Neurodevelopmental outcome in very preterm and very-low-birthweight infants born over the past decade: a meta-analytic review

AURELIE PASCAL[1,2] | PAUL GOVAERT[1] | ANN OOSTRA[3] | GUNNAR NAULAERS[2] | ELS ORTIBUS[2] | CHRISTINE VAN DEN BROECK[1]

1 Department of Rehabilitation Sciences and Physiotherapy, Ghent University, Ghent; **2** Department of Development and Regeneration, Katholieke Universiteit Leuven, Leuven; **3** Center for Developmental Disorders, University Hospital Ghent, Ghent, Belgium.

Correspondence to Aurelie Pascal, Department of Rehabilitation Sciences and Physiotherapy, Ghent University, Campus Heymans (UZ), 1B3, De Pintelaan 185, B-9000 Ghent, Belgium. E-mail: Aurelie.pascal@ugent.be

This article is commented on by Marlow on page 330 of this issue.

PUBLICATION DATA
Accepted for publication 2nd December 2017.
Published online 19th January 2018.

ABBREVIATIONS

Bayley-III	Bayley Scales of Infant and Toddler Development, Third Edition
BSID-II	Bayley Scales of Infant Development, Second Edition
ELBW	Extremely low birthweight
EPT	Extremely preterm
M-ABC	Movement Assessment Battery for Children
VLBW	Very-low-birthweight
VPT	Very preterm

AIM The purpose of this systematic review was to provide an up-to-date global overview of the separate prevalences of motor and cognitive delays and cerebral palsy (CP) in very preterm (VPT) and very-low-birthweight (VLBW) infants.
METHOD A comprehensive search was conducted across four databases. Cohort studies reporting the prevalence of CP and motor or cognitive outcome from 18 months corrected age until 6 years of VPT or VLBW infants born after 2006 were included. Pooled prevalences were calculated with random-effects models.
RESULTS Thirty studies were retained, which included a total of 10 293 infants. The pooled prevalence of cognitive and motor delays, evaluated with developmental tests, was estimated at 16.9% (95% confidence interval [CI] 10.4–26.3) and 20.6% (95% CI 13.9–29.4%) respectively. Mild delays were more frequent than moderate-to-severe delays. Pooled prevalence of CP was estimated to be 6.8% (95% CI 5.5–8.4). Decreasing gestational age and birthweight resulted in higher prevalences. Lower pooled prevalences were found with the Third Edition of the Bayley Scales of Infant Development than with the Second Edition.
INTERPRETATION Even though neonatal intensive care has improved over recent decades, there is still a wide range of neurodevelopmental disabilities resulting from VPT and VLBW births. However, pooled prevalences of CP have diminished over the years.

It is estimated that preterm birth occurs in 11.1% of all worldwide deliveries, of which 10% are very preterm (VPT) infants (28–31wks gestational age) and 5% extremely preterm (EPT) (<28wks gestational age). This represents almost 15 million infants annually and the number keeps rising.[1,2] Such trends could be explained by enhanced reproductive technology, which is commonly associated with multiple gestations, increased age of the mother, and changes in clinical practice as an increase in Caesarean sections before term age.[3] With the more prevalent use of antenatal steroids, surfactants, advanced ventilator techniques, and a drastic reduction in postnatal steroid use over the past two decades,[4] not only have survival rates of VPT, especially EPT, infants increased, but neonatal morbidity has also decreased.[5,6] Furthermore, the frequency and severity of adverse outcomes seem to be related to a decreased gestational age, birthweight, and structural brain changes.[7–9] At present, a considerable number of infants born before 25 weeks gestational age do survive. Nevertheless, fewer than half of those infants survive without neurodevelopmental impairment around 2 years corrected age (20% for infants born at 22–24wks' gestation,[10] and 34–48.5% for infants born at 22–26wks' gestation).[11–13] Proportionally, the prevalence of EPT is low; however, on the basis of their high rate of mortality and morbidity, this may affect the overall impairment rates in the wider VPT population group.

A wide range of neurodevelopmental outcomes of infants born EPT and VPT have been described in the literature; however, just a few articles have provided unified data through a global meta-analysis.[7,14–16] Neurodevelopmental outcomes, often defined as a combination of cognitive delays, motor delays, cerebral palsy (CP), blindness, and/or hearing impairment, have been the historical results of interest as they are the most commonly reported disabilities of infants born preterm. A recent meta-analysis by Blencowe et al.[14] was based on articles with a median birth year of 2000 or later and estimated that worldwide 52% and 24% of EPT and VPT infants respectively develop a certain degree of neurodevelopmental impairment. Yet this provides no detailed information on specific outcomes. In the past decade, two meta-analyses provided data on

 DOI: 10.1111/dmcn.13675

separate outcomes. One was performed by Mwaniki et al.,[16] which included articles from 1966 until 2011. They reported a median prevalence for CP and motor, cognitive, and overall neurodevelopmental impairment in 11.6%, 18.9%, 20.7%, and 27.9% of infants born preterm respectively (<37wks gestational age). The other meta-analysis was performed by Oskoui et al.,[15] which featured articles published from 1985 until 2011; they revealed that the pooled prevalence of CP was 14.5% and 11.2% in infants born EPT and VPT respectively.

On the basis of continuous advances in obstetric and neonatal care, which has affected the morbidities and neurodevelopmental outcomes of those VPT or very-low-birthweight (VLBW) infants, it is important to collect and unify recent data. Accurate prognostic information is valuable for clinicians and families who are exposed to VPT infants or those with a VLBW, as well as for benchmarking hospitals. The purpose of this systematic review is to provide an up-to-date overview of the separate prevalences of motor and cognitive delay and CP in relation to gestational age and birthweight.

METHOD

Search strategy

This systematic review was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.[17] A systematic literature search was conducted with the Embase, MEDLINE, Web of Science, and CINAHL databases in August 2016. The search strategy comprised free keywords combined with Medical Subject Headings (MeSH) terms or Emtree terms, as detailed in Appendix S1 (online supporting information). Searches were restricted to English, French, or Dutch publications (i.e. languages understood by the review authors) and strictly human studies. Only consecutive cohort studies (prospective and retrospective) investigating and reporting the prevalence of CP and motor or cognitive outcomes from 18 months, or if started earlier going up to at least 20 months, until the age of 6 years of VPT or VLBW infants were included. The participants had to be born within the past decade (2006 or after, or at least two-thirds of the total cohort born after 2006) and before 32 weeks gestational age (or mean gestational age <30.5wks), and/or have a VLBW (<1500g). Follow-ups had to be performed in at least 50 eligible infants by professionals. Outcomes based exclusively on questionnaires for parents or parental interviews were excluded. We also decided to dismiss studies with only outcomes for working memory, language, behaviour, or executive functioning. If different papers were based on the same cohort, only the article representing the largest population or reporting most data was retained. The titles and abstracts of the studies were screened by two authors (AP and CVdB) to identify all potentially eligible studies. Full texts of the remaining articles were read and assessed thoroughly to exclude articles that did not meet our inclusion criteria. Any discrepancy in the suitability for inclusion of a study was resolved by discussion among the authors. A flowchart, summarizing the article selection process and the reasons for exclusion, is presented in Figure 1.

What this paper adds

- The Bayley Scales of Infant and Toddler Development, Third Edition reported lower pooled prevalences of motor and cognitive delays than the Second Edition.
- The pooled prevalence of cerebral palsy in infants born extremely preterm was reduced compared with previous meta-analyses.

Quality assessment

Each study was evaluated by two independent authors (AP and CVdB) for methodological quality. As only cohort studies were included, the Quality Assessment Tool for Observational Cohort and Cross-Sectional Studies was used for all studies. A total score was generated by summation of all criteria that were fulfilled and this score was transformed into a percentage. The limit to be included in the meta-analytic review was set at 50%. The results are found in Appendix S2 (online supporting information).

Data extraction and processing

Study characteristics (see Appendix S3, online supporting information) and outcome measurements of each included article were collected using our data extraction form, which included (1) first author, year of publication, country, neonatal death rate, and prevalences of active neonatal care; (2) participant characteristics (birth year, inclusion criteria, mean and range of gestational age and birthweight, exclusion criteria and sample size); (3) outcomes (number of patients at follow-up, mean age at follow-up, outcome measurements, and cut-off values). Outcome measurements were divided into developmental scales, as well as motor and cognitive tests.

Mild delays were considered to be scores between one and two units of standard deviation (SD) and moderate-to-severe delays had a score of two SD below standard norms or the comparison group. If other cut-off values were used, the described criteria were adopted.

