The challenges of heterogeneity in gestational age and birthweight inclusion criteria for research synthesis on very preterm birth and childhood cognition: An umbrella review and meta-regression analysis

Mariane Sentenac, Anna Chaimani, Sabrina Twilhaar, Valérie Benhammou, Samantha Johnson, Andrei Morgan, Jennifer Zeitlin

The challenges of heterogeneity in gestational age and birthweight inclusion criteria for research synthesis on very preterm birth and childhood cognition: An umbrella review and meta-regression analysis

Short running title: Study criteria on very preterm births

Mariane Sentenac[1], Anna Chaimani[2], Sabrina Twilhaar[1], Valérie Benhammou[1], Samantha Johnson[3], Andrei Morgan[1,4], Jennifer Zeitlin[1]

[1] Université de Paris, CRESS, Obstetrical Perinatal and Pediatric Epidemiology Research Team, EPOPé, INSERM, INRA, F-75004 Paris, France

[2] Université de Paris, Research Centre of Epidemiology and Statistics (CRESS-U1153), INSERM, Paris, France

[3] Department of Health Sciences, University of Leicester, Leicester, United Kingdom.

[4] Department of Neonatal Medicine, Maternité Port-Royal, Paris, 75014, France

Corresponding author

Mariane Sentenac

INSERM 1153, Obstetrical Perinatal and Pediatric Epidemiology Research Team

Maternité Port-Royal

53 Avenue de l'Observatoire

75014 Paris, France

E-mail: mariane.sentenac@inserm.fr

Word count: 3052 words

SOCIAL MEDIA QUOTE

These results suggest that meta-analyses on the consequences of very preterm birth would benefit from using gestational age and birthweight criteria to increase the number of studies and the generalizability of results

SYNOPSIS

Study question

Do gestational age (GA) and birthweight (BW) inclusion criteria in studies on very preterm birth and cognition affect the results of meta-analyses?

What's already known

Meta-analyses consistently show the negative consequences of very preterm birth on cognitive development in childhood but with large unexplained between-study heterogeneity.

What this study adds

This study showed that the use of gestational age vs birthweight criteria was more common in European and recent studies, with no substantial effect on meta-analysis results after adjustment for degree of prematurity. These results suggest that meta-analyses on the consequences of very preterm birth would benefit from broad inclusion criteria to increase the number of studies and the generalisability of results.

ABSTRACT

Background: Meta-analyses of studies on very preterm (VPT) birth and childhood cognition are essential for informing clinical practice and policy. Some reviews select primary studies using gestational age inclusion criteria only, while others also include birthweight criteria. The consequences of this choice are unknown.

Objective: To describe the gestational age (GA) and birthweight (BW) criteria used in studies of VPT birth and cognition, and investigate whether meta-analysis results differ based on these criteria.

Data sources: Primary studies from five systematic reviews on VPT birth and childhood IQ.

Study selection and data extraction: Country, birth years, GA/BW selection criteria and participant IQ were extracted from 156 studies representing 103 birth cohorts.

Synthesis: Pooled standardized mean differences (SMD) in IQ between cases and controls were estimated by sub-group based on GA/BW criteria (GA, BW, GA/BW combined) and degree of prematurity: extremely preterm (EPT, <28 weeks(w)) and extremely low (EL)BW(<1000 grams(g)); VPT(<32w) and very low (VL)BW(<1500g); and moderately (M)PT(<34w) and moderately low (ML)BW(<1800g).

Results

Cohorts used 27 distinct GA/BW inclusion criteria. Most common criteria were BW<1500g (24 cohorts), BW<1000g (12), GA<32w (12) and GA<33w (12); 23 studies used GA/BW

combinations. BW-only criteria were more frequent in North America than Europe (63% vs 24%), and for cohorts before rather than after 1990 (67% vs 26%). Pooled SMD in IQ varied with the degree of prematurity ($SMD_{EPT/ELBW}$=-0.94, 95%CI -1.07, -0.82; $SMD_{VPT/VLBW}$=-0.78, 95%CI -0.85, -0.71; $SMD_{MPT/MLBW}$=-0.68, 95%CI -0.79, -0.57), but there was no difference in SMD between cohorts using BW compared to GA criteria after adjustment on risk group.

Conclusions

Our results support the inclusion of studies using GA and/or BW criteria in meta-analyses on VPT birth and cognition to increase the geographical and temporal generalisability of the results and to allow investigation of the impact of the heterogeneous inclusion criteria in this literature on outcomes.

KEY WORDS

Preterm infants; cognition; meta-analysis; very low birth weight

BACKGROUND

Since the 1990s, advances in obstetrical and neonatal care have led to major survival gains very preterm (VPT, <32 weeks of gestation) infants,[1-3] but have also raised concerns about high levels of impairment among survivors. In this context, studies of the association between VPT birth and neurodevelopment, especially cognitive development, have increased in number over the past four decades. More recently, this voluminous literature has been combined in systematic reviews and summarized using meta-analysis.[4-8]

One of the challenges for systematic reviews on VPT and cognition is establishing criteria for study selection given the heterogeneity in gestational age (GA) and birthweight (BW) inclusion criteria in primary studies.[4-9] There has been a longstanding debate about the best way to define populations for research on the sequelae of VPT birth. BW is considered a more reliable measure than GA, but leads to the inclusion of infants born at later GAs with growth pathologies.[6] In contrast, GA is a direct measure of the degree of prematurity and allows investigation of the distinct effects of prematurity and growth restriction, but its measurement may be imprecise when ultrasounds are not routinely used to date pregnancies.[10] This leads to misclassification, which can be particularly acute among disadvantaged groups with sub-optimal antenatal care.[11] This is of concern because disadvantaged families face higher risks of VPT birth[11] and social disadvantage is associated with worse long-term cognitive outcomes.[12]

Systematic reviews have adopted different approaches, with some including all studies based on GA and/or BW, while others have aimed to reduce heterogeneity by selecting only studies that used GA inclusion criteria. In a review of five systematic reviews with meta-analyses on the topic of VPT birth and cognition, we previously showed that investigators'

methodological choices with regard to inclusion and exclusion criteria for primary studies resulted in minimal overlap in the included studies across systematic reviews.[9] While these meta-analyses consistently showed the negative consequences of VPT birth on cognitive development in childhood,[9] they reported substantial unexplained between-study heterogeneity and some conflicting results regarding whether this effect varied with the degree of prematurity.[4-6,8] This inconsistency causes confusion and may represent a missed opportunity to synthesize and analyse all available data.

