ORIGINAL ARTICLE

Sensitivity and specificity of General Movements Assessment for diagnostic accuracy of detecting cerebral palsy early in an Australian context

Catherine Morgan,[1] Cathryn Crowle,[2] Traci-Anne Goyen,[3] Caroline Hardman,[4] Michelle Jackman,[5] Iona Novak[1] and Nadia Badawi[1,2]

[1]Cerebral Palsy Alliance Research Institute, University of Notre Dame, [2]Grace Centre for Newborn Care, The Children's Hospital Westmead, [3]Neonatal Intensive Care Unit, Westmead Hospital, [4]Newborn Care, Royal Prince Alfred Hospital and [5]Occupational Therapy Department, John Hunter Children's Hospital, Sydney, New South Wales, Australia

Aim: The aim of this study was to calculate the sensitivity and specificity of the General Movements Assessment (GMA) for estimating diagnostic accuracy in detecting cerebral palsy (CP) in an Australian context by a newly established NSW rater network.

Methods: A prospective longitudinal cross-sectional study was conducted. The GMA was blind-rated from conventional video by two independent certified raters, blinded to medical history. A third rater resolved disagreements. High-risk population screening for CP using the GMA during the fidgety period (12–20 weeks) was carried out in four neonatal intensive care units and one CP service over a 30-month period (2012–2013). Participants were 259 high-risk infants. Sensitivity and specificity values were calculated with true positives defined as a confirmed diagnosis of CP from a medical doctor.

Results: Of the 259 infants assessed, 1-year follow-up data were available for 187. Of these, $n = 48$ had absent fidgety (high risk for CP), $n = 138$ had normal fidgety (low risk for CP), and $n = 1$ had abnormal fidgety (high risk for a neurological disorder). Of the 48 with absent fidgety movements, 39 had received a diagnosis of CP by 18 months and another 6 had an abnormal outcome. Of the $n = 138$ normal fidgety cases, $n = 99$ cases had a normal outcome, $n = 38$ had an abnormal outcome but not CP, and $n = 1$ had CP. For detecting CP, we had a sensitivity of 98% and specificity of 94%.

Conclusion: GMA was feasible in an Australian context and accurately identified CP with a sensitivity and specificity comparable with European standards and published neuroimaging data.

Key words: cerebral palsy; infant; General Movements Assessment.

What is known on this topic
- CP is most often diagnosed around 18 months of age.
- The General Movements Assessment at 3–4 months of age has the highest sensitivity for detecting CP.

What this paper adds
- A new rater network in Australia achieved comparable sensitivity and specificity results for early detection of CP as those found in Europe.
- Recommendations for use in Australian clinical practice are suggested.

Introduction

Cerebral palsy (CP) is defined as a group of disorders of movement and posture that results from a lesion to the developing brain[1] and is the most common physical disability of childhood. Early detection of CP is important as it allows referral to early interventions aimed at maximising motor and cognitive outcomes in children and providing support to families.[2,3]

Data from the Australian Cerebral Palsy Register shows that the average age for diagnosis of CP is 17 months although the range varies from a few weeks to 4 years of age (I. Novak, unpubl data, 2014). About half of all children diagnosed with CP have identifiable markers that enable them to be labelled 'at risk' during the neonatal period, for example prematurity or neonatal encephalopathy.[2] These infants are typically cared for in neonatal intensive care units (NICUs) and are often enrolled in follow-up programmes to ascertain their long-term outcome. These programmes follow protocols to monitor infants for evidence of developmental delay or disabilities, referring for early intervention once signs become apparent.

Correspondence: Ms Catherine Morgan, Cerebral Palsy Alliance Research Institute, PO Box 6427, Frenchs Forest, NSW 2086, Australia. Fax: +61 2 9479 7291; email: cmorgan@cerebralpalsy.org.au