Statistical analyses

Prevalence calculations were consistently based on the number of infants with a certain degree of mild or moderate-to-severe delays divided by the total number of infants assessed during the same follow-up period with the same outcome measures. The confidence intervals of the prevalences were calculated by using a logit transformation (with back-transformation). The overall pooled prevalences, with their 95% confidence limits, were estimated with a random-effects model that accounted for between-study heterogeneity. Using a random-effects model allows a higher generalization of the results than a fixed-effects model.[18] Heterogeneity between studies was evaluated with a χ^2 test (Cochran Q statistic) and quantified with the I^2 statistic, which represents the percentage of between-study variation that emanates from heterogeneity rather than from chance. A value of 0% indicated no observed heterogeneity, whereas I^2

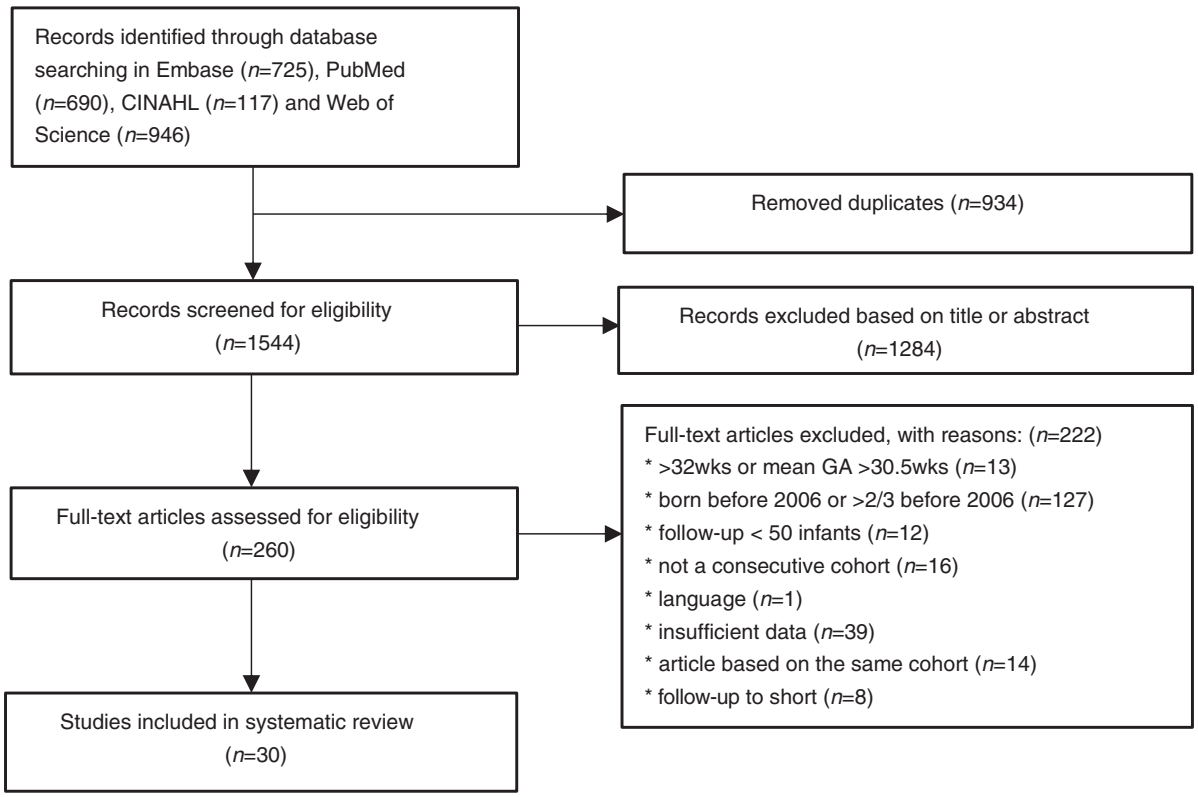

Figure 1: Flow chart outlining literature selection process. GA, gestational age.

values greater than or equal to 50% suggested a substantial level of heterogeneity, and a value greater than 75% was interpreted as high heterogeneity.[19] It is known that this test has low power for the purposes of detecting heterogeneity and, therefore, it is advised to use a *p* value of 0.10 as a cut-off for significance.[19]

The potential sources of heterogeneity were investigated by stratification of the studies according to potentially relevant characteristics. Subgroup analyses were performed on the basis of the mean gestational age, mean birthweight, age at follow-up, follow-up ratio, sample size, outcome measures and cut-off values, country income level, and geographical region. The significance threshold was set at 0.05 for variability in terms of prevalences. The prevalence of CP in relation to gestational age and birthweight was evaluated by meta-regression using weighted-linear regression. All statistical analyses were conducted with the Comprehensive Meta-Analysis program (Biostat, Inc., Englewood, NJ, USA), version 3.3.070.

RESULTS

Study selection process

A total of 2478 publications were initially identified (Fig. 1). After removing duplicates, 1284 citations were excluded on the basis of the screening of titles and abstracts, and 222 citations after detailed assessment of the full text. In total, 44 articles met our inclusion criteria. Cohort information was carefully verified and 14 studies were excluded as their results were based on the same cohort. Finally, 30 studies were retained for the work featured herein.[20–49]

Study characteristics and population

All included articles had a level of evidence B. The quality of the articles varied between 58.3% and 92.3%. No articles were dismissed as a result of the quality assessment.

The characteristics of the included articles are listed in Appendix S2. There were 20 prospective,[20,21,23–26,29–31,33,34,37,40–42,44–48] and 10 retrospective cohort studies.[22,27,28,32,35,36,38,39,43,49] Altogether, 10 293 infants were included for the follow-up, representing different continents. Eleven studies were conducted in Europe,[20,24,25,29,33,37,40,45–48] nine in North America,[23,30,32,34,35,39,41–43] five in Asia,[27,28,31,44,49] two each in Africa[21,36] and Oceania,[22,38] and one in South America.[26]

Six articles featured exclusively infants born EPT.[23,24,32–34,38] Twelve articles used VLBW as an inclusion criterion.[21,25–28,31,37–39,42,44,49] Eleven articles reported a study sample with a mean gestational age of less than 28 weeks,[23,24,28,30,32–35,38,41,43] and 17 articles between 28 weeks and 32 weeks.[20–22,25,26,28,29,31,37,39,42,44–49] Three articles reported no mean gestational age.[27,36,40] The mean birthweight was lower than 1000g in 11 articles,[23,24,28,30–32,34,35,38,41,43] and between 1000g and 1501g in 15

studies.[20–22,25,26,29,37,39,42,44–49] Four articles did not report a mean birthweight.[27,33,36,40]

Neonatal mortality and active neonatal care

The reported prevalences of neonatal mortality and active neonatal care are given in Appendix S3. Of the 30 included articles, just 19 (63.3%) reported the number of infants who died before discharge[21–26,34–38,41,43,44,46,49] or exclusively during the neonatal period, the first 28 days of life.[20,47,48] Four articles featured a subdivision between the period of death (0–7d, 7–28d, and after 28d[22,34,49] or ≤12h and >12h–3d[23]). Administration of antenatal and/or postnatal corticosteroids was described in, 21 articles[20,21,23,24,26,29–33,35–37,39–41,43–45,47,49] and eight articles[20,29,33,37–39,44,46] respectively and varied between 41% and 95% versus 5% and 29%. Only two articles noted specific limitations for active reanimation, which was set at a minimum of 900g in the study by Ballot et al.,[21] and a minimum of 26 weeks gestational age in the work of Besnard et al.[22]

Outcome measurements

The length of follow-up varied between 18 months and 5 years 6 months, except for one study where the authors started follow-up at 8 months up to 22 months.[21] Only five articles reported longer-term outcomes, in particular between 3 years and 6 years.[22,24,27,29,33] The most commonly used outcome measure was the Bayley Scales of Infant Development (BSID). The Bayley Scales of Infant Development, Second Edition (BSID-II) and Bayley Scales of Infant and Toddler Development, Third Edition (Bayley-III) were used in four studies[26,28,33,49] and 18 studies[21,23,29–35,38,39,41–43,45–48] respectively. Other developmental tests used were the Griffiths Development Scales[25,29,37] and Brunet–Lézine test.[40]

The Movement Assessment Battery for Children (M-ABC),[27,29] as a motor outcome measure, was used in two studies. As cognitive assessment tools, the Wechsler Preschool and Primary Scale of Intelligence Test, Revised[27] and Third Editions, were used. The Amiel-Tison and Hempel neurological examinations were performed in six studies[20,24,32,37,39,44] and one study[45] respectively. The remaining investigations featured standard neurological examinations.

Cognitive outcomes

Prevalences

Cognitive outcomes were divided into developmental scales with a cognitive subscore and proper cognitive tests.

The cognitive subscore of the BSID was reported in 20 of the included studies. Five studies did not distinguish between any level of delay,[26,29,32,45,47] and six studies strictly described the prevalence of moderate-to-severe cognitive delay.[23,30,33,35,43,49] Age at follow-up varied between 8 months and 3 years, with most evaluating cognitive outcomes around 2 years corrected age.

Overall, few studies[29,35,37,48] described very low prevalences (<5%) of cognitive delay, whereas Rogers et al.[41] reported the highest prevalence, with nearly 70% of those infants demonstrating a cognitive delay.

Three studies reported the outcome of cognitive tests (Wechsler Preschool and Primary Scale of Intelligence Test) at a later age (27mo–5y 6mo; see Table I). Mild cognitive delay (<1SD), as reported by Keunen et al.,[29] was present in 25% of those infants, while moderate-to-severe delay (<2SD) was present in 11.9% to 16.3% of the infants.[27,33]

Meta-analysis

Figure 2 illustrates the individual and pooled prevalences. The random-effects pooled prevalence of overall cognitive delay among VPT/VLBW infants on the basis of the developmental scales was estimated at 16.9% (95% CI 10.4–26.3, I^2=94.22, p<0.001). The pooled prevalence of mild cognitive delay was higher than moderate-to-severe cognitive delay and reached an overall prevalence of 14.3% (95% CI 8.3–23.5, I^2=90.49, p<0.001) versus 8.2% (95% CI 5.5–12.0, I^2=92.20, p<0.001) respectively.

On the basis of cognitive tests at a later age, only the pooled prevalences of moderate-to-severe delay could be calculated. This was estimated at 14.7% (95% CI 10.9–19.5, I^2=46.99, p=0.170), on the basis of just two studies.[27,33]

Subgroup analysis

Table II summarizes the prevalence rate calculations and 95% CIs based on mean birthweight and gestational age for the cognitive score of the developmental scales. The prevalence of overall cognitive delay increased with a decreasing mean gestational age, although this was not found to be statistically significant (p=0.305). The estimated pooled prevalence of cognitive delay was higher in infants born EPT than infants born VPT, at 29.4% (95% CI 7.5–68.0, I^2=96.91, p<0.001) and 14.3% (95% CI 8.2–23.7%, I^2=93.75, p<0.001) respectively. ELBW infants had higher prevalences of cognitive delay than VLBW infants (22.4%, 95% CI 9.7–43.6, I^2=94.88 vs 14.3, 95% CI 7.3–25.4, I^2=94.32, p=0.368). Moderate-to-severe cognitive delay was also found to be higher in EPT and ELBW infants than VPT and VLBW infants. Other subgroup analyses are represented in Table III. Sample sizes and follow-up ratios were not significant moderators for prevalence variability, whereas geographical region, country income, and age at follow-up were observed to be significant.