Therefore, using the primary studies included in these five systematic reviews, our aim is to describe the GA and BW inclusion criteria in studies investigating VPT birth and cognition and to determine whether between-study heterogeneity in meta-analyses can be explained by the GA and BW criteria used in the selected studies.

METHODS

This study is registered with PROSPERO, under the registration number CRD42020176193.

Eligibility criteria, information sources and search strategy

Our eligibility criteria were primary studies included in systematic reviews carrying out meta-analyses based on observational studies investigating general cognitive ability (IQ), regardless of the instrument used, in childhood (<18 years of age) for VPT infants in comparison with a term-born control group. Systematic reviews were considered eligible if other neurodevelopmental outcomes or populations were also investigated. Systematic reviews without meta-analysis were excluded, as well as studies focusing only on other neurodevelopmental outcomes such as executive function. Eligible systematic reviews were

searched in PubMed, Web of Science, the Cochrane Database of Systematic Reviews, and the PROSPERO databases from January 2000 to February 2020 without any language restrictions, as previously described.[9]

Data extraction and study selection

Data were extracted in two stages. First, two researchers (MS and JZ) independently extracted data from eligible systematic reviews, including the study objectives, selection criteria and statistical methods. In a second stage, each primary study included in the selected systematic reviews was independently reviewed by two of the co-authors (MS, ST, VB, JZ, AM) to extract information on country of origin, range of birth years, and study design features (whether the study was population-based, multi-centric or single-centre). Additionally, we extracted selection criteria for VPT participants regarding GA and/or BW (cut-off and combination of criteria); IQ mean scores (unadjusted as well as adjusted for potential confounders, if available), standard deviations and sample sizes were extracted for cases and controls. The reviewers identified which primary studies came from the same cohort (i.e. follow-up studies giving rise to several publications either at the same age point or at multiple age points). They also identified situations where data from two cohorts were presented in the same primary study. Disagreements were resolved by consensus.

For this study, we selected one primary study per cohort. In cases of multiple studies from the same cohort, we selected the study with the longest follow-up but before 18 years of age. This convention was adopted because the stability of cognitive abilities increase with age.[13] If two studies from the same cohort had the same follow-up year, we selected the study with the larger sample size (less selective).

Cohort classification

Cohorts were first classified according to their selection criteria used to define the study population (GA/BW criteria: GA-only, BW-only, GA/BW combined). Then, we derived three subgroups of cohorts reflecting the degree of prematurity using the GA and BW upper thresholds from the WHO definitions[14-16]: EPT (GA<28 weeks' GA) or extremely low birthweight (ELBW, <1000 grams); VPT (<32 weeks) or very low birthweight (VLBW, <1500 grams); and as many of the included studies used 32 or 33 weeks as an upper threshold for inclusions, these studies were retained in a third group classified as moderate preterm (MPT, <34 weeks' GA) or moderate low birthweight (MLBW, <1800 grams). Within each of these three preterm/low birthweight subgroups, studies with higher perinatal risk infants were identified based on GA and BW thresholds (e.g. infants with BW<800 grams or GA<27 weeks were classified as higher risk EPT/ELBW).

Statistical analysis

We estimated the standardized mean difference (SMD) (Hedge's formula) in IQ between VPT and term-born controls for each study, and DerSimonian and Laird random effects models were performed to generate pooled SMDs and their 95% confidence interval (95%CI).[17] Unadjusted scores were used, except for one study which provided scores adjusted for parental education. To investigate whether thresholds and combinations of BW and GA criteria explain heterogeneity in study results, we computed pooled SMDs in sub-group meta-analyses, and we performed meta-regressions based on random-effect models using as covariates GA/BW criteria, degree of prematurity, and level of perinatal risk. The percentage of the between-study variance explained by the meta-regression model (R^2_{meta})

was calculated as a proportion of the between-study variance unexplained by the model (Tau^2, DerSimonian-Laird estimator) related to the total between-study variance.[18] We did not reevaluate the quality of the primary studies using a standardized instrument, as performed in four reviews.[4-6,8] No studies were removed from the reviews based on quality assessments; two reviews[6,8] investigated whether the results were affected by study quality and found no impact on results. The potential for small study-effects was investigated graphically using funnel plot, with asymmetry suggesting a differentiated effect between large and small studies. Analyses were performed using R version 4.0.3 (The R Foundation for Statistical Computing, Vienna, Austria) with the "meta" package version 4.11-0.[19,20]

RESULTS

The search yielded five eligible systematic reviews with meta-analyses of observational studies investigating the association between VPT birth and IQ in childhood. **Table 1** summarizes criteria specified for the target population (i.e. VPT children) for the five reviews. Criteria differed regarding GA boundaries (<32 weeks of gestation or <34 weeks) and consideration of studies using BW limits (<1000 or <1500 grams), which were either explicitly included, excluded or not mentioned.

After removing duplicates, 156 primary studies were identified from these reviews. Of these, 55 were considered as non-eligible and were excluded **(eFigure 1)**. The reasons for exclusion were: age at assessment >17 years (8), no full text (1), no IQ scores provided (3), studies on the same cohort at the same age (23), or a different age (11), late preterm birth (GA<35; GA<36 and BW<2500 grams) (3) and other (7). The analysis sample included 101 primary studies reporting results on 103 unique cohorts of children.

The GA/BW inclusion criteria in these studies were highly varied as shown in **Table 2** with 27 different combinations. The most common were BW <1500 g (24 cohorts), GA <32 weeks (12), GA <33 weeks (12) and BW<1000 grams (12). Few cohorts (2) had GA <28 weeks as an inclusion criterion. Combinations of BW and GA criteria were frequent, but combinations differed. Some combinations were not considered to be consequential and they were combined with other criteria; for instance, BW<1000 grams and GA<34 weeks will not differ much from BW<1000 grams since having a BW<1000 grams and a gestational age of 34 weeks or more is rare. Out of the 103 unique cohorts, 25 were classified as reporting results on children born EPT or ELBW including 5 cohorts targeting a population at higher risk. A further 58 studies reported results on children born VPT or VLBW from which 20 were classified as at higher risk, and 20 studies reported results on children born MPT or MLBW, including 16 studies classified as at higher risk (see **Table 2** for detailed criteria).