Recent systematic reviews have demonstrated that in fact, CP can reliably be detected as early as 3 months post term age using Prechtl's Qualitative Assessment of General Movements Assessment (GMA) and medical resonance imaging (MRI).[4–6] The GMA was developed by Professor Heinz Prechtl in the early 1990s and is an assessment of the spontaneous movement patterns ['general movements' (GMs)] of young infants.[7,8] Two periods of GMs are described; the 'writhing period' from preterm until 6–9 weeks post term age, and 'fidgety period' from 9 to 20 weeks post term age.[9] Normal GMs are shown to have a high correlation with a normal outcome, while abnormal GMs, in particular absent fidgety GMs (F−), are highly predictive of CP (sensitivity as high as 98% and specificity 91%,[6]). Thus the GMA is considered the reference standard for early detection of CP. Validity of the tool is established[9] and inter-rater reliability of the GMA has been repeatedly demonstrated.[9–11] Importantly, a number of studies have demonstrated that the predictive validity of the GMA is superior to neuroimaging,[6] while the combination of abnormal GMA and white matter injury evident on MRI has been shown to be 100% predictive of an outcome of CP in a cohort of preterm infants.[12] Studies in infants with hypoxic ischaemic encephalopathy (HIE) showed a high correlation between abnormal GMs and lesions of the basal ganglia and thalamus.[13]

Despite the compelling psychometric data, implementation of GMA in clinical practice outside of Europe has been ad hoc and is a 'know-do' evidence to practice gap. Systematic reviews on the predictive validity of the GMA have proposed that the lack of non-European data, especially outside the expert group (General Movements Trust), is a potential limitation to the generalizability of findings and possible explanation for the know-do gap.[5] Use of the GMA has been growing in Australia in the last 7–8 years. Spittle and colleagues from Melbourne Australia have demonstrated sensitivity and specificity results similar to European rates in very preterm children. In addition, their work has demonstrated important associations between neuroimaging findings and the GMA in predicting later neurodevelopmental outcomes.[12,14,15] These important studies have focused on preterm infants, a population that make up about 30% of all CP.[2] To date, little published data exist on the diagnostic accuracy of the GMA for a more heterogeneous clinical population of high-risk infants in an Australian context. In 2011, a knowledge translation programme to close the GMA know-do gap was implemented in New South Wales (NSW) Australia. First, European trainers were brought to Australia to remove the barrier of needing overseas rater training. Second, educational scholarships were provided to remove the costs of obtaining rater training. Third, a new rater network was established in NSW for the purpose of providing peer-to-peer support for maintaining GMA scoring reliability and troubleshooting any difficulties embedding the GMA in clinical practice. Network meetings are held twice a year and trained assessors from all participating centres present cases for blind scoring to help maintain inter-rater reliability. Between network meetings, de-identified videos are shared for blind scoring purposes to arbitrate any discrepancies.

The aim of this study was to calculate the sensitivity and specificity of the GMA for diagnostic accuracy of detecting CP at 3–5 months of age in high-risk infants, in an Australian context when scored by the NSW rater network.

Methods

Participants

Inclusion criteria

(i) All infants included were those prospectively enrolled in follow – up clinics and screened using the GMA from the study sites: four NICUs in NSW Australia (Westmead Hospital, the Children's Hospital at Westmead, John Hunter Children's Hospital and Royal Prince Alfred Hospital) and the Cerebral Palsy Alliance (CPA); (ii) All infants were designated high-risk of poor neurodevelopmental outcome based on their medical history and /or neuroimaging by at least one member of their treating team. This included infants admitted to NICUs post surgery or with neurological risk factors (e.g. severe intraventricular haemorrhage, periventricular leukomalacia, neonatal stroke), HIE (stages II–III), or due to prematurity (i.e. <29 weeks, one unit enrolled <32 weeks); or infants referred to CPA with motor delay or neurological signs suggestive of CP. Recruitment via voluntary participation was offered to all infants meeting the inclusion criteria, unless there was a competing concurrent study in which case they were offered enrolment to both studies, with researchers respecting the parent's choices.

Exclusion criteria

Nil.

Methodology

High-risk population screening for CP was conducted at study sites, predominantly in the NICU follow-up clinic over a 30-month period, resulting in a prospective longitudinal and cross-sectional study. The CPA received referrals from concerned parents and professionals in the community to screen infants for signs of CP.