Table IV offers a summary of the pooled prevalences by different outcome measures and the cut-off values used. The results indicate that studies making use of the BSID-II are associated with reports of higher, but not statistically significant (p=0.104), overall cognitive delay prevalence compared with the Bayley-III, when using the same standard cut-off values.

Motor outcomes

Prevalences

Motor outcomes were divided into developmental scales with a motor (sub)score and proper motor tests at a

Table I: Pooled prevalences of motor and cognitive delay, based on motor and cognitive tests

Article	Outcome measures	Cut-off	Age at follow-up	Event rate (%)	95% CI	Heterogeneity I^2 (%), (*p*)
Motor						
Howe et al.[27]	M-ABC	<5th centile	5y	33.8 (52/154)	26.7–41.6	
Keunen et al.[29]	M-ABC	<1SD	5y 6mo	40.0 (34/85)	30.2–50.7	
			Total	36.0 (86/239)	30.2–42.3	0.0 (p=0.337)
Cognitive						
Keunen et al.[29]	WPPSI-III	<1SD	5y 6mo	25.0 (10/40)	14.0–40.5	
Howe et al.[27]	WPPSI-R	<2SD	5y	11.9 (19/160)	7.7–17.9	
Moore et al.[33]	WPPSI-III	<2SD	27–48mo	16.3 (94/576)	13.5–19.6	
			Total <2SD	14.7 (113/736)	10.9–19.5	47.0 (p=0.170)

Random-effects analysis. CI, confidence interval; M-ABC, Movement Assessment Battery for Children; WPPSI (-III, -R), Wechsler Preschool and Primary Scale of Intelligence Test, Third Edition (Revised).

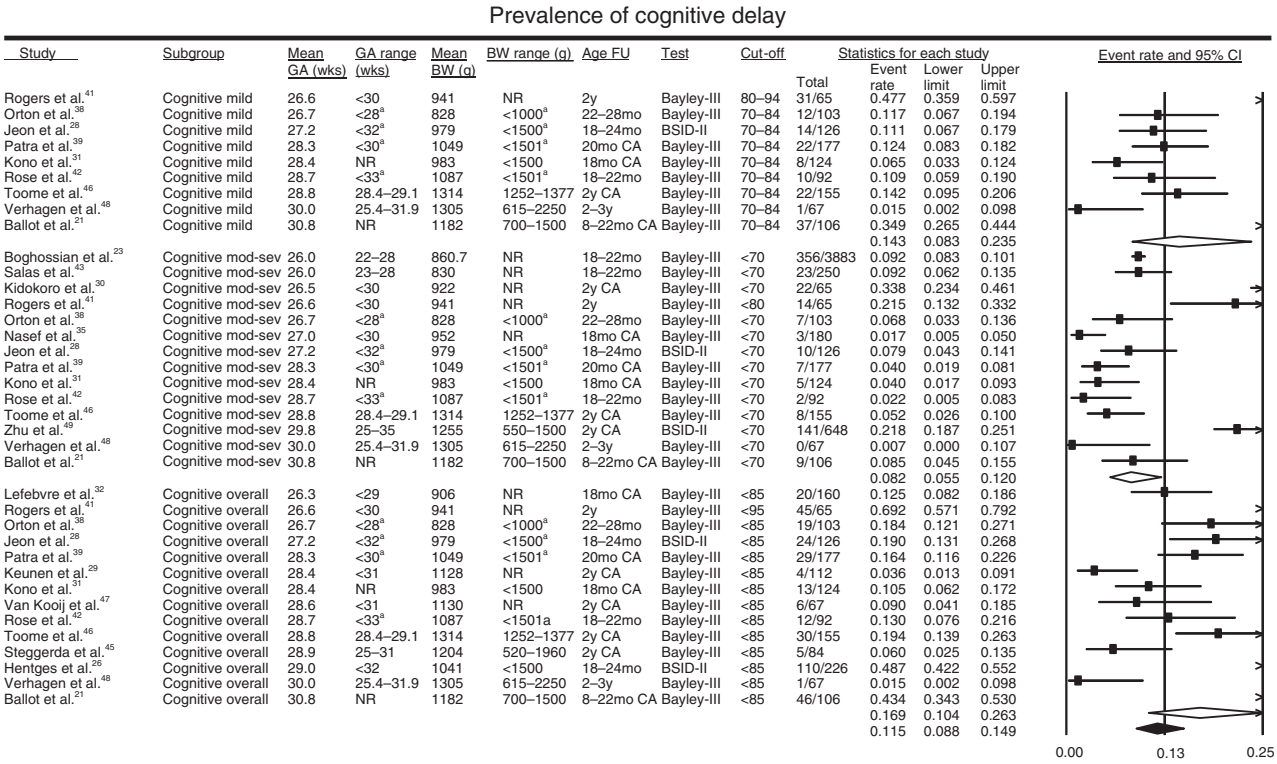

Figure 2: Prevalences of cognitive delays. Forest plot depicting the random-effects proportion meta-analysis for cognitive delays, on the basis of developmental scales (cognitive subscore of the Bayley Scales of Infant Development, Second and Third Editions). Studies are ordered on mean gestation. Black squares denote the reported prevalence of each study and the horizontal line represents the 95% confidence interval (CI). The pooled prevalence estimate is marked with a diamond. [a]Birthweight range and/or gestational age range. GA, gestational age; BW, birthweight; FU, follow-up; NR, not reported; CA, corrected age; Bayley-III, Bayley Scales of Infant and Toddler Development, Third Edition; BSID-II, Bayley Scales of Infant Development, Second Edition.

preschool age. Five studies[26,29,32,45,47] did not discriminate between any level of delay, and another four studies just described the prevalence of moderate-to-severe motor delay.[23,30,33,35]

Most of the articles reported motor outcomes at approximately 2 years corrected age. The prevalence of motor delays, based on developmental tests, varied considerably between the included studies. Moderate-to-severe delays were observed to be less than 5% in four of 17 studies,[31,35,42,48] and reached as high as 34% in the work of Rogers et al.,[41] where higher cut-off values for the motor scale of the Bayley-III were applied.

The prevalence of motor delays evaluated with the M-ABC increased until 33% or 40% at the age of 5 to 5 years 6 months, investigated by Howe et al.[27] and Keunen et al.[29] respectively.

Meta-analysis

Pooled prevalences and a corresponding forest plot are featured in Figure 3. An overall motor delay, based on

Table II: Pooled prevalences of cerebral palsy and cognitive and motor delays (based on developmental tests) by mean birthweight and gestational age

Subgroup	Categories	Number of studies	Event rate	Pooled prevalence (%)	95% CI (%)	Heterogeneity I^2 (%), (p)	p value for difference[a]
Overall cognitive delay							
Mean gestational age	<28wks (26–27wks)	3	84/328	29.4	7.5–68.0	96.91 (p<0.001)	0.305
	28–32wks	11	280/1336	14.3	8.2–23.7	93.75 (p<0.001)	
Mean birthweight	<1000g	5	121/578	22.4	9.7–43.6	94.88 (p<0.001)	0.368
	1000–1500g	9	243/1086	14.1	7.3–25.4	94.32 (p<0.001)	
Moderate-to-severe cognitive delay							
Mean gestational age	<28wks (26–27wks)	6	425/4546	10.9	6.1–18.6	91.30 (p<0.001)	0.184
	28–32wks	8	182/1495	5.8	2.7–12.0	91.10 (p<0.001)	
Mean birthweight	<1000g	8	440/4796	9.5	5.9–15.0	88.74 (p<0.001)	0.372
	1000–1500g	6	167/1245	5.6	2.1–14.1	91.84 (p<0.001)	
Overall motor delay							
Mean gestational age	<28wks (26–27wks)	3	120/327	44.5	14.2–79.5	96.70 (p<0.001)	0.093
	28–32wks	10	236/1181	16.4	11.1–23.7	85.92 (p<0.001)	
Mean birthweight	<1000g	5	177/577	34.4	18.5–54.6	94.12 (p<0.001)	0.021
	1000–1500g	8	179/931	13.3	7.6–22.2	88.79 (p<0.001)	
Moderate-to-severe motor delay							
Mean gestational age	<28wks	6	394/3340	11.2	7.0–17.4	89.82 (p<0.001)	0.227[a]
	<26wks	1	45/576	7.8	5.9–10.3	0.0	
	26–27wks	5	349/2764	12.0	6.5–21.1	89.15 (p<0.001)	
	28–32wks	6	41/692	6.3	4.3–9.3	34.87 (p=0.175)	
Mean birthweight	<1000g	7	368/3014	10.6	6.5–16.7	86.44 (p<0.001)	0.185
	1000–1500g	4	22/442	5.5	3.3–9.1	23.53 (p=0.270)	
	Not reported	1	45/576	7.8	5.9–10.3	0.0	
Cerebral palsy							
Mean gestational age	<28wks	9	603/5416	10.0	8.1–12.2	61.76 (p=0.007)	<0.001[a]
	<26wks	3	103/769	13.2	10.6–16.4	12.1 (p=0.320)	
	26–27wks	6	500/4647	8.6	6.4–11.6	64.8 (p=0.014)	
	28–32wks	15	117/2373	4.5	3.3–6.3	57.4 (p=0.003)	
	Not reported	1	8/60	13.3	6.8–24.5	0.0	<0.001
Mean birthweight	<1000g	10	534/5090	8.4	6.6–10.7	58.5 (p=0.100)	
	1000–1500g	13	103/2123	4.2	2.9–6.2	62.2 (p=0.002)	
	Not reported	2	91/636	14.3	11.8–17.3	0.0 (p=0.821)	

Random-effects analysis. [a]p value for the mean gestational age is based on the three categories (<26wks, 26–27wks, and 28–32wks gestational age); the category 'not reported' is not included for the calculations of the p value. CI, confidence interval.

developmental scales, was documented at 20.6% (95% CI 13.9–29.4, I^2=90.91, p<0.001) among all VPT or VLBW infants. Mild delays (18.0%, 95% CI 11.1–27.8, I^2=88.53, p<0.001) were more common than moderate-to-severe motor delays (8.6%, 95% CI 6.0–12.1, I^2=84.77, p<0.001; Fig. 3).