The characteristics of cohorts using BW versus GA criteria differed (**Table 3**). Cohorts from North America were more likely to use BW only (63%) to select the study population compared to European cohorts, which were more likely to include infants based on GA only (59%). Cohorts with birth years from 1977 to 1990 were more likely to be based on BW only (67%) compared to more recent cohorts which were more likely to include infants based on GA only.

Figure 1 and eFigure 2 provide results of meta-analyses in sub-groups of cohorts defined by their GA and BW inclusion criteria. Children born EPT or ELBW scored lower on IQ measures compared to term-born children (all GA/BW criteria; SMD=-0.94, 95%CI -1.07, -0.82; equivalent to a deficit of 14.1 IQ points). For children born VPT or VLBW, the pooled SMD varied between -0.73 (95%CI -0.90, -0.55; equivalent to a deficit of 10.9 IQ points) when

computed among the 11 cohorts with combined GA and BW criteria to -0.83 (95%CI -0.93, -0.74; equivalent to a deficit of 12.5 IQ points) among the 28 cohorts using BW criteria only. A smaller effect size was found for the full set of 20 cohorts including children born MPT or MLBW (SMD=-0.68, 95%CI -0.79, -0.57; equivalent to a deficit of 10.2 IQ points).

Table 4 provides the results from random-effect meta-regressions. No clear association was found between the type of GA/BW criteria and the effect size, before (model 1) and after adjustment (models 2 and 3). In contrast, the effect size increased with the degree of prematurity, as well as with the level of perinatal risk (models 2 and 3). The estimated amount of residual heterogeneity between studies (Tau^2) reduced from 0.0486 in model 1 to 0.0380 in model 3, corresponding to less than 3% of between-study variance explained (R^2 meta) in model 1 to 24% in the fully adjusted model 3.

No asymmetry was observed in the funnel plot, suggesting that the results were not distorted by smaller studies (**eFigure 3**).

COMMENT

Principal findings

This study illustrated the wide range of different GA and BW criteria that are used to define longitudinal cohorts for research on the developmental consequences of VPT birth. There were geographical and temporal patterns in inclusion criteria, with BW criteria more commonly used in the North America versus Europe and in older versus more recent studies. In synthesising the results from these studies, we found similar results within preterm/low birthweight subgroups, reflecting the degree of prematurity, regardless of whether GA or BW criteria were used. Taken together, these results provide support for a more inclusive

approach to primary study selection for reviews on cognition and VPT birth, allowing for systematic and transparent reporting of primary study inclusion/exclusion criteria and making it possible to increase the number of studies and generalisability of results.

Strengths of the study

Strengths of our study are inclusion of all primary studies from 5 systematic reviews, with data extraction by 2 reviewers making it possible to reanalyse the data using standardised definitions.

Limitations of the data

Limitations of the study are small sample sizes in some GA/BW combinations which limited sub-group analysis. Because we used data from published meta-analyses, the last primary study included in this analysis is from 2017. Further, the large remaining unexplained between-study variance tends to reduce certainty in the evidence, suggesting the possible effect of other study design characteristics (e.g. type of IQ assessment, study period, analysis of children with severe conditions, follow-up rates) not considered in this study.

Interpretation

The variability in inclusion criteria for the target populations in primary studies of cognitive outcomes after VPT birth represents a methodological challenge for systematic reviews and meta-analyses. Heterogeneity is a problem for combining study results because pooled measures of effect are difficult to interpret in the presence of high between-study heterogeneity.[21-23] The investigation of sources of heterogeneity is crucial,[23,24] but it is often not possible to disentangle effects due to sample characteristics (i.e. clinical heterogeneity) and those attributable to study design and methods (i.e. methodological heterogeneity). Therefore, attempts to include primary studies with similar inclusion criteria in systematic

reviews in order to reduce methodological heterogeneity may be justified. However, in our study, applying more stringent inclusion criteria limited the number of included studies, their geographic and temporal generalizability and the precision of results, without reducing statistical heterogeneity.

Systematic reviews are increasingly undertaken to summarize research results before starting new studies or to underpin recommendations and guidelines and having several systematic reviews on the same topic is a common occurrence. Differences in authors' methodological choices can lead to situations where results diverge or even conflict.[25] All five systematic reviews in our study produced consistent pooled effects showing lower IQ among children born VPT compared to term-born controls but there was variability in findings related to the impact of GA.[4-8] Several found no clear association between GA and IQ.[4-6] While the effect of GA on the cognitive development of children born VPT will be more limited if the included GA range is narrow,[26] this finding does not align with clinical knowledge or results from large population-based cohort studies.[27,28] By combining all available information from the five reviews, we showed a significant gradient related to degree of prematurity/low birthweight, when using grouped categories of GA/BW and when including an additional measure of perinatal risk within categories.

Our results did not reveal a significant impact on effect sizes of using either GA or BW to define the study population when the degree of prematurity/low birthweight was taken into account. While both GA and restricted growth have an impact on cognition, this finding should not be interpreted to mean that their effect on cognition is the same or that there are shared underlying causal mechanisms. Rather it illustrates that, at the study level, there is high overlap between GA and BW groups.[29] This overlap has been the justification for

using BW, considered more reliable, as a proxy for GA. Until the beginning of the 1980s these terms were used interchangeably, although as early as 1961 WHO recommended not using LBW to define prematurity.[14,30,31] However, babies with BW under 1500 grams are principally preterm,[31,32] although they can be moderate or late preterm. Use of BW may also overestimate prematurity in some countries, such as south Asia, where there are a high proportion of small for GA term births.[14] Even though there are differences between the aetiology and consequences of low BW and low GA, a systematic review is not an appropriate study design for distinguishing between these effects, although the use of participant level meta-analyses would make it possible to create sub-groups defined at the individual level and allow investigation of this source of heterogeneity.[33] Initiatives, such as the RECAP Preterm platform (https://recap-preterm.eu/) of European cohorts of children and adults born VPT, provide opportunities for research using similarly defined sub-populations.