Instrument: GMA

Infants were assessed during the fidgety movement period at the developmental follow-up clinic or in the family home. Since GMs in the fidgety period are the most predictive for a later diagnosis of CP, our outcome of interest, we focused on results from this GMA period. GMAs for 259 infants were collected on conventional video following the protocol outlined by Einspieler *et al.*[9]

All study sites used certified GM assessors to score the videos blinded to medical and clinical history. Although all sites had certified blind raters there was a number of minor pragmatic practice variations across the study sites in relation to the processes for arranging the scoring. Despite uniformity being preferable, in the clinical setting local variations was deemed allowable as the greater knowledge translation goal was for as many raters as possible to be using the GMA and all study sites to develop feasible and acceptable local processes that led to routine GMA use. For instance, one service had a number of raters who scored independently and were blinded, another had

two raters but only one blinded, and the other services had two blinded raters. A third rater, unaware of medical and clinical history and part of the GM Network, resolved disagreements for any case at any site. There were no scoring accuracy differences between the study sites, despite the differing processes.

Neurodevelopmental outcome

Infants were followed to 12–24 months post term age. True positives were defined as a confirmed diagnosis of CP from a medical doctor. The diagnosis was typically given at a follow-up time point by a developmental paediatrician or neonatologist based on neurological examination, clinical history and developmental motor assessment. For infants not diagnosed with CP, an abnormal outcome was defined as having scores on one or more domains of the Bayley Scales of Infant and Toddler Development - third edition (BSID-III) greater than 1 standard deviation (SD) below the mean at follow-up. If the only delay on the BSID-III was in the domain of language, the outcome was not coded as abnormal, given these children can go on to have a normal outcome despite delayed speech and language.[16]

Ethics

Ethics approval was obtained from all study sites, with The Royal Prince Alfred Hospital Human Research Ethics Committee as the lead committee and site specific approval from all other participating institutions. Parental consent was previously obtained from families at the point of GMA. The accompanying history data was abstracted from medical records.

Statistical analysis

This was a prospective study; with data analysis planned a priori to data collection. Study design and data analysis were reported in accordance with the STARD checklist for reporting of studies of diagnostic accuracy. Neurodevelopmental outcome data were compared with GMA results from the fidgety period. Statistical analysis was completed using the Statistical Package for the Social Sciences (SPSS; SPSS, Inc., Chicago, IL, USA) using conventional sensitivity and specificity calculation methods. Confidence intervals were calculated for sensitivity and specificity for predicting an outcome of CP and for any abnormal outcome.

Results

Participants

Data were collected on all infants recruited and screened between 2011 and 2013, although some study sites did not collect data for the full study period while awaiting Ethics Clearance. Infants were all between 10 and 20 weeks post term age at the time of their GMA fidgety assessment and within 2 weeks of their first or second birthday at their 1- or 2-year aged follow-up. The most common reason for a GMA was prematurity followed by neonatal encephalopathy.

Complete 1-year follow-up data were available for 187 infants. Partial data were available for another 72 infants who had not yet reached 12–24 months or were lost to follow up ($n = 62$) and for $n = 10$ whose GMA was not able to be scored. Reasons for conducting GMA are presented in Fig. 1.

Quality of GMs

Data were analysed when both the fidgety GMA and 12-month outcomes were available, all other cases were treated as missing and excluded from the analysis. Of the 187 complete cases, $n = 138$ were scored as normal fidgety (F+) that is low risk for CP, $n = 48$ were scored as absent fidgety (F−) that is high-risk of CP, and $n = 1$ were scored abnormal fidgety (AF) that is high-risk for a neurological disorder. No adverse events were reported as a result of testing.

Neurological outcome

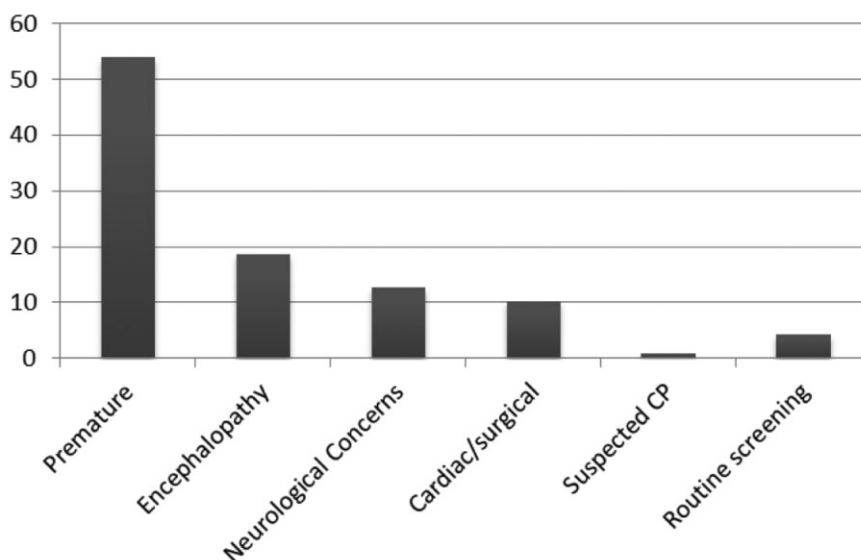At 1-year follow-up, of the $n = 187$ cases: 102 children had a normal outcome and 40 children had a diagnosis of CP. A