At preschool age, a pooled prevalence of 36.0% (95% CI 30.2–42.3) was estimated for motor delay, established with the M-ABC (Table I).

Subgroup analysis

Subgroup analyses based on the results of developmental tests are presented in Tables II–IV. The prevalence of motor delays among infants born EPT was considerably higher than in infants born VPT (44.5%, 95% CI 14.2–79.5, I^2=96.70 vs 16.4%, 95% CI 11.1–23.7, I^2=85.92), although this was not statistically significant (p=0.093). Motor delays were also significantly (p=0.021) more present in ELBW infants than in VLBW infants (34.4%, 95% CI 18.5–54.6, I^2=94.12 vs 13.3%, 95% CI 7.6–22.2, I^2=88.79).

Country income, geographical region, and age at follow-up were identified as significant moderators of prevalence variability (p<0.05). On the other hand, the variability in prevalence estimates was not explained by follow-up rate and sample size (p>0.05).

Stratification by outcome measure (see Table IV) showed that studies using the BSID-II had significantly (p=0.010) higher prevalence rates than studies using the Bayley-III, when using the same cut-off values.

General developmental quotient

Some developmental scales only provide an overall general developmental quotient, such as the Griffith's developmental scales, which was used in three studies,[25,29,37] and the Brunet–Lézine test used in the study of Perivier et al.[40]

The estimated pooled prevalence for a general developmental quotient less than 1SD is 11.2% (I^2=96.30, p<0.001) and is reported in Table IV. The 95% prediction interval ranged from 4.7% to 24.6%, reflecting the between-study heterogeneity.

CP

Prevalences

In total, 25 of the included studies reported prevalences of CP.[20–24,26,28–39,42–47,49] Only six studies made a distinction between mild and moderate-to-severe CP, on the basis of

Table III: Pooled prevalences of cerebral palsy and motor and cognitive delay (based on developmental tests) by subgroup analysis

Subgroup	Categories	Number of studies	Event rate	Pooled prevalence (%)	95% CI	Heterogeneity I^2 (%), (p)	p value for difference
Overall cognitive delay							
Age at follow-up	18–24mo	11	298/1388	16.8	9.5–28.0	94.89 (p<0.001)	<0.001
	24–36mo	1	1/67	1.5	0.2–9.8	0.0	
	8–22mo	1	46/106	43.4	34.3–53.0	0.0	
	22–28mo	1	19/103	18.4	12.1–27.1	0.0	
Follow-up rate	<40%	1	24/126	19.0	13.1–26.8	0.0	0.276
	40%–70%	2	48/280	17.2	13.2–22.0	<0.001 (p=0.659)	
	70%–100%	10	279/1134	17.0	8.8–30.5	95.16 (p<0.001)	
	Not reported	1	13/124	10.5	6.2–17.2	0.0	
Sample size follow-up	<100	5	69/375	12.3	2.6–42.4	95.44 (p<0.001)	0.581
	100–500	9	295/1289	18.7	11.2–29.5	94.17 (p<0.001)	
Geographical region	Africa	1	46/106	43.4	34.3–53.0	0.0	<0.001
	Asia	2	37/150	14.5	7.9–25.2	71.77 (p=0.060)	
	Europe	5	46/485	6.9	2.9–15.5	82.11 (p<0.001)	
	North America	4	106/494	23.7	8.4–51.5	96.01 (p<0.001)	
	Oceania	1	19/103	18.4	12.1–27.1	0.0	
	South America	1	110/226	48.7	42.2–55.2	0.0	
Country income	High-income economy	12	208/1332	13.8	8.5–21.5	90.42 (p<0.001)	<0.001
	Upper-middle-income economy	2	156/332	47.0	41.7–52.4	<0.001 (p=0.369)	
Overall motor delay							
Age at follow-up	18–24mo	10	296/1233	21.2	13.4–31.9	91.98 (p<0.001)	<0.001
	24–36mo	1	1/67	1.5	0.2–9.8	0.0	
	8–22mo	1	40/106	37.7	29.0–47.3	0.0	
	22–28mo	1	19/102	18.6	12.2–27.4	0.0	
Follow-up rate	<40%	1	32/126	25.4	18.6–33.7	0.0	0.664
	40%–70%	2	63/279	22.4	16.9–29.0	30.00 (p=0.232)	
	70%–100%	9	236/979	17.7	9.1–31.5	93.52 (p<0.001)	
	Not reported	1	25/124	20.2	14.0–28.1	0.0	
Sample size follow-up	<100	5	80/375	12.1	1.8–51.0	95.99 (p<0.001)	0.423
	100–500	8	276/1133	24.1	18.8–30.4	78.38 (p<0.001)	
Geographical region	Africa	1	40/106	37.7	29.0–47.3	0.0	<0.001
	Asia	2	55/250	22.9	18.1–28.5	<0.001 (p=0.325)	
	Europe	4	11/330	3.0	1.0–8.5	57.67 (p<0.069)	
	North America	4	160/494	37.8	17.6–63.4	95.44 (p<0.001)	
	Oceania	1	19/102	18.6	12.2–27.4	0.0	
	South America	1	69/226	30.5	24.9–36.8	0.0	
Country income	High-income economy	11	247/1176	17.4	10.4–27.7	91.68 (p<0.001)	0.012
	Upper-middle-income economy	2	109/332	33.4	26.9–40.6	40.92 (p=0.193)	
Cerebral palsy							
Age at follow-up	18–24mo	19	622/6749	6.7	5.2–8.5	74.56 (p<0.001)	0.224
	22–28mo	2	8/273	2.4	0.2–20.3	80.13 (p=0.025)	
	24–48mo	3	94/721	10.6	5.7–18.6	66.47 (p=0.051)	
	8–22mo	1	4/106	3.8	1.4–9.6	0.0	
Follow-up rate	<40%	2	12/205	6.1	3.5–10.4	0.0 (p=0.328)	0.423
	40%–70%	5	111/988	9.1	5.3–15.3	77.5 (p=0.001)	
	70%–100%	17	600/6532	6.3	4.8–8.3	76.5 (p<0.001)	
	Not reported	1	5/124	4.0	1.7–9.3	0.0	
Sample size follow-up	<100	8	40/570	7.7	4.8–12.0	51.3 (p=0.045)	0.003
	100–500	14	113/2172	5.4	4.0–7.2	56.5 (p=0.005)	
	>500	3	575/5107	10.9	8.1–14.4	86.8 (p=0.001)	
Geographical region	Africa	2	12/166	7.5	2.1–23.6	78.4 (p=0.032)	0.345
	Asia	4	65/953	6.9	5.5–8.7	0.0 (p=0.574)	
	Europe	8	121/1380	5.9	3.2–10.5	80.4 (p<0.001)	
	North America	8	513/4936	7.6	5.4–10.5	74.2 (p<0.001)	
	Oceania	2	10/188	5.5	3.0–9.9	0.0 (p<0.001)	
	South America	1	7/226	3.1	1.5–6.4	0.0	
Country income	High-income economy	20	658/6754	7.0	5.5–8.9	75.1 (p<0.001)	0.419
	Upper-middle-income economy	3	59/980	4.9	2.5–9.1	67.8 (p=0.045)	
	Lower-middle-income economy	2	11/115	9.5	3.9–21.1	48.4 (p=0.164)	

Random-effects analysis. CI, confidence interval.

the Gross Motor Function Classification System.[23,32,37,38,45,46] As a consequence of the small number of articles reporting the degree of disability and the differences in classification (moderate-to-severe Gross Motor Function Classification System >2 or ≥2), no separated pooled prevalences were calculated. Neurological assessment was completed between 18 months and 3 years.

Three investigations found a CP prevalence lower than 1%,[20,29,47] while five others observed it to be more than 10%.[23,24,30,33,36]

Meta-analysis

The overall prevalence of CP in the 25 retrieved studies was 6.8% (95% CI 5.5–8.4, I^2=76.1%, p<0.001; see Fig. 4).

Subgroup analysis and meta-regression

Significant differences (p<0.001) in the overall prevalence rates were documented according to the mean gestational age, being significant higher for infants born EPT (10.0%, 95% CI 8.1–12.2, I^2=61.7, p=0.007) than for infants born VPT (4.5%, 95% CI 3.3–6.3, I^2=57.4, p=0.003), and to the mean birthweight, being greater for ELBW infants (8.4%, 95% CI 6.6–10.7, I^2=58.5, p=0.10) than for VLBW infants (4.2%, 95% CI 2.9–6.2, I^2=62.2, p=0.002). The pooled differences in prevalence rates following the possible comparisons within studies were all non-significant (p<0.05) except for sample size (p=0.003) (Table III).

The results of random-effects meta-regression analyses that assessed the relationship between the selected covariates and the observed prevalences in each single study are presented in Figures 5 and 6. There was a statistically significant linear trend that explained prevalence variation by mean birthweight and mean gestational age (p<0.001) with 33% and 35% respectively, of variance accounted for.

DISCUSSION

Main findings

This systematic review and meta-analysis was performed to supply an overview of separate prevalences of CP, as well as motor and cognitive delays in VPT and VLBW infants born in the past decade, and to evaluate the influence of gestational age and VLBW on these prevalences. Most of the included studies assessed motor and cognitive development near the age of 2 years corrected age. At this age, there is still a high proportion of parents motivated to attend follow-up visits and this is also the age where neurological problems can be reliably detected and most children with CP are diagnosed.[50]

Herein, it was estimated that 20.6% (95% CI 13.9–29.4) and 16.9% (95% CI 10.4–26.3) of VPT or VLBW infants respectively, developed a certain degree of motor or cognitive delay, on the basis of developmental scales at approximately 2 years corrected age. As expected, mild motor or cognitive delays were more frequent than moderate-to-severe delays (18.0% vs 8.6% and 14.3% vs 8.2% respectively).