The problems for systematic reviews posed by methodological heterogeneity in primary studies are the justification for initiatives such as COMET[34] to define common data sets for clinical trials,[35,36] and several of these have focused on pregnancy outcomes and neonates.[37-39] Similar initiatives would be helpful for establishing guidelines for inclusion criteria for very preterm or very low birthweight longitudinal cohorts. The WHO terminology based on GA[15] may need to be expanded to include higher risk sub-groups used for some studies, such as births <27 weeks, or births close to the limits of viability at <24 weeks (sometimes termed periviable).[36] Furthermore, while GA may be the preferred inclusion criteria, investigators in

low- and middle-income countries may continue to opt for BW criteria because of availability and quality of GA measurement.[14,40]

Our results also suggest that specific guidelines for reviews with meta-analyses of VPT birth are needed. While the five systematic reviews reported on characteristics of primary studies, including mean GA and BW, none provided a comprehensive description of the GA and BW inclusion criteria. One easily applicable recommendation for improving systematic reviews would be to systematically report and analyse this information. This recommendation would have relevance for future reviews of cognition as well as those focusing on other outcomes after preterm birth. The consequences of preterm birth are diverse, affecting multiple facets of cognition, and also motor function, sensory capacities and behaviour and mental health. Adults born preterm continue to experience challenges in the labour market and forming families, and may be more vulnerable to a wide range of non-communicable diseases, notably cardiovascular and psychiatric problems.[41] Birth cohorts are constituted to respond to a range of different research questions and include measures of multiple outcomes. Therefore, our results demonstrating heterogeneity in the inclusion criteria of primary studies likely apply to reviews of other outcomes which have grown in number over past years.[42-45] By transparently reporting and evaluating inclusion criteria in new reviews, it will be possible to assess if the results found for cognition can be transposed to other outcomes.

Conclusions

Reviews to synthesize the literature on VPT populations should consider using broad inclusion criteria based on both GA and BW to encompass studies from all countries, older studies and those from settings where GA data are not available or of poor quality. This approach will make it possible to maximize the number and representativeness of primary

studies included in reviews. However, investigators should abstract and report GA and BW inclusion criteria as well as any other potential effect modifiers and use these in analyses for investigating heterogeneity.

REFERENCES

1. Ancel PY, Goffinet F, Kuhn P, Langer B, Matis J, Hernandorena X, et al. Survival and morbidity of preterm children born at 22 through 34 weeks' gestation in France in 2011: results of the EPIPAGE-2 cohort study. *JAMA Pediatrics.* 2015;169(3):230-238.
2. Costeloe KL, Hennessy EM, Haider S, Stacey F, Marlow N, Draper ES. Short term outcomes after extreme preterm birth in England: comparison of two birth cohorts in 1995 and 2006 (the EPICure studies). *BMJ (Clinical research ed.).* 2012;345:e7976.
3. Fellman V, Hellstrom-Westas L, Norman M, Westgren M, Kallen K, Lagercrantz H, et al. One-year survival of extremely preterm infants after active perinatal care in Sweden. *Jama.* 2009;301(21):2225-2233.
4. Allotey J, Zamora J, Cheong-See F, Kalidindi M, Arroyo-Manzano D, Asztalos E, et al. Cognitive, motor, behavioural and academic performances of children born preterm: a meta-analysis and systematic review involving 64 061 children. *BJOG : an international journal of obstetrics and gynaecology.* 2018;125(1):16-25.
5. Arpi E, D'Amico R, Lucaccioni L, Bedetti L, Berardi A, Ferrari F. Worse global intellectual and worse neuropsychological functioning in preterm-born children at preschool age: a meta-analysis. *Acta Paediatrica.* 2019;108(9):1567-1579.
6. Brydges CR, Landes JK, Reid CL, Campbell C, French N, Anderson M. Cognitive outcomes in children and adolescents born very preterm: a meta-analysis. *Developmental Medicine & Child Neurology.* 2018;60(5):452-468.
7. Kerr-Wilson CO, Mackay DF, Smith GC, Pell JP. Meta-analysis of the association between preterm delivery and intelligence. *Journal of public health.* 2012;34(2):209-216.
8. Twilhaar E, Wade RM, de Kieviet JF, van Goudoever JB, van Elburg RM, Oosterlaan J. Cognitive outcomes of children born extremely or very preterm since the 1990s and associated risk factors: A meta-analysis and meta-regression. *JAMA Pediatrics.* 2018.
9. Sentenac M, Boutron I, Draper ES, Kajantie E, Maier RF, Wolke D, et al. Defining Very Preterm Populations for Systematic Reviews With Meta-analyses. *JAMA Pediatrics.* 2020;174(10):997-999.
10. Hall ES, Folger AT, Kelly EA, Kamath-Rayne BD. Evaluation of gestational age estimate method on the calculation of preterm birth rates. *Maternal and child health journal.* 2014;18(3):755-762.
11. van Oppenraaij RH, Eilers PH, Willemsen SP, van Dunné FM, Exalto N, Steegers EA. Determinants of number-specific recall error of last menstrual period: a retrospective cohort study. *BJOG : an international journal of obstetrics and gynaecology.* 2015;122(6):835-841.
12. Sentenac M, Benhammou V, Aden U, Ancel PY, Bakker LA, Bakoy H, et al. Maternal education and cognitive development in 15 European very preterm birth cohorts from the RECAP Preterm platform. *International journal of epidemiology.* [published online Sept 15, 2021]. doi:10.1093/ije/dyab170.
13. Deary IJ, Whiteman MC, Starr JM, Whalley LJ, Fox HC. The impact of childhood intelligence on later life: following up the Scottish mental surveys of 1932 and 1947. *Journal of personality and social psychology.* 2004;86(1):130-147.
14. Blencowe H, Cousens S, Chou D, Oestergaard M, Say L, Moller AB, et al. Born too soon: the global epidemiology of 15 million preterm births. *Reproductive health.* 2013;10 Suppl 1(Suppl 1):S2.