Fig. 1 Reasons for conducting General Movements Assessment (%).

Table 1   GMA fidgety results and 12 month outcome results

GMA result		12 month outcome		
Type of fidgety		Normal	CP	Abnormal
Normal (F+)	*n* = 138 (74%)	*n* = 99 (72%)	*n* = 1 (<1%)	*n* = 38 (28%)
Abnormal (AF)	*n* = 1 (<1%)	*n* = 0 (0%)	*n* = 0 (0%)	*n* = 1 (100%)
Absent (F−)	*n* = 48 (26%)	*n* = 3 (6%)	*n* = 39 (81%)	*n* = 6 (13%)

NB, shading indicates the predicted outcome from General Movements Assessment (GMA).

further 45 children had an abnormal outcome (not CP) (Table 1).

First, in the *n* = 138 with F + movements there were *n* = 99 with a normal outcome, *n* = 1 later diagnosed with CP, and *n* = 38 with a neurodevelopmental delay other than CP. Abnormal neurodevelopmental outcomes that were not CP included; *n* = 1 with Prader–Willi syndrome, *n* = 2 with hearing impairments and *n* = 35 with global developmental delay, including *n* = 1 suspected autism. The infants with global developmental delay ranged from mild motor and/or cognitive delay at 12 months to significant delays in both the cognitive and motor domains, as scored on the BSID-III.

Second, in the *n* = 48 with F− movements, *n* = 39 infants were diagnosed with CP by 12–18 months. Of the *n* = 9 with F− and not diagnosed with CP all had a suspected or confirmed diagnosis of an abnormal outcome, including: *n* = 1 had a genetic disorder; *n* = 1 had a mitochondrial disorder; *n* = 1 was recovering from meningitis; *n* = 3 had moderate-severe global developmental delay; and *n* = 3 had suspected CP at 12 months, but had not yet been formally diagnosed but were undergoing monitoring for a diagnosis of CP (coded as normal outcomes however at 12 months).

Third, the *n* = 1 with AF movements had a motor delay at 12 months > 1 SD below the mean.

Sensitivity and specificity

Sensitivity and specificity scores were calculated for predicting CP and for predicting an abnormal outcome. Sensitivity for detecting CP was 98% [95% confidence interval (CI): 86.79–99.58] and specificity 94% (95% CI: 88.69–97.16). Sensitivity for detecting any abnormal outcome with abnormal or absent fidgety GMs was 54% (95% CI: 42.66–64.98) and specificity 97% (95% CI: 91.63–99.36).

The mean age of CP diagnosis for children identified at high risk of CP by the NSW GMA rater network was 8.5 months (SD = 4 months). All infants identified as high risk of CP by F− GMs at 3–4 months were referred to early intervention services. The child later diagnosed with CP but with normal fidgety movements was also referred for early intervention due to concerns about motor development that were identified at follow-up from tests other than the GMA.

Discussion

The GMA has consistently been shown to be a sensitive method for early detection of adverse neurodevelopmental outcomes especially CP. Although clinical use has generally been lacking outside the European context, this study confirmed that the GMA had excellent sensitivity and specificity to predict infants who would later be diagnosed with CP as well as those with normal outcomes. Our results are comparable with previous Australian and European studies demonstrating that the reliability of the GMA can be replicated in different parts of the world.