Contrary to the definition of motor or cognitive delay, which varied considerably between articles, CP is a clearly defined criterion used as a touchstone for neurodevelopmental outcomes after preterm births and the quality of neonatal care. The overall estimated pooled prevalences for CP was 6.8% (95% CI 5.5–8.4) for all included articles. Since the heterogeneity of the articles was substantial (I^2>50), the specific rates should be interpreted with care.

A secondary objective of this study was to perform meta-regression to discern potential associations between the prevalence of CP, as well as motor and cognitive delays and mean birthweight and gestational age. Although overall prevalences of CP along with motor and cognitive delays were higher in ELBW infants than VLBW infants, this variability was only statically significant for CP (p<0.001) and motor delays (p=0.012). On the other hand, the subgroup analyses clearly indicated that the overall prevalence of CP and motor and cognitive delays rose with decreasing gestational age. However, this was only statistically significant for CP (p<0.001). These findings are in line with the results of a considerable number of studies where EPT and ELBW infants exhibited greater neurodevelopmental impairment than their older peers.[8,14,51,52]

As the prevalence of CP was elevated with decreasing gestational age[7] and more infants born EPT survived to discharge, this could be proportionally the most concerning group. In this systematic review, three of the included articles had a mean gestational age of less than 26 weeks in their cohort,[24,33,34] which also included infants born at 22 to 23 weeks' gestation and reported prevalences varying between 9.4% and 14.4%. In total, the overall pooled prevalence rate of CP in EPT was 10.0% (95% CI 8.1–12.2). Himpens et al.[7] reported a weighted prevalence of 14.4% in infants born EPT, on the basis of all articles with a birth year earlier than 2006. An update of this meta-analysis by Oskoui et al.[15] included articles from 1985 until 2011 and showed that the overall rate remained constant. As such, our meta-analysis, which only included recent articles, could verify the decreasing trend of CP over recent years in infants born EPT as a direct consequence of improved neonatal care.[53–56]

Only five articles fulfilling our inclusion criteria reported longer-term outcomes up to the age of 6 years. Large differences in prevalences were seen between articles reporting motor delays based on developmental scales, such as the BSID-II and Bayley-III tests, and motor tests such as the M-ABC. Motor delays assessed by those developmental scales were estimated to be 20.6% (95% CI 13.9–29.4) and rose to between 34% and 40% of motor delays when evaluated with the M-ABC at a preschool age.[27,29] This corresponds to the study of Spittle et al.,[57] which concluded that the Bayley-III underestimates later rates of motor performance delays evaluated with the M-ABC. This could be understood by the fact that the Bayley-III assesses current levels of motor development rather than basic milestones whereas the M-ABC focuses on specific motor function tasks in various categories (e.g. manual dexterity, aiming and catching, balance tasks). The previous literature has suggested that motor milestones could be more easily attained than advanced motor skills.[58] The same underestimation with the BSID-II was observed for later cognitive delays.[59] In accordance with this, we found that the pooled prevalences of moderate-to-severe cognitive delay evaluated with the BSID-II and Bayley-III were considerably lower than moderate-to-severe delay determined with the Wechsler Preschool and Primary Scale of Intelligence Test, Third Edition (8.2%, 95% CI 5.5–12.0 vs 14.7%, 95% CI 10.9–19.5).

Table IV: Pooled prevalences of motor, cognitive, and general developmental delay (based on developmental tests) by outcome measures and cut-off values

Cut-off value	Cut-off value	Outcome measures	Number of studies	Event rate	Pooled prevalence (%)	95% CI	Heterogeneity I^2 (%), (p)	p value for difference	
Cognitive delay									
Mild delay	Index/CS 70–84	BSID-II	1	14/126	11.1	6.7–17.9	0.0	0.817	
		Bayley-III	7	112/824	12.1	7.0–20.2	86.69 (p<0.001)		<0.001
		BSID-II+Bayley-III	8	126/950	12.1	7.5–19.0	85.07 (p<0.001)		
	CS 80–94	Bayley-III	1	31/65	47.7	35.9–59.7	0.0		
	Total	BSID-II+Bayley-III	9	157/1015	14.3	8.3–23.5	90.49 (p<0.001)		
Moderate-to-severe delay	Index/CS<70	BSID-II	2	151/774	13.9	4.9–33.7	91.42 (p=0.001)	0.185	
		Bayley-III	11	442/5202	6.5	4.9–10.9	85.36 (p<0.001)		<0.001
		BSID-II+Bayley-III	13	593/5976	7.5	4.9–11.2	92.49 (p<0.001)		
	CS<80	Bayley-III	1	14/65	21.5	13.2–33.2	0.0		
	Total	BSID-II+Bayley-III	14	607/6041	8.2	5.5–12.0	92.20 (p<0.001)		
Overall delay	Index/CS<85	BSID-II	2	134/352	32.3	10.9–65.2	96.44 (p<0.001)	0.104	
		Bayley-III	11	185/1247	12.6	8.1–19.0	87.56 (p<0.001)		<0.001
		BSID-II+Bayley-III	13	319/1599	14.6	9.1–22.6	93.11 (p<0.001)		
	CS<95	Bayley-III	1	45/65	69.2	57.1–79.2	0.0		
	Total	BSID-II+Bayley-III	14	364/1664	16.9	10.4–26.3	94.22 (p<0.001)		
Motor delay									
Mild delay	Index/CS 70–84	BSID-II	1	19/126	15.1	9.8–22.4	0.0	0.633	
		Bayley-III	5	107/601	17.2	11.9–24.3	74.67 (p=0.003)		<0.001
		BSID-II+Bayley-III	6	126/727	16.9	12.4–22.7	69.97 (p=0.005)		
	CS 80–94	Bayley-III	1	34/65	52.3	40.3–64.1	0.0		
	Scaled score 4–7	Bayley-III	1	1/67	1.5	0.2–9.8	0.0		
	Total	BSID-II+Bayley-III+PDMS-II	8	161/859	18.0	11.1–27.8	88.53 (p<0.001)		
Moderate-to-severe delay	Index/CS<70	BSID-II	1	13/126	10.3	6.1–17.0	0.0	0.291	
		Bayley-III	9	400/3774	7.4	5.2–10.4	80.18 (p<0.001)		<0.001
		BSID-II+Bayley-III	10	413/3900	7.7	5.6–10.5	77.76 (p<0.001)		
	CS<80	Bayley-III	1	22/65	33.8	23.4–46.1	0.0		
	Scaled score<4	Bayley-III	1	0/67	0.7	0.0–10.7	0.0		
	Total	BSID-II+Bayley-III+PDMS-II	12	435/4032	8.6	6.0–12.1	84.77 (p<0.001)		
Overall delay	Index/CS<85	BSID-II	2	101/352	28.7	24.1–33.8	3.82 (p=0.308)	0.010	
		Bayley-III	9	198/1024	17.0	11.5–24.4	84.72 (p<0.001)		<0.001
		BSID-II+Bayley-III	11	299/1376	19.8	14.9–25.9	82.64 (p<0.001)		
	CS<95	Bayley-III	1	56/65	86.2	75.5–92.6	0.0		
	Scaled score ≤7	Bayley-III	1	1/67	1.5	0.2–9.8	0.0		
	Total	BSID-II+Bayley-III+PDMS-II	13	356/1508	20.6	13.9–29.4	90.91 (p<0.001)		
General development									
Mild delay	GQ<1SD>2SD	GDS	2	58/432	7.8	0.9–42.6	94.35 (p<0.001)		
Moderate-to-severe delay	GQ<2SD	GDS	2	33/432	4.3	0.5–29.3	89.53 (p=0.002)		
Overall delay	GQ<1SD	GDS	3	100/530	11.2	2.7–36.4	95.25 (p<0.001)		
		BL	1	101/968	10.4	8.7–12.5	0.0		
	Total	GDS+BL	4	201/1498	11.2	4.7–24.6	96.30 (p<0.001)		

Random-effects analysis. CI, confidence interval; CS, composite score; BSID-II, Bayley Scales of Infant Development, Second Edition; Bayley-III, Bayley Scales of Infant and Toddler Development, Third Edition; PDMS-II, Peabody Developmental Motor Scales, Second Edition; GQ, general quotient; GDS, Griffith Developmental Scale; BL, Brunet–Lézine test.

It seems that as VPT and VLBW infants get older, more cognitive and motor delays become apparent as a result of increasing functional demands in daily life and school activities and because of the use of more specific function-related assessment tools instead of developmental outcome measurements. Therefore the results of developmental tests performed before or around 2 years of age should be treated with caution as the predictive value for later motor or cognitive delays is limited.[60]

To our knowledge, no similar previously unified data on separate motor or cognitive outcome prevalences in VPT or VLBW infants exist. Consequently, no possible evolution in time can be reported.

Study strengths and limitations

The major strength of this systematic review is that a literature search was performed in four different databases to identify all relevant articles. Articles were included on the basis of birthweight and gestational age, which may have caused more heterogeneity between studies but ensured that more pertinent publications were included. Certain groups within the preterm population are at greater risk of developing neurodevelopmental delays; therefore only sequential total cohort studies of VPT and/or VLBW infants were included, which reduced sampling bias. Additionally, because the sample size was set at a minimum of 50 infants, less representative studies were left out.