15. World Health Organization. Preterm birth fact sheet. Assessed: 19 February 2018. https://www.who.int/news-room/fact-sheets/detail/preterm-birth.
16. World Health Organization. ICD-11 for mortality and morbidity statistics (ICD-11 MMS). 2018;https://icd.who.int/browse11/l-m/en -/http%3a%2f%2fid.who.int%2ficd%2fentity%2f1950379483. Accessed 05/2021.
17. Sterne JA, Juni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Statistics in medicine.* 2002;21(11):1513-1524.
18. Aloe AM, Becker BJ, Pigott TD. An alternative to R2 for assessing linear models of effect size. *Research synthesis methods.* 2010;1(3-4):272-283.
19. Balduzzi S, Rücker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. *Evidence-based mental health.* 2019;22:153-160.
20. *R: A language and environment for statistical computing* [computer program]. Vienna: Austria; 2020.
21. Metelli S, Chaimani A. Challenges in meta-analyses with observational studies. *Evidence-based mental health.* 2020;23(2):83-87.
22. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *Jama.* 2000;283(15):2008-2012.
23. Mueller M, D'Addario M, Egger M, Cevallos M, Dekkers O, Mugglin C, et al. Methods to systematically review and meta-analyse observational studies: a systematic scoping review of recommendations. *BMC medical research methodology.* 2018;18(1):44.
24. Thompson SG. Why sources of heterogeneity in meta-analysis should be investigated. *BMJ (Clinical research ed.).* 1994;309(6965):1351-1355.
25. Ioannidis JP. The Mass Production of Redundant, Misleading, and Conflicted Systematic Reviews and Meta-analyses. *The Milbank quarterly.* 2016;94(3):485-514.
26. Linsell L, Malouf R, Morris J, Kurinczuk JJ, Marlow N. Prognostic Factors for Poor Cognitive Development in Children Born Very Preterm or With Very Low Birth Weight: A Systematic Review. *JAMA Pediatrics.* 2015;169(12):1162-1172.
27. Pierrat V, Marchand-Martin L, Marret S, Arnaud C, Benhammou V, Cambonie G, et al. Neurodevelopmental outcomes at age 5 among children born preterm: EPIPAGE-2 cohort study. *BMJ (Clinical research ed.).* 2021;373:n741.
28. Draper ES, Zeitlin J, Manktelow BN, Piedvache A, Cuttini M, Edstedt Bonamy AK, et al. EPICE cohort: two-year neurodevelopmental outcomes after very preterm birth. *Archives of disease in childhood. Fetal and neonatal edition.* 2020;105(4):350-356.
29. Koller-Smith LI, Shah PS, Ye XY, Sjörs G, Wang YA, Chow SSW, et al. Comparing very low birth weight versus very low gestation cohort methods for outcome analysis of high risk preterm infants. *BMC pediatrics.* 2017;17(1):166.
30. Wilcox AJ. On the importance--and the unimportance--of birthweight. *International journal of epidemiology.* 2001;30(6):1233-1241.
31. Quinn JA, Munoz FM, Gonik B, Frau L, Cutland C, Mallett-Moore T, et al. Preterm birth: Case definition & guidelines for data collection, analysis, and presentation of immunisation safety data. *Vaccine.* 2016;34(49):6047-6056.
32. Villar J, Cheikh Ismail L, Victora CG, Ohuma EO, Bertino E, Altman DG, et al. International standards for newborn weight, length, and head circumference by

gestational age and sex: the Newborn Cross-Sectional Study of the INTERGROWTH-21st Project. *Lancet.* 2014;384(9946):857-868.
33. Gagnier JJ, Morgenstern H, Altman DG, Berlin J, Chang S, McCulloch P, et al. Consensus-based recommendations for investigating clinical heterogeneity in systematic reviews. *BMC medical research methodology.* 2013;13:106.
34. Williamson PR, Altman DG, Bagley H, Barnes KL, Blazeby JM, Brookes ST, et al. The COMET Handbook: version 1.0. *Trials.* 2017;18(Suppl 3):280.
35. Williamson PR, Altman DG, Blazeby JM, Clarke M, Devane D, Gargon E, et al. Developing core outcome sets for clinical trials: issues to consider. *Trials.* 2012;13:132.
36. Pavlek LR, Rivera BK, Smith CV, Randle J, Hanlon C, Small K, et al. Eligibility Criteria and Representativeness of Randomized Clinical Trials That Include Infants Born Extremely Premature: A Systematic Review. *The Journal of Pediatrics.* 2021;235:63-74.e12.
37. Khan K, O'Donovan P. The CROWN initiative: journal editors invite researchers to develop core outcomes in women's health. *BMC Pregnancy Childbirth.* 2014;14:199.
38. Devane D, Begley CM, Clarke M, Horey D, C OB. Evaluating maternity care: a core set of outcome measures. *Birth (Berkeley, Calif.).* 2007;34(2):164-172.
39. Kajantie E, Johnson S, Heinonen K, Anderson PJ, Wolke D, Evensen KAI, et al. Common Core Assessments in follow-up studies of adults born preterm-Recommendation of the Adults Born Preterm International Collaboration. *Paediatric and perinatal epidemiology.* 2021;35(3):371-387.
40. Lawn JE, Gravett MG, Nunes TM, Rubens CE, Stanton C, Group GR. Global report on preterm birth and stillbirth (1 of 7): definitions, description of the burden and opportunities to improve data. *BMC pregnancy and childbirth.* 2010;10 Suppl 1(Suppl 1):S1-S1.
41. Wolke D, Johnson S, Mendonça M. The Life Course Consequences of Very Preterm Birth. *Annual Review of Developmental Psychology.* 2019;1(1):69-92.
42. Barre N, Morgan A, Doyle LW, Anderson PJ. Language abilities in children who were very preterm and/or very low birth weight: a meta-analysis. *The Journal of Pediatrics.* 2011;158(5):766-774.e761.
43. Aarnoudse-Moens CS, Weisglas-Kuperus N, van Goudoever JB, Oosterlaan J. Meta-analysis of neurobehavioral outcomes in very preterm and/or very low birth weight children. *Pediatrics.* 2009;124(2):717-728.
44. Pascal A, Govaert P, Oostra A, Naulaers G, Ortibus E, Van den Broeck C. Neurodevelopmental outcome in very preterm and very-low-birthweight infants born over the past decade: a meta-analytic review. *Developmental medicine and child neurology.* 2018.
45. Sømhovd MJ, Hansen BM, Brok J, Esbjørn BH, Greisen G. Anxiety in adolescents born preterm or with very low birthweight: a meta-analysis of case-control studies. *Developmental medicine and child neurology.* 2012;54(11):988-994.

ACKOWLEDGEMENTS

Authors' contributions

Study concept and design: Mariane Sentenac, Anna Chaimani, Jennifer Zeitlin

Data extraction: Mariane Sentenac, Sabrina Twilhaar, Andrei Morgan, Valérie Benhammou, Jennifer Zeitlin

Analysis of data: Mariane Sentenac

Interpretation of data: Mariane Sentenac, Anna Chaimani, Sabrina Twilhaar, Samantha Johnson, Jennifer Zeitlin

Drafting of the manuscript: Mariane Sentenac, Jennifer Zeitlin

Final approval and overall agreement: All authors

FUNDING

This study is part of research on the consequences of very preterm birth for use in developing a research agenda for the RECAP Preterm project, an EU Horizon 2020 (grant agreement No 733280) study to construct a platform combining data from very preterm cohort studies in Europe. The funders had no role in the design and conduct of the study.

