematic review of the predictive validity of neurobehavioral assessments during the preterm period. *Phys Occup Ther Pediatr* 2017; **37**: 292–307.
28. Oudgenoeg-Paz O, Mulder H, Jongmans MJ, van der Ham IJM, Van der Stigchel S. The link between motor and cognitive development in children born preterm and/or with low birth weight: a review of current evidence. *Neurosci Biobehav Rev* 2017; **80**: 382–93.
29. Noble Y, Boyd R. Neonatal assessments for the preterm infant up to 4 months corrected age: a systematic review. *Dev Med Child Neurol* 2012; **54**: 129–39.
30. 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.
31. Nunan D, Aronson J, Bankhead C. Catalogue of bias: attrition bias. *BMJ Evid Based Med* 2018; **23**: 21–2.
32. Hack M. Dilemmas in the measurement of developmental outcomes of preterm children. *J Pediatr* 2012; **160**: 537–8.
33. Connolly BH, McClune NO, Gatlin R. Concurrent validity of the Bayley-III and the Peabody developmental motor scale-2. *Pediatr Phys Ther* 2012; **24**: 345–52.
34. Yu Y-T, Hsieh W-S, Hsu C-H, et al. A psychometric study of the Bayley Scales of Infant and Toddler Development – 3rd Edition for term and preterm Taiwanese infants. *Res Dev Disabil* 2013; **34**: 3875–83.
35. Dorling JS, Field DJ, Manktelow B. Neonatal disease severity scoring systems. *Arch Dis Child Fetal Neonatal Ed* 2005; **90**: F11–6.
36. Whiting PF, Rutjes AWS, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med* 2011; **155**: 529–36.
37. Beccaria E, Martino M, Briatore E, et al. Poor repertoire General Movements predict some aspects of development outcome at 2 years in very preterm infants. *Early Hum Dev* 2012; **88**: 393–6.
38. De Bock F, Will H, Behrenbeck U, Jarczok MN, Hadders-Algra M, Philippi H. Predictive value of General Movement Assessment for preterm infants' development at 2 years: implementation in clinical routine in a non-academic setting. *Res Dev Disabil* 2017; **62**: 69–80.
39. de Vries NKS, Bos AF. The quality of general movements in the first ten days of life in preterm infants. *Early Hum Dev* 2010; **86**: 225–9.
40. Feldman R, Eidelman AI. Neonatal state organization, neuromaturation, mother-infant interaction, and cognitive development in small-for-gestational-age premature infants. *Pediatrics* 2006; **118**: e869–78.
41. Harijan P, Beer C, Glazebrook C, et al. Predicting developmental outcomes in very preterm infants: validity of a neonatal neurobehavioral assessment. *Acta Paediatr* 2012; **101**: e275–81.
42. Maas YG, Mirmiran M, Hart AA, Koppe JG, Ariagno RL, Spekreijse H. Predictive value of neonatal neurological tests for developmental outcome of preterm infants. *J Pediatr* 2000; **137**: 100–6.
43. Peyton C, Yang E, Msall ME, et al. White matter injury and general movements in high-risk preterm infants. *AJNR Am J Neuroradiol* 2017; **38**: 162–9.
44. Romeo DMM, Cioni M, Scoto M, Pizzardi A, Romeo MG, Guzzetta A. Prognostic value of a scorable neurological examination from 3 to 12 months post-term age in very preterm infants: a longitudinal study. *Early Hum Dev* 2009; **85**: 405–8.
45. Spittle AJ, Spencer-Smith MM, Cheong JLY, et al. General movements in very preterm children and neurodevelopment at 2 and 4 years. *Pediatrics* 2013; **132**: e452–8.
46. 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, 2004.
47. Kwong AKL, Fitzgerald TL, Doyle LW, Cheong JLY, Spittle AJ. Predictive validity of spontaneous early infant movement for later cerebral palsy: a systematic review. *Dev Med Child Neurol* 2018; **60**: 480–9.
48. Šimundić A-M. Measures of diagnostic accuracy: basic definitions. *EJIFCC* 2009; **19**: 203–11.
49. Adde L, Helbostad J, Jensenius AR, Langaas M, Støen R. Identification of fidgety movements and prediction of CP by the use of computer-based video analysis is more accurate when based on two video recordings. *Physiother Theory Pract* 2013; **29**: 469–75.
50. Bosanquet M, Copeland L, Ware R, Boyd R. A systematic review of tests to predict cerebral palsy in young children. *Dev Med Child Neurol* 2013; **55**: 418–26.
51. 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.
52. Spittle AJ, Spencer-Smith MM, Eeles AL, et al. Does the Bayley-III Motor Scale at 2 years predict motor outcome at 4 years in very preterm children? *Dev Med Child Neurol* 2013; **55**: 448–52.
53. Potharst ES, Houtzager BA, van Sonderen L, et al. Prediction of cognitive abilities at the age of 5 years using developmental follow-up assessments at the age of 2 and 3 years in very preterm children. *Dev Med Child Neurol* 2012; **54**: 240–6.
54. Bode MM, D'Eugenio DB, Mettelman BB, Gross SJ. Predictive validity of the bayley, third edition at 2 years for intelligence quotient at 4 years in preterm infants. *J Dev Behav Pediatr* 2014; **35**: 570–5.