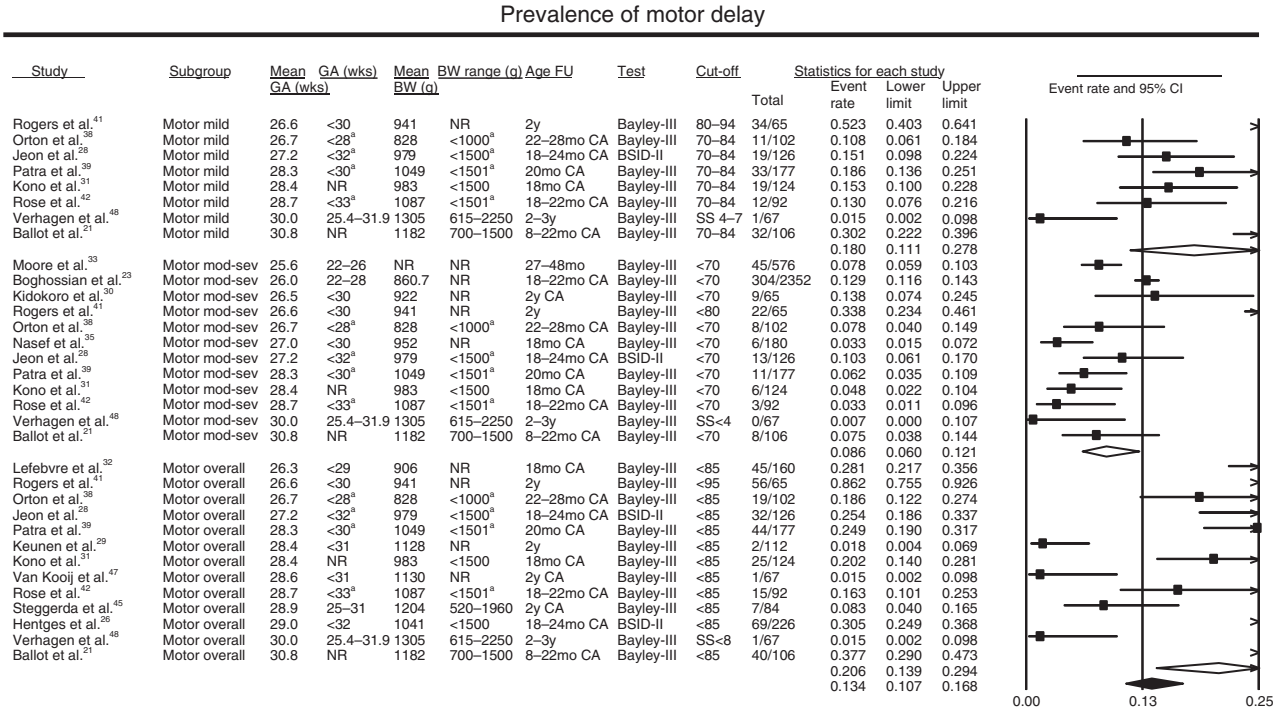

Figure 3: Prevalences of motor delays. Forest plot depicting the random-effects proportion meta-analysis for motor delays, on the basis of developmental scales (motor subscore of the BSID-II and Bayley-III). Studies are ordered on mean gestation. Black squares denote the reported prevalence of each study and the horizontal line represents the 95% confidence interval. The pooled prevalence estimate is marked with a diamond. [a]Birthweight range and/or gestational age range. GA, gestational age; BW, birthweight; FU, follow-up; CI, confidence interval; NR, not reported; Bayley-III, Bayley Scales of Infant and Toddler Development, Third Edition; CA, corrected age; BSID-II, Bayley Scales of Infant Development, Second Edition; SS, scaled scores.

This systematic review also had several limitations that need to be considered when interpreting our findings. As our review featured clinically and methodologically very diverse studies, it is not surprising that high heterogeneity (I^2>75) was found for each individual outcome. This heterogeneity could have arisen from many different factors such as inclusion and exclusion criteria, length of follow-up, outcome measures used, etc. An example worth mentioning in this regard is the fact that just a few studies excluded infants with congenital malformations and genetic disorders. Some of those disorders could be associated with an increased risk of an adverse neurodevelopmental outcome, skewing the results towards higher prevalences of neurodevelopmental delay.

Several other sources of heterogeneity were explored into more detail, although this could never explain the entirety of the variance in the outcomes.

First, our review featured articles from all over the world, varying between low- and high-income countries and representative of important differences in religions, health systems, and norms surrounding active neonatology care. Most of the included articles originated from high-income countries where, in general, the prevalence of preterm infants was lower and the survival rate was higher.[1] Upper-middle-income countries reported significant (p<0.001) higher pooled prevalences of cognitive and motor delays than high-income countries. Stratification by region resulted in a significant variance for motor and cognitive delays (p<0.001). The reported prevalences were systematically the lowest in Europe and the highest in South America for cognitive delay, and North America and Africa for motor delay. No consistent results could be determined for income level and geographical region with respect to the prevalence of CP. Nevertheless, the number of studies or the sample size, in particular covariate subgroups, may be too sparse to arrive at robust conclusions.

Second, different outcome measurements and cut-off values were used, creating a serious challenge for this review in terms of cataloguing all outcome data into mild-to-severe developmental delays. Even with the most widely used assessment tool, the BSID, inequity is often observed between the Second and Third Editions. Recent studies have reported higher scores for the Bayley-III than the BSID-II.[13,61] Consequently, fewer infants were classified as moderately and severely impaired. Our results were consistent with these observations. It was found that pooled prevalences for moderate-to-severe motor and cognitive delays were higher when evaluated using the BSID-II than the Bayley-III, while the opposite was observed within the mild category. It is unclear whether the BSID-II

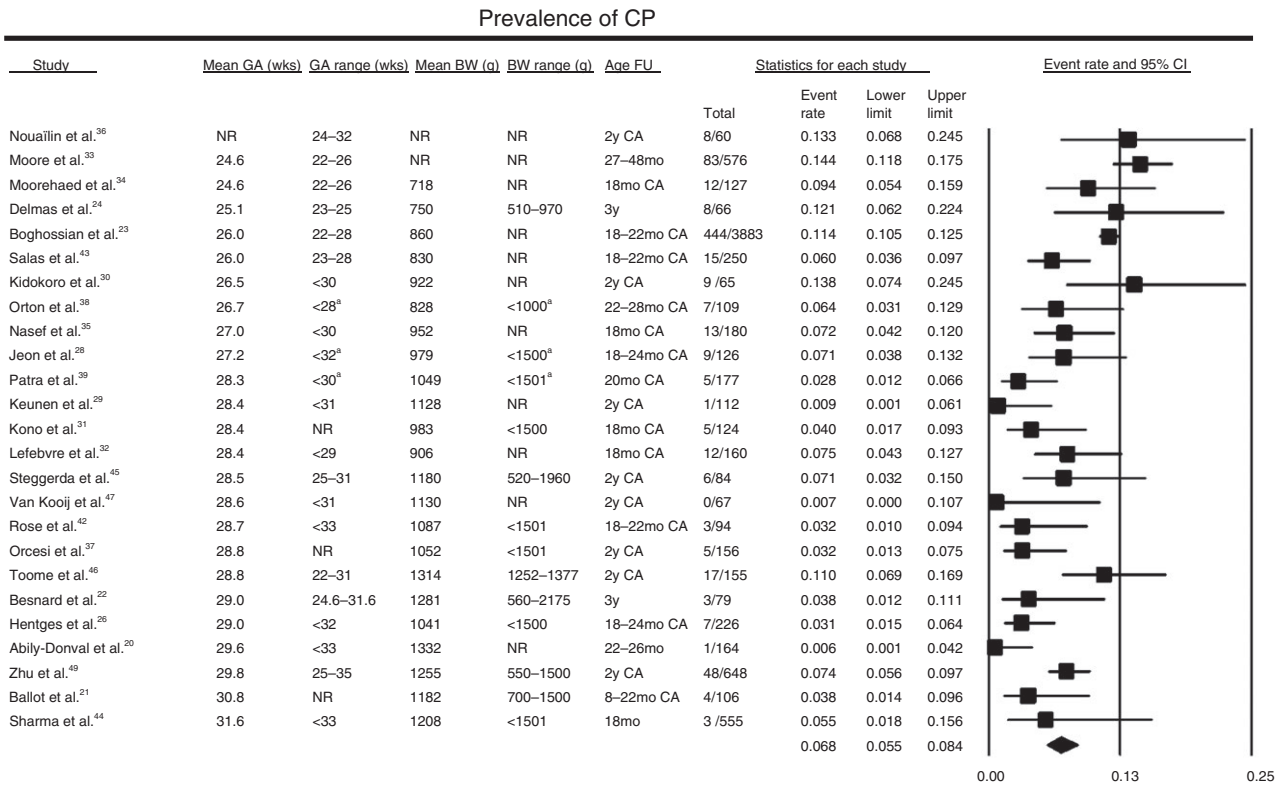

Figure 4: Prevalences of cerebral palsy (CP). Forest plot depicting the random-effects proportion meta-analysis for CP. Studies are ordered on mean gestation. Black squares denote the reported prevalence of each study and the horizontal line represents the 95% confidence interval. The pooled prevalence estimate is marked with a diamond. [a]Birthweight range and/or gestational age range. GA, gestational age; BW, birthweight; FU, follow-up; NR, not reported.

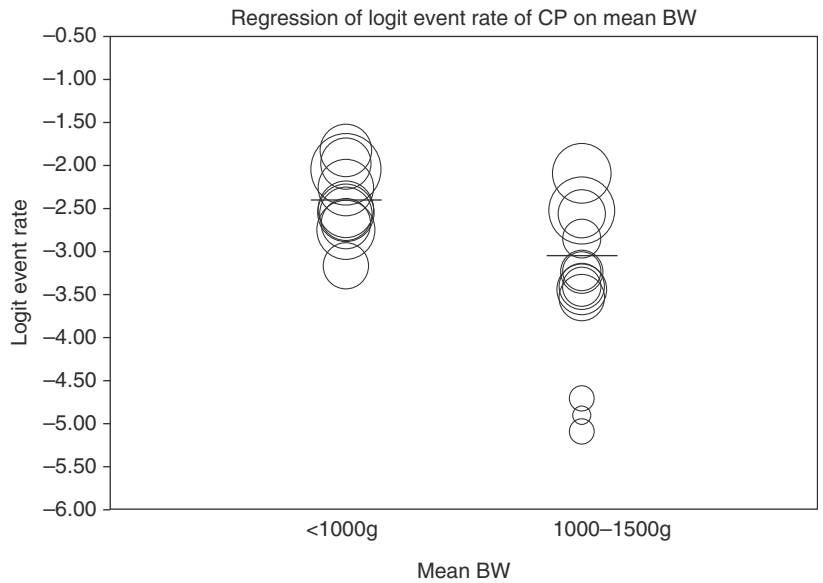

Figure 5: Regression of mean birthweight (BW) on logit event rate of cerebral palsy (CP). Scatter-plot representation of the relationship between gestational age and the prevalence of CP. Each circle represents the results of a study.