TABLES

Table 1- Gestational age and birthweight inclusion criteria used to select primary studies in systematic reviews of very preterm (VPT) birth and IQ in childhood

Systematic review (first author, year)	**Gestational age (GA) criteria**	**Birthweight (BW) criteria**	**Number of studies (VPT children only)**
Kerr-Wilson et al. 2012	Inclusion of studies on all preterm (with sub-group analyses: GA<28 weeks; 28-31 weeks; ≥32 weeks);	Birthweight not mentioned as part of inclusion or exclusion criteria	27 (na)[a]
Allotey et al. 2018	Inclusion of studies on all preterm (with sub-group analyses very: <28 weeks; moderate: 28-33+6 weeks; late: 34-36+6 weeks);	Birthweight not mentioned as part of inclusion or exclusion criteria	57 (50)
Brydges et al. 2018	Inclusion of studies on children with GA <32 weeks;	Exclusion of studies examining exclusively low BW children	44 (44)
Twilhaar et al. 2018	Inclusion of studies on children with GA< 32 weeks	Inclusion of studies with BW<1000g or <1500g	71 (71)
Arpi et al. 2019	Inclusion of studies on children with GA< 32 weeks	Inclusion of studies with BW<1500g	7 (7)

Abbreviations: na=not available; GA=gestational age; BW=birthweight

[a] Studies on VPT not identified in the systematic review

Table 2- Classification of cohorts regarding the study criteria on gestational age (GA) and birthweight (BW)

	GA/BW criteria	**Study criteria (27 different criteria, including 18 denoting higher perinatal risk)**	**Number of cohorts**
Extremely preterm birth (EPT, <28 weeks) or extremely low birthweight (ELBW, <1000 grams)	GA	GA <28 weeks	2
		GA <27 weeks[a]	1
		GA <26 weeks[a]	2
	BW	BW<1000g [and (GA<34 wks)]	12
		BW <800g[a]	2
	GA or BW	GA<28 or BW<1000g	6
Very preterm birth (VPT, <32 weeks) or very low birthweight (VLBW, <1500 grams)	GA	GA<32 weeks	12
		GA <31 weeks[a]	5
		GA <30 weeks[a]	1
		GA<29 weeks[a]	1
	BW	BW<1500g [and GA<36/37]	24
		BW < 1250g[a]	4
	GA or BW	GA<32 weeks or BW<1500g	2
		GA <30 weeks or BW<1000g[a]	1
		GA <30 weeks or BW<1250g[a]	1
	GA and BW	BW < 1500 g and GA<29 weeks[a]	1
		BW < 1500 g and GA<30 weeks[a]	1
		BW < 1500 g and GA<32 weeks[a]	1
		BW < 1500 g and GA<33 weeks[a]	2
		BW < 1500 g and GA<34 weeks[a]	1
		BW < 1500 g and GA<35 weeks[a]	1
Moderate preterm birth (MPT, <34 weeks) or moderate low birthweight (MLBW, <2000 grams)	GA	GA <34 weeks	2
		GA <33 weeks[a]	12
	GA or BW	GA <34 weeks or BW<1800g	1
		GA <33 weeks or BW<1500g[a]	3
	GA and BW	GA <34 weeks and BW<1800g	1
		GA <37 weeks and BW<1600g[a]	1

[a] criteria classified as targeting a population of children with higher perinatal risk

Table 3 – Characteristics of cohorts by their use of gestational age (GA) or birthweight (BW) inclusion criteria

Cohort characteristics	Number of cohorts	Combination of the gestational age (GA) and birthweight (BW) criteria			
		GA only n=38	BW only n=42	GA or BW n=14	GA and BW n=9
Geographical area, n (%)					
North America	32	5 (16%)	20 (63%)	2 (6%)	5 (16%)
Europe	49	29 (59%)	12 (24%)	6 (12%)	2 (4%)
Other	22	4 (18%)	10 (45%)	6 (27%)	2 (9%)
Birth year, n (%)					
<=1990	27	5 (19%)	18 (67%)	2 (7%)	2 (7%)
1991-2000	38	18 (47%)	10 (26%)	7 (18%)	3 (8%)
2001-2010	26	12 (46%)	8 (31%)	4 (15%)	2 (8%)
Study design, n (%)					
Single centre	53	21 (40%)	18 (34%)	8 (15%)	9 (11%)
Multi centre	17	4 (24%)	11 (65%)	0	2 (12%)
Population based	30	12 (40%)	12 (40%)	5 (17%)	1 (3%)
Unknown	3	1 (33%)	1 (33%)	1 (33%)	0

Note: Chi2 tests were performed to evaluate differences among groups yielding p-values <0.001 (geographical area), =0.051 (birth year), or =0.404 (study design)

Table 4 – Results from random-effect meta-regressions of the association between GA/BW criteria and SMDs in IQ between EPT/VPT and full term children