In the clinical setting, making a diagnosis of CP utilises a combination of robust, evidence-based tools including neuroimaging, neurological and standardised motor testing.[2] The GMA is a highly predictive, non-invasive assessment that would be a valuable tool to add to the diagnostic work-up. Results of this study suggest that one benefit of early detection using GMA was that diagnosis occurred earlier, on average at 8.5 months compared with the Australian CP Register convention of 17 months. Previous Australian studies of preterm infants have followed infants until 4 years and demonstrated the value of the GMA in predicting adverse neurodevelopmental outcomes.[15] The current study builds on this work with a broader group of high-risk infants, indicating very early identification of infants at the highest risk of motor impairment is possible and clinicians can be confident in referring those most in need of early intervention in the first few months of life. Clinical application of the GMA is useful to build a clinical profile of high-risk infants over time. It allows early entry of infants into targeted treatment programmes and enrolment into intervention studies during the period of greatest neuroplastic change.

All study infants designated 'high risk of CP' on the basis of F− GMs were referred for early intervention. Unless there were very clear markers such as severe MRI findings, a definitive diagnosis was not given at this time, because of diagnostician-preferred practices.[2] Importantly, our high sensitivity rates confirm that parents were not 'worried unnecessarily', given that almost all the infants with F− movements were found to have an abnormal outcome.

The high rate of abnormal outcomes found in this study is consistent with previous studies reporting outcomes in this high-risk population.[17–19] We defined 'abnormal' as a delay in at least one developmental domain of the BSID-III, which is the commonly used criteria in some follow-up services, although some services prefer to define an abnormal outcome as one where at least two domains of the BSID are > 1 SD below the mean. In our analysis, only delays in language alone were not counted as abnormal because of the high level of variability in

the emergence of these skills and high prevalence of early language delays that resolve.[20] Not surprisingly, the GMA did not detect infants with developmental delay, highlighting the importance of using complementary assessments when following high-risk infants. Detecting a probable CP outcome versus one of mild developmental delay is important as it allows referral for diagnosis-specific intervention.[21]

The GMA has now been embedded in clinical practice across NICU follow-up services in NSW, Australia. Use of the GM Rater Network has provided support for use of this tool and for the accuracy of results. To accommodate timing of peak fidgety period, a number of services have brought forward their initial follow-up clinic visit to 3 months of age rather than the conventional 4 months of age, in order to capture the GMs of at-risk infants during the ideal fidgety period.

It is recommended that the following high-risk groups of infants be screened using the GMA; preterm (including late preterm), all with neonatal encephalopathy, cardiac and surgical infants, those with stroke and neurological signs such as seizures, growth restriction and those with birth defects.[22]

Limitations

There are several limitations related to our study. First, as has been noted in previous publications, sampling is a potential source of bias.[23] All infants in this study were already considered at high risk of adverse neurodevelopmental outcome. Within our group the level of risk for CP specifically was variable. For example, the sample included term infants with HIE (very high risk for CP) and those with congenital heart defects, very preterm and late preterm infants. In addition, cases were recruited for the most part sequentially; however, some cases were excluded as they were recruited to other studies, and some study sites did not collect data for the full study period owing to the differing timelines for study approval from Ethics.

Second, outcome data were mostly only at 12 months and it is known that milder forms of CP may only be diagnosed later in childhood when the diagnostician is sure that the motor impairment is permanent. Indeed $n = 3$ infants were suspected to have a mild CP because of tone abnormalities, but had not yet been formally diagnosed, but were being closely monitored by allied health practitioners who suspected they had CP. Potentially the rate of CP therefore has been under identified in this sample, and that the sensitivity of the GMA might have been even higher. Future studies should report 2-year outcomes in this high-risk cohort, as has been done previously in very preterm groups.[15] Third, additional analysis of sensitivity and specificity of GMs in the earlier writhing period might lead to the development of effective very early interventions that could be applied in the NICU within first 2–3 months of life closest to the timing of the brain injury. Fourth, sex of participants was deliberately not recorded so as to protect the anonymity of children with an absent fidgety score from small study sites, where the n-value was below the conventional $n = 4$ cut-off for anonymity. The accuracy of GMs is not however known to be affected by gender and therefore this is unlikely to have influenced the results. Finally, as previously outlined, the practice variation between sites in terms of number of blinded GMs scorers is a further limitation of the study.