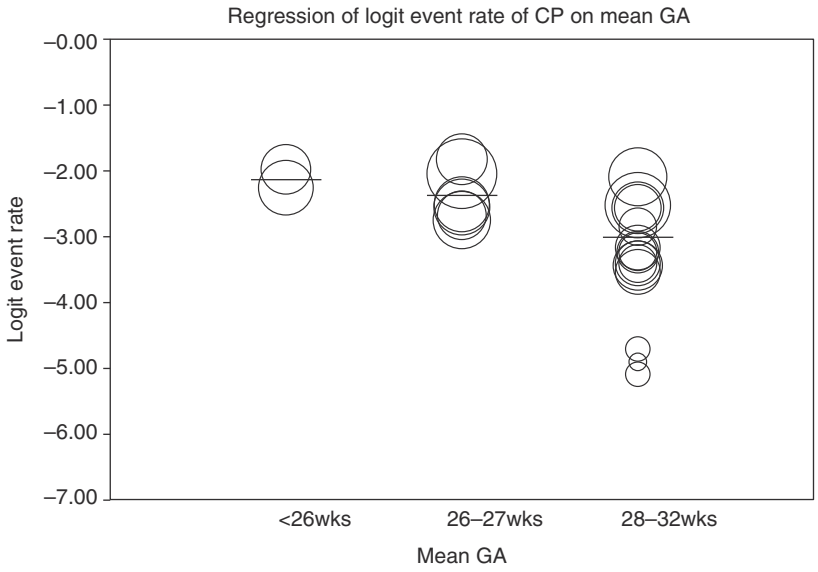

Figure 6: Regression of mean gestational age (GA) on logit event rate of cerebral palsy (CP). Scatter-plot representation of the relationship between birthweight and the prevalence of CP. Each circle represents the results of a study.

underestimates or the Bayley-III overestimates development. To correct for this, it is advised to increase the cut-off values for the Bayley-III.[62] Only one study followed this advice and applied higher cut-off values for the Bayley-III, specifically composite scores of 80 and 65 instead of the 70 and 55 values generally used respectively.[41] This could explain why this study exhibited the highest prevalence of all the included studies on the subscales of motor and cognitive delays of the Bayley-III. Moreover, as described in a recent systematic review of the cross-cultural validity of assessment tools, the use of standardized norms should be used and interpreted with caution across varying cultures versus the initial samples.[63] For example, one included study was based on an Australian sample,[37] and used standardized American Bayley-III norms, although it is suggested that this would considerably underestimate developmental delays.[64]

Third, the follow-up rate and sample size varied considerably between the studies. The mean follow-up rate was 77.7%, meaning that nearly one out of five infants was not seen at follow-up. Furthermore, two of the 28 articles reporting the number of eligible infants for follow-up had a follow-up rate of less than 50%.[22,28] However, subgroup analyses based on the percentage of eligible infants that had follow-up were not statistically different for CP or motor and cognitive delays ($p>0.005$). Only the prevalence of CP was significantly influenced by the sample size ($p=0.003$), but no linear trend could be observed. Of all included articles, only 12 reported information about statistics between the infants followed up and the groups lost to follow-up. Six articles found no significant differences in neonatal characteristics,[20,28,29,32,37,47] and one reported no difference in maternal characteristics.[32] In contrast, six articles noted significant differences between both groups.[24,25,29,38–40] Infants included in the follow-up had significantly ($p<0.05$) lower mean birthweight[25,29,38] and gestational age,[25,38,40] and were more severely ill (days on mechanical ventilation,[25,29] sepsis,[25,29,38] infection,[24,40] bronchopulmonary dysplasia,[25,39] chorioamnionitis,[24] chronic lung disease,[38] and inferior neuromotor examination at discharge).[40] Two studies observed differences in maternal characteristics.[24,39] Delmas et al.[24] found non-significant differences between both groups with respect to higher maternal education (60% in the follow-up group and 35.3% for the lost-to-follow-up group, $p=0.074$); and in the study of Patra et al.,[39] mothers were slightly younger in the lost-to-follow-up group (not significant). One investigation noted that, in the follow-up group, significantly more parents living in metropolitan areas were represented ($p=0.02$) compared with parents living in rural areas.[38] This could validate the hypothesis that infants with no or mild disabilities may be more likely to be lost to follow-up as parents determine there is less of a benefit from returning for it. Further, this could potentially have biased the results towards a greater prevalence of more severe delays.

Implications for future research

In line with a recent paper,[65] this systematic review has highlighted the strong need for uniformization of the used assessment tools and cut-off values to be able to compare studies more accurately. Recent large epidemiological studies such as EXPRESS,[66] EPICure,[33] and EPIPAGE[67] have demonstrated how this is necessary to reach solid conclusions. Finally, more long-term follow-up is required at preschool ages, since other difficulties can be observed, such as visual–motor integration or coordination problems.

CONCLUSION

To our knowledge, this is the first systematic review and meta-analysis to separately demarcate the prevalence of both motor and of cognitive delays in VPT or VLBW infants born over the past decade. Even though neonatal intensive care has improved over the previous few decades, the data from this meta-analysis suggest that, overall, nearly one out of six and one out of five VPT or VLBW infants had a cognitive or motor delay respectively, assessed with developmental scales at approximately 2 years corrected age and roughly one out of fifteen developed CP. Decreasing birthweight and gestational age led to higher prevalences of CP, as well as motor and cognitive delays. It was also shown that overall prevalences of CP diminished over the years in infants born EPT. As a result of the notable heterogeneity between the articles and the wide confidence intervals, the results should be interpreted with care.

ACKNOWLEDGEMENTS

The authors have stated that they had no interests that might be perceived as posing a conflict or bias.

SUPPORTING INFORMATION

The following additional material may be found online:

Appendix S1: Search strategy

Appendix S2: Quality assessment of the included studies

Appendix S3: Characteristics of the included studies

REFERENCES

1. Blencowe H, Cousens S, Chou D, et al. Born too soon: the global epidemiology of 15 million preterm births. *Reprod Health* 2013; **10**(Suppl. 1): S2.
2. Harrison MS, Goldenberg RL. Global burden of prematurity. *Semin Fetal Neonatal Med* 2016; **21**: 74–9.
3. Grisaru-Granovsky S, Reichman B, Lerner-Geva L, et al. Population-based trends in mortality and neonatal morbidities among singleton, very preterm, very low birth weight infants over 16 years. *Early Hum Dev* 2014; **90**: 821–7.
4. Hack M, Costello DW. Trends in the rates of cerebral palsy associated with neonatal intensive care of preterm children. *Clin Obstet Gynecol* 2008; **51**: 763–74.
5. Stoll BJ, Hansen NI, Bell EF, et al. Trends in care practices, morbidity, and mortality of extremely preterm neonates, 1993-2012. *JAMA* 2015; **314**: 1039–51.
6. Wilson-Costello D. Is there evidence that long-term outcomes have improved with intensive care? *Semin Fetal Neonatal Med* 2007; **12**: 344–54.
7. Himpens E, Van den Broeck C, Oostra A, Calders P, Vanhaesebrouck P. Prevalence, type, distribution, and severity of cerebral palsy in relation to gestational age: a meta-analytic review. *Dev Med Child Neurol* 2008; **50**: 334–40.
8. Sansavini A, Savini S, Guarini A, Broccoli S, Alessandroni R, Faldella G. The effect of gestational age on developmental outcomes: a longitudinal study in the first 2 years of life. *Child Care Health Dev* 2011; **37**: 26–36.
9. Saigal S, Doyle LW. An overview of mortality and sequelae of preterm birth from infancy to adulthood. *Lancet* 2008; **371**: 261–9.
10. Younge N, Goldstein RF, Bann CM, et al. Survival and neurodevelopmental outcomes among periviable infants. *N Engl J Med* 2017; **376**: 617–28.
11. Moore GP, Lemyre B, Barrowman N, Daboval T. Neurodevelopmental outcomes at 4 to 8 years of children born at 22 to 25 weeks' gestational age: a meta-analysis. *JAMA Pediatr* 2013; **167**: 967–74.
12. Pierrat V, Marchand-Martin L, Arnaud C, et al. Neurodevelopmental outcome at 2 years for preterm children born at 22 to 34 weeks' gestation in France in 2011: EPIPAGE-2 cohort study. *BMJ* 2017; **358**: j3448.
13. Serenius F, Källén K, Blennow M, et al. Neurodevelopmental outcome in extremely preterm infants at 2.5 years after active perinatal care in Sweden. *JAMA* 2013; **309**: 1810–20.
14. Blencowe H, Lee AC, Cousens S, et al. Preterm birth-associated neurodevelopmental impairment estimates at regional and global levels for 2010. *Pediatr Res* 2013; **74** (Suppl. 1): 17–34.
15. Oskoui M, Coutinho F, Dykeman J, Jetté N, Pringsheim T. An update on the prevalence of cerebral palsy: a systematic review and meta-analysis. *Dev Med Child Neurol* 2013; **55**: 509–19.
16. Mwaniki MK, Atieno M, Lawn JE, Newton CR. Long-term neurodevelopmental outcomes after intrauterine and neonatal insults: a systematic review. *Lancet* 2012; **379**: 445–52.
17. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg* 2010; **8**: 336–41.
18. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. A basic introduction to fixed-effect and random-effects models for meta-analysis. *Res Synth Meth* 2010; **1**: 97–111.
19. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003; **327**: 557–60.
20. Abily-Donval L, Pinto-Cardoso G, Chadie A, et al. Comparison in outcomes at two-years of age of very preterm infants born in 2000, 2005 and 2010. *PLoS ONE* 2015; **10**: e0114567.
21. Ballot DE, Potterton J, Chirwa T, Hilburn N, Cooper PA. Developmental outcome of very low birth weight infants in a developing country. *BMC Pediatr* 2012; **12**: 11.
22. Besnard M, Kuo P, Pawlotsky F, Guyot D, Elie V, Papouin-Rauzy M. [Very preterm births in French Polynesia: update and proposal for follow-up](In French). *Arch Pediatr* 2015; **22**: 160–5.
23. Boghossian NS, Hansen NI, Bell EF, et al. Outcomes of extremely preterm infants born to insulin-dependent diabetic mothers. *Pediatrics* 2016; **137**: e20153424.
24. Delmas O, Garcia P, Bernard V, et al. [Neurodevelopmental outcome at 3 years of age of infants born at less than 26 weeks] (In French). *Arch Pediatr* 2016; **23**: 927–34.
25. Gibertoni D, Corvaglia L, Vandini S, et al. Positive effect of human milk feeding during NICU hospitalization on 24 month neurodevelopment of very low birth weight infants: an Italian cohort study. *PLoS ONE* 2015; **10**: e0116552.
26. Hentges CR, Silveira RC, Procianoy RS, et al. Association of late-onset neonatal sepsis with late neurodevelopment in the first two years of life of preterm infants with very low birth weight. *J Pediatr (Rio J)* 2014; **90**: 50–7.
27. Howe TH, Sheu CF, Wang TN, Hsu YW, Wang LW. Neuromotor outcomes in children with very low birth weight at 5 yrs of age. *Am J Phys Med Rehabil* 2011; **90**: 667–80.
28. Jeon TY, Kim JH, Yoo SY, et al. Neurodevelopmental outcomes in preterm infants: comparison of infants with and without diffuse excessive high signal intensity on MR images at near-term-equivalent age. *Radiology* 2012; **263**: 518–26.
29. Keunen K, Išgum I, van Kooij BJM, et al. Brain volumes at term-equivalent age in preterm infants: imaging biomarkers for neurodevelopmental outcome through early school age. *J Pediatr* 2016; **172**: 88–95.
30. Kidokoro H, Anderson PJ, Doyle LW, Woodward LJ, Neil JJ, Inder TE. Brain injury and altered brain growth in preterm infants: predictors and prognosis. *Pediatrics* 2014; **134**: e444–53.
31. Kono Y, Yonemoto N, Kusuda S, et al. Developmental assessment of VLBW infants at 18 months of age: a comparison study between KSPD and Bayley III. *Brain Dev* 2016; **38**: 377–85.
32. Lefebvre F, Gagnon MM, Luu TM, Lupien G, Dorval V. In extremely preterm infants, do the Movement Assessment of Infants and the Alberta Infant Motor Scale predict 18-month outcomes using the Bayley-III? *Early Hum Dev* 2016; **94**: 13–7.
33. 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.
34. Moorehead PA, Martin EM, Shepherd EG, et al. Factors associated with survival of <27 week infants in an all-referral neonatal intensive care unit. *J Neonat-Perinat Med* 2012; **5**: 105–11.