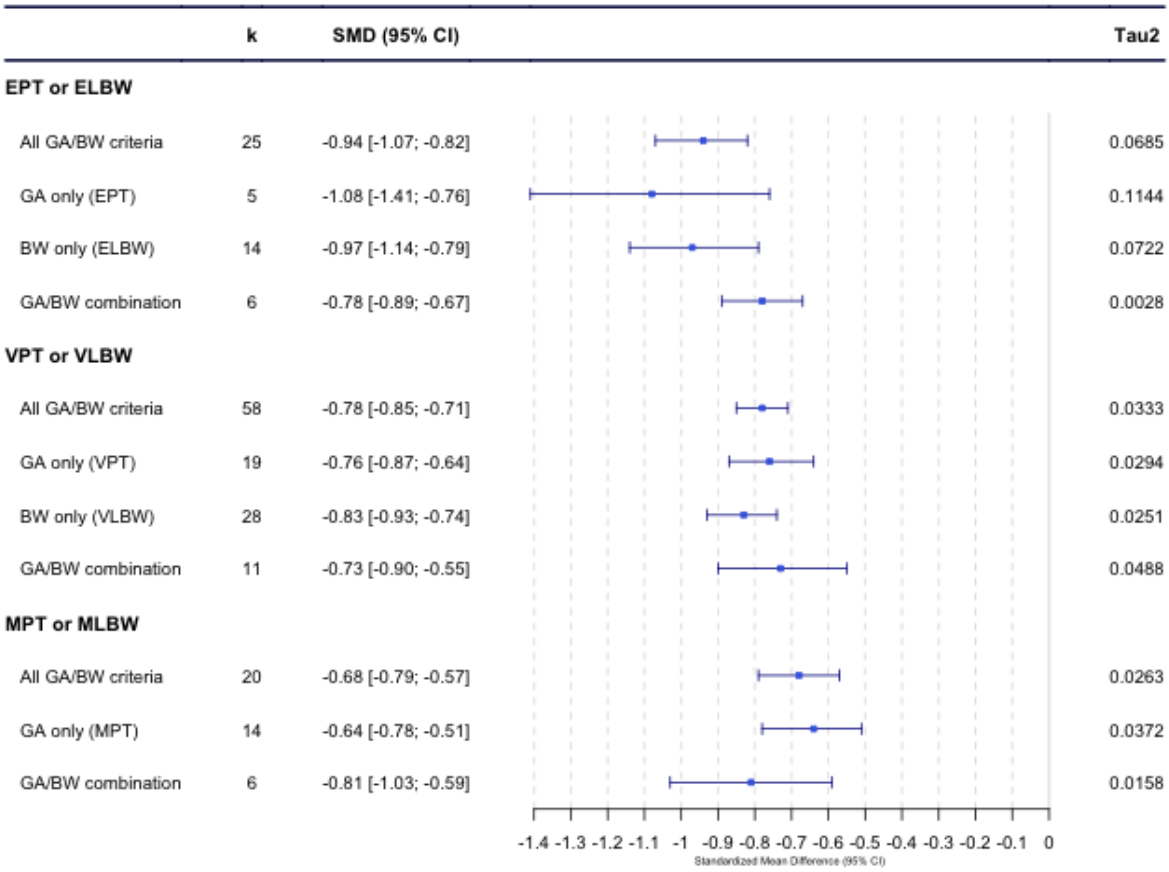
		Model 1[a]		Model 2[a]		Model 3[a]	
	k	**coef.**	**95%CI**	**coef.**	**95%CI**	**coef.**	**95%CI**
GA/BW criteria							
GA only	38	ref		ref		ref	
BW only	42	-0.10	(-0.23, -0.03)	-0.04	(-0.18, 0.10)	-0.11	(-0.25, 0.03)
GA or BW	14	-0.03	(-0.20, 0.14)	0.03	(-0.14, 0.20)	-0.01	(-0.18, 0.16)
GA and BW	9	0.11	(-0.12, 0.34)	0.10	(-0.13, 0.33)	0.17	(-0.05, 0.40)
Degree of prematurity							
EPT or ELBW	25	-		ref		ref	
VPT or VLBW	58	-		0.15	(0.01, 0.28)	0.16	(0.03, 0.29)
MPT or MLBW	20	-		0.23	(0.05, 0.41)	0.30	(0.12, 0.49)
Level of perinatal risk							
Studies without higher risk criteria	62	-		-		ref	
Studies with higher risk criteria	41	-		-		-0.19	(-0.32, -0.06)
R^2_{meta} (percentage of between-study variance explained by the model)		2.47%		11.59%		23.77%	
Tau^2 (residual between-study variance)[b]		0.0486		0.0440		0.0380	

[a] Model 1 adjusted for the type of GA/BW criteria, model 2 adjusted for the type of GA/BW criteria and the degree of prematurity, and model 3 adjusted for the type of GA/BW criteria, the degree of prematurity and the level of perinatal risk.

[b] Total between-study variance estimated from model without covariate = 0.0498

FIGURE LEGENDS

Figure 1 – Forest plot of standardized mean differences (SMDs) by degree of prematurity and by inclusion criteria for GA and BW



Abbreviations: k=number of cohorts; SMD= Standardized mean difference; EPT=extremely preterm; ELBW=Extremely low birthweight; VPT=very preterm; VLBW=very low birthweight; MPT=moderate preterm; MLBW=moderate low birthweight; GA=gestational age; BW=birthweight