Conclusion

The GMA is an accurate, important and feasible assessment tool. It is non-invasive and therefore should be used regularly in the NICU environment and in follow-up programmes for early identification of infants at the highest risk of CP. It is clinically feasible to use and has excellent predictive validity when used by certified Australian assessors. Early detection of CP is possible and implementation of screening high-risk infants will allow those identified timely access to intervention services that aim to optimise their developmental outcomes. In conclusion, we recommend that the GMA be widely adopted into clinical practice, to close the know-do gap about late diagnosis of CP, which is potentially harmful to infants.

Acknowledgements

References

1 Rosenbaum P, Paneth N, Leviton A *et al.* The definition and classification of cerebral palsy. *Dev. Med. Child Neurol.* 2007; **19**: 1–44.

2 McIntyre S, Morgan C, Walker K, Novak I. Cerebral palsy – don't delay. *Dev Disabil. Res. Rev.* 2011; **17**: 114–29.

3 Herskind A, Greisen G, Nielsen JB. Early identification and intervention in cerebral palsy. *Dev. Med. Child Neurol.* 2015; **57**: 29–36.

4 Burger M, Louw QA. The predictive validity of general movements – a systematic review. *Eur. J. Paediatr. Neurol.* 2009; **13**: 408–20.

5 Darsaklis V, Snider L, 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.

6 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.

7 Prechtl H, Ferrari F, Cioni G. Predictive value of general movements in asphyxiated fullterm infants. *Early Hum. Dev.* 1993; **35**: 91–120.

8 Ferrari F, Cioni G, Einspieler C *et al.* Cramped synchronized general movements in preterm infants as an early marker for cerebral palsy. *Arch. Pediatr. Adolesc. Med.* 2002; **156**: 460–7.

9 Einspieler C, Prechtl H, Bos AF, Ferrari F, Cioni G, eds. *Prechtl's Method on the Qualitative Assessment of General Movements in Preterm, Term and Young Infants*. London: Mac Keith Press, 2004.

10 Spittle A, Doyle L, Boyd R. 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.

11 Valentin T, Uhl K, Einspieler C. The effectiveness of training in Prechtl's method on the qualitative assessment of general movements. *Early Hum. Dev.* 2005; **81**: 623–7.

12 Spittle A, Brown N, Doyle L *et al.* Quality of general movements is related to white matter pathology in very preterm infants. *Pediatrics* 2008; **121**: e1184–9.

13 Ferrari F, Todeschini A, Guidotti I *et al.* General movements in full-term infants with perinatal asphyxia are related to basal ganglia and thalamic lesions. *J. Pediatr.* 2011; **158**: 904–11.

14 Spittle A, Boyd R, Inder T, Doyle L. 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**: 512–17.

15 Spittle A, Spencer-Smith J, Cheong J *et al*. General movements in very preterm children and neurodevelopment at 2 and 4 years. *Pediatrics* 2013; **132**: e452–8.

16 Tomblin JB, Records NL, Buckwalter P *et al*. Prevalence of specific language impairment in kindergarten children. *J. Speech Lang. Hear. Res.* 1997; **40**: 1245–60.

17 Hoekstra RE, Ferrara TB, Couser RJ *et al*. Survival and long-term neurodevelopmental outcome of extremely premature infants born at 23–26 weeks' gestational age at a tertiary center. *Pediatrics* 2004; **113**: e1–6.

18 Walker K, Badawi N, Halliday R *et al*. Early developmental outcomes following major non-cardiac and cardiac surgery in term infants: a population-based study. *J. Pediatr.* 2012; **161**: 748–52.

19 Walker K, Holland AJ, Halliday R, Badawi N. Which high-risk infants should we follow-up and how should we do it? *J. Paediatr. Child Health* 2012; **48**: 789–93.

20 Dale PS, Price TS, Bishop DV, Plomin R. Outcomes of early language delay: predicting persistent and transient language difficulties at 3 and 4 years. *J. Speech Lang. Hear. Res.* 2003; **46**: 544–60.

21 Novak I. Evidence-based diagnosis, health care, and rehabilitation for children with cerebral palsy. *J. Child Neurol.* 2014; **29**: 1141–56.

22 McIntyre S, Taitz D, Keogh J *et al*. A systematic review of risk factors for cerebral palsy in children born at term in developed countries. *Dev. Med. Child Neurol.* 2013; **55**: 499–508.

23 Hanna SE. High and variable prevalence does matter in reviews of diagnostic accuracy in cerebral palsy. *Dev. Med. Child Neurol.* 2013; **55**: 397–8.