35. Nasef N, Shabaan AE, Schurr P, et al. Effect of clinical and histological chorioamnionitis on the outcome of preterm infants. *Am J Perinatol* 2013; **30**: 59–68.

36. Ben Hamida Nouaïli E, Chaouachi S, Bouziri A, Ben Jeballah N, Marrakchi Z. [Outcome at 2 years of very premature infants. A Tunisian series] (In French). *Tunisie Medicale* 2011; **89**: 682–5.

37. 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.

38. Orton JL, McGinley JL, Fox LM, Spittle AJ. Challenges of neurodevelopmental follow-up for extremely preterm infants at two years. *Early Hum Dev* 2015; **91**: 689–94.

39. Patra K, Greene MM, Patel AL, Meier P. Maternal education level predicts cognitive, language, and motor outcome in preterm infants in the second year of life. *Am J Perinatol* 2016; **33**: 738–44.

40. Périvier M, Rozé JC, Gascoin G, et al. Neonatal EEG and neurodevelopmental outcome in preterm infants born before 32 weeks. *Arch Dis Child Fetal Neonatal Ed* 2016; **101**: F253–9.

41. Rogers CE, Smyser T, Smyser CD, Shimony J, Inder TE, Neil JJ. Regional white matter development in very preterm infants: perinatal predictors and early developmental outcomes. *Pediatr Res* 2016; **79**: 87–95.

42. Rose J, Cahill-Rowley K, Vassar R, et al. Neonatal brain microstructure correlates of neurodevelopment and gait in preterm children 18-22 mo of age: an MRI and DTI study. *Pediatr Res* 2015; **78**: 700–8.

43. Salas AA, Faye-Petersen OM, Sims B, et al. Histological characteristics of the fetal inflammatory response associated with neurodevelopmental impairment and death in extremely preterm infants. *J Pediatr* 2013; **163**: 652–7.e1–2.

44. Sharma PK, Sankar MJ, Sapra S, et al. Growth and neurosensory outcomes of preterm very low birth weight infants at 18 months of corrected age. *Indian J Pediatr* 2011; **78**: 1485–90.

45. Steggerda SJ, De Bruïne FT, van den Berg-Huysmans AA, et al. Small cerebellar hemorrhage in preterm infants: perinatal and postnatal factors and outcome. *Cerebellum* 2013; **12**: 794–801.

46. Toome L, Varendi H, Männamaa M, Vals MA, Tänavsuu T, Kolk A. Follow-up study of 2-year-olds born at very low gestational age in Estonia. *Acta Paediatr* 2013; **102**: 300–7.

47. van Kooij BJ, van Pul C, Benders MJ, van Haastert IC, de Vries LS, Groenendaal F. Fiber tracking at term displays gender differences regarding cognitive and motor outcome at 2 years of age in preterm infants. *Pediatr Res* 2011; **70**: 626–32.

48. Verhagen EA, Van Braeckel KN, van der Veere CN, et al. Cerebral oxygenation is associated with neurodevelopmental outcome of preterm children at age 2 to 3 years. *Dev Med Child Neurol* 2015; **57**: 449–55.

49. Zhu JJ, Bao YY, Zhang GL, Ma LX, Wu MY. No relationship between mode of delivery and neonatal mortality and neurodevelopment in very low birth weight infants aged two years. *World J Pediatr* 2014; **10**: 227–31.

50. Ashwal S, Russman BS, Blasco PA, et al. Practice parameter: diagnostic assessment of the child with cerebral palsy: report of the Quality Standards Subcommittee of the American Academy of Neurology and the Practice Committee of the Child Neurology Society. *Neurology* 2004; **62**: 851–63.

51. Larroque B, Ancel PY, Marret S, et al. Neurodevelopmental disabilities and special care of 5-year-old children born before 33 weeks of gestation (the EPIPAGE study): a longitudinal cohort study. *Lancet* 2008; **371**: 813–20.

52. Agarwal P, Lim SB. Long-term follow-up and outcome of extremely-low-birth-weight (ELBW) infants. *Ann Acad Med Singapore* 2003; **32**: 346–53.

53. Platt MJ, Cans C, Johnson A, et al. Trends in cerebral palsy among infants of very low birthweight (<1500 g) or born prematurely (<32 weeks) in 16 European centres: a database study. *Lancet* 2007; **369**: 43–50.

54. Robertson CM, Watt MJ, Yasui Y. Changes in the prevalence of cerebral palsy for children born very prematurely within a population-based program over 30 years. *JAMA* 2007; **297**: 2733–40.

55. Sellier E, Platt MJ, Andersen GL, et al. Decreasing prevalence in cerebral palsy: a multi-site European population-based study, 1980 to 2003. *Dev Med Child Neurol* 2016; **58**: 85–92.

56. van Haastert IC, Groenendaal F, Uiterwaal CS, et al. Decreasing incidence and severity of cerebral palsy in prematurely born children. *J Pediatr* 2011; **159**: 86–91 e1.

57. 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.

58. 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.

59. Hack M, Taylor HG, Drotar D, et al. Poor predictive validity of the Bayley Scales of Infant Development for cognitive function of extremely low birth weight children at school age. *Pediatrics* 2005; **116**: 333–41.

60. Leversen KT, Sommerfelt K, Elgen IB, et al. Prediction of outcome at 5 years from assessments at 2 years among extremely preterm children: a Norwegian national cohort study. *Acta Paediatr* 2012; **101**: 264–70.

61. Moore T, Johnson S, Haider S, Hennessy E, Marlow N. Relationship between test scores using the second and third editions of the Bayley Scales in extremely preterm children. *J Pediatr* 2012; **160**: 553–8.

62. Jary S, Whitelaw A, Walløe L, Thoresen M. Comparison of Bayley-2 and Bayley-3 scores at 18 months in term infants following neonatal encephalopathy and therapeutic hypothermia. *Dev Med Child Neurol* 2013; **55**: 1053–9.

63. Mendonça B, Sargent B, Fetters L. Cross-cultural validity of standardized motor development screening and assessment tools: a systematic review. *Dev Med Child Neurol* 2016; **58**: 1213–22.

64. Anderson PJ, De Luca CR, Hutchinson E, Roberts G, Doyle LW; Victorian Infant Collaborative Group. Underestimation of developmental delay by the new Bayley-III Scale. *Arch Pediatr Adolesc Med* 2010; **164**: 352–6.

65. Rysavy MA, Marlow N, Doyle LW, et al. Reporting outcomes of extremely preterm births. *Pediatrics* 2016; **138**: e20160689.

66. EXPRESS Group, Fellman V, Hellström-Westas L, et al. One-year survival of extremely preterm infants after active perinatal care in Sweden. *JAMA* 2009; **301**: 2225–33.

67. Larroque B, Bréart G, Kaminski M, et al. Survival of very preterm infants: Epipage, a population based cohort study. *Arch Dis Child Fetal Neonatal Ed* 2004; **89**: F139–44.