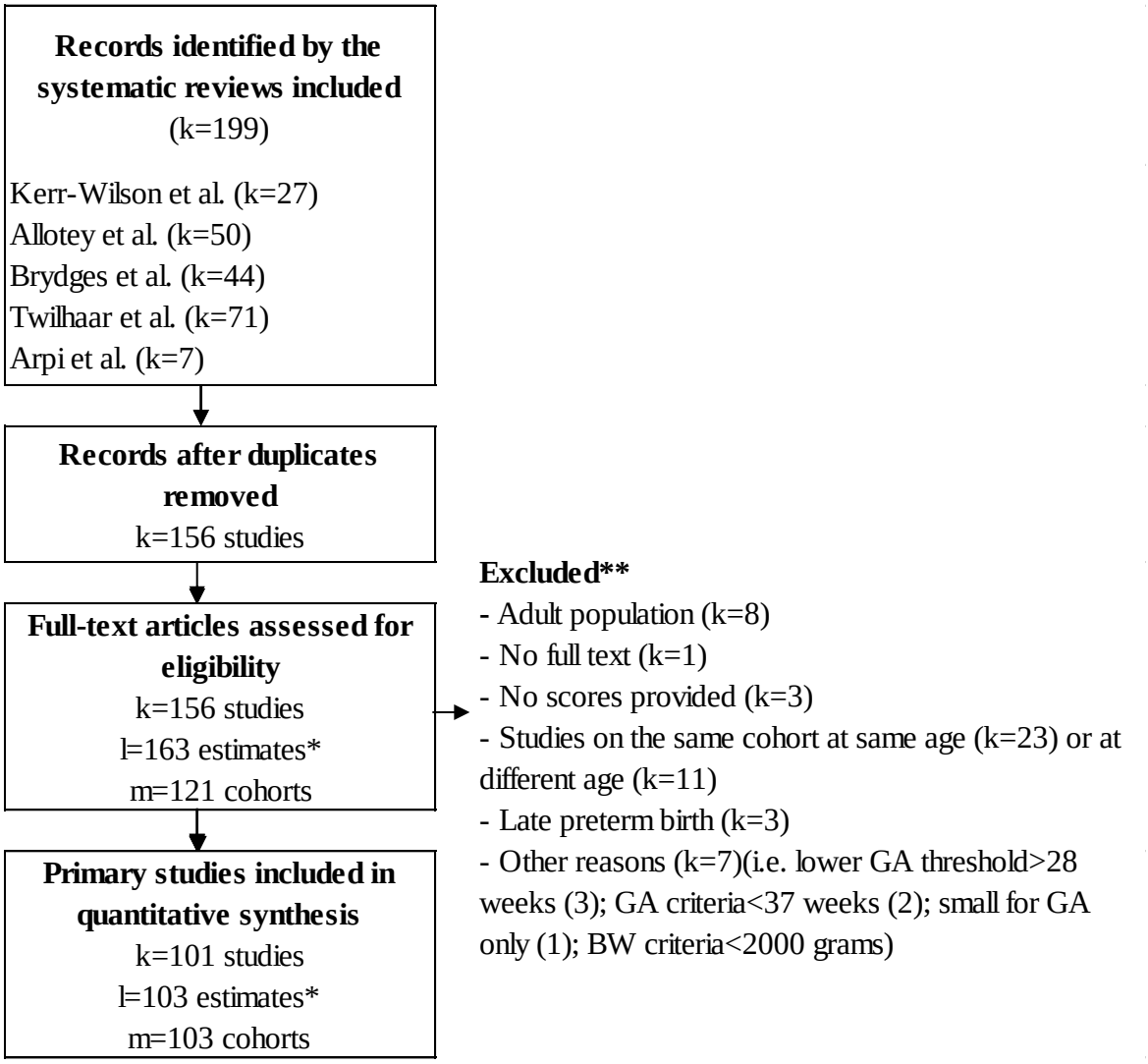
SUPPLEMENTAL TABLES AND FIGURES

eFigure 1 : Flowchart

eFigure 2 - Forest plot of standardized mean differences (SMDs) in IQ by degree of prematurity subgroups (103 cohorts)

eFigure 3 : Funnel plot of standardized mean differences (SMDs) in IQ (103 cohorts)

eFigure 1 : Flowchart



* results from several cohorts reported by paper

** One study shared two exclusion criteria

References

Allotey J, Zamora J, Cheong-See F, et al. Cognitive, motor, behavioural and academic performances of children born preterm: a meta-analysis and systematic review involving 64 061 children. BJOG : an international journal of obstetrics and gynaecology. 2018;125(1):16-25.

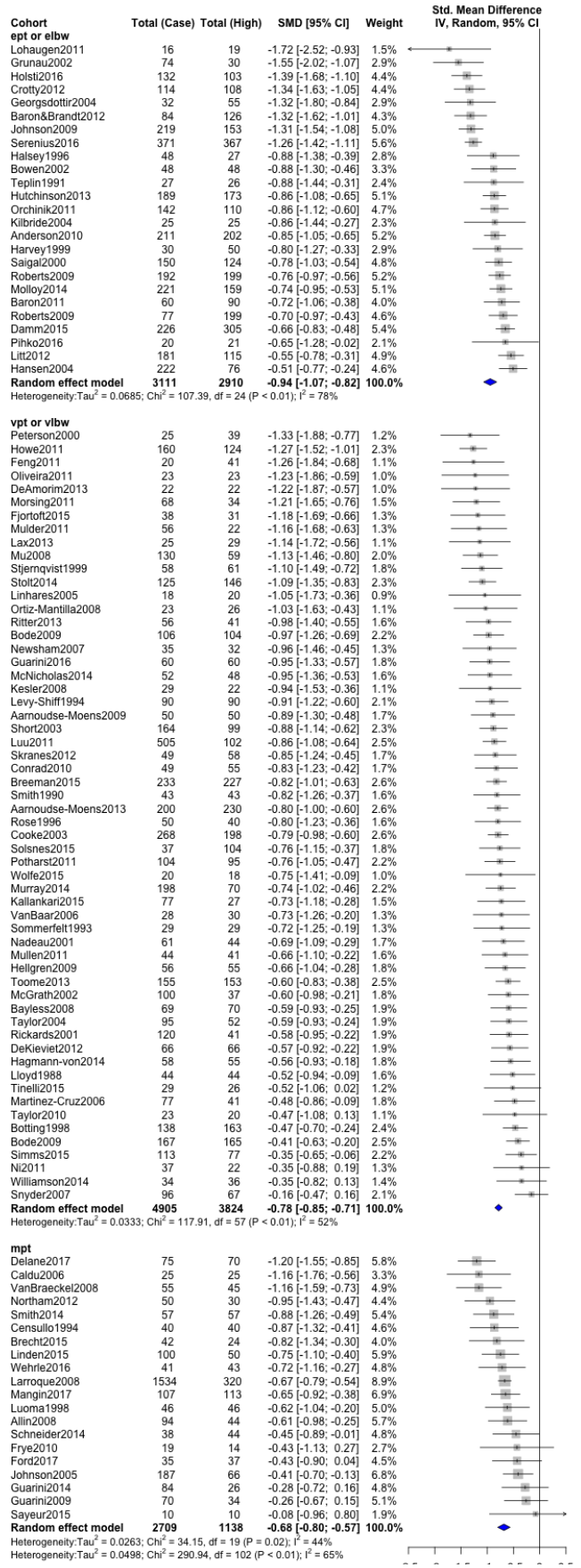
Arpi E, D'Amico R, Lucaccioni L, Bedetti L, Berardi A, Ferrari F. Worse global intellectual and worse neuropsychological functioning in preterm-born children at preschool age: a meta-analysis. Acta Paediatrica. 2019;108(9):1567-1579.

Brydges CR, Landes JK, Reid CL, Campbell C, French N, Anderson M. Cognitive outcomes in children and adolescents born very preterm: a meta-analysis. Developmental Medicine & Child Neurology. 2018;60(5):452-468.

Kerr-Wilson CO, Mackay DF, Smith GC, Pell JP. Meta-analysis of the association between preterm delivery and intelligence. Journal of public health. 2012;34(2):209-216.

Twilhaar E, Wade RM, de Kieviet JF, van Goudoever JB, van Elburg RM, Oosterlaan J. Cognitive outcomes of children born extremely or very preterm since the 1990s and associated risk factors: A meta-analysis and meta-regression. JAMA Pediatrics. 2018.

eFigure 2 - Forest plot of standardized mean differences (SMDs) in IQ by degree of prematurity subgroups (103 cohorts)



eFigure 3 : Funnel plot of standardized mean differences (SMDs) in IQ (103 cohorts)