REVIEW

Paediatric and Perinatal Epidemiology · WILEY

Heterogeneity of design features in studies included in systematic reviews with meta-analysis of cognitive outcomes in children born very preterm

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

[1]Centre of Research in Epidemiology and Statistic (CRESS), Inserm, INRAE, Université Paris Cité, Paris, France

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

[3]Department of Health Sciences, University of Leicester, Leicester, UK

Correspondence
Mariane Sentenac, Inserm UMR 1153, Obstetrical, Perinatal and Pediatric Epidemiology Research Team (EPOPé), Paris, France.
Email: mariane.sentenac@inserm.fr

[Correction added on 18 February, after first online publication: The figure on page 3 has been removed in this version.]

Abstract

Background: Meta-analyses of the voluminous scientific literature on the impact of very preterm (VPT, <32 weeks' gestation) birth on cognition find a marked deficit in intelligence quotient (IQ) among children born VPT relative to term-born peers, but with unexplained between-study heterogeneity in effect size.

Objectives: To conduct an umbrella review to describe the design and methodology of primary studies and to assess whether methodological heterogeneity affects the results of meta-analyses.

Data Sources: Primary studies from five systematic reviews with meta-analysis on VPT birth and childhood IQ.

Study Selection and Data Extraction: Information on study design, sample characteristics and results was extracted from studies. Study features covered study type, sample size, follow-up rates, adjustment for social context, management of severe impairments and test type.

Synthesis: We used random-effects subgroup meta-analyses and meta-regressions to investigate the contribution of study features to between-study variance in standardised mean differences (SMD) in IQ between groups.

Results: In 58 cohorts (56%), children with severe impairments were excluded, while 23 (22%) cohorts accounted for social factors. The least reported feature was the follow-up rate (missing in 38 cohorts). The largest difference in SMDs was between studies using full scale IQ tests (61 cohorts, SMD −0.89, 95% CI −0.96, −0.82) versus short-form tests (27 cohorts, SMD −0.68, 95% CI −0.79, −0.57). The proportion of between-study variance explained by the type of test was 14%; the other features explained less than 1% of the variance.

Conclusions: Study design and methodology varied across studies, but most of them did not affect the variance in effect size, except the type of cognitive test. Key features, such as the follow-up rate, were not consistently reported limiting the evaluation of

their potential contribution. Incomplete reporting limited the evaluation of the full impact of this methodological diversity.

KEYWORDS
cognition, heterogeneity, meta-analysis, preterm infants, very low birthweight

1 | BACKGROUND

Over past decades, the increasing survival of infants born very pre-term (VPT, <32 weeks' gestation) propelled research into the long-term health and neurodevelopmental consequences of VPT birth. More recently, researchers have synthesised this literature, especially studies focusing on cognitive development, in systematic reviews and meta-analyses.[1–5] These meta-analyses have documented a marked negative association of VPT birth with cognitive abilities, reporting standardised mean difference (SMD) larger than −0.75 in intelligence quotient (IQ) (corresponding to a deficit of at least 11 IQ points) between children and adolescents born VPT and term-born controls.[6] However, I^2 values of at least 60% in these meta-analyses[6] raised questions about between-study heterogeneity, the sources of this heterogeneity and the relevance of the pooled results for all VPT populations.

Further sub-group and meta-regression analyses to explore this heterogeneity found that sample characteristics such as age at assessment,[1–3] parental education[3] and birth year[1,3] were not associated with effect sizes and findings were inconsistent for gestational age (GA).[1,4,7,8] Some of these null findings, especially related to parental education and GA, go against well-documented associations in the literature. Possible explanations for these results might be other sources of heterogeneity linked to study design features, which could obscure these associations or that the use of individual characteristics averaged across studies (e.g. mean GA) creates an ecological fallacy.[9] In a previous study, we found support for the first hypothesis: The GA and birthweight (BW) criteria used to define VPT populations were important sources of heterogeneity across studies included in these meta-analyses, but were not described fully. By taking this diversity into consideration and using more uniform groups describing the children's risk status, the degree of preterm birth was strongly related to effect sizes and explained up to 24% of between-study heterogeneity.[10]

The purpose of this study is to investigate whether design and methodological characteristics, beyond the criteria used to define VPT populations, contribute to the heterogeneity in the results of meta-analyses of this research.

2 | METHODS

This study is part of a larger project on information synthesis of studies of VPT birth and cognition. We carried out an umbrella review by searching for systematic reviews with meta-analyses of cognitive

Synopsis

Study Question

Do the methodological characteristics of primary studies on very preterm birth and cognition affect the results of meta-analyses?

What's Already Known

Differences in the methodological characteristics of studies on very preterm birth and childhood cognitive function, such as exclusions of severe impairments and follow-up rates, can affect results, but are rarely considered in meta-analyses.

What this Study Adds

This study illustrated wide variability of methodological characteristics in studies on very preterm birth and cognition, although many design features were poorly reported. However, available characteristics only explained a small part of between-study heterogeneity in results, with the exception of the use of long- versus short-form cognitive assessments.

outcomes in VPT children and then pooling the primary studies included in these reviews and conducting new meta-analyses. The methods have been described previously.[6,10]

2.1 | Search strategy and eligibility criteria

We first searched PubMed, Web of Science, the Cochrane Database of Systematic Reviews and the PROSPERO databases from January 2000 to February 2020 for systematic reviews with meta-analysis of observational studies investigating general cognition (intelligence quotient; IQ) in children and adolescents (<18 years) born VPT compared to term-born controls (the search strategy including search terms was published elsewhere[6]). We excluded systematic reviews that investigated only subdomains of cognition, that did not include studies with control groups or that did not provide quantitative synthesis of study results for

VPT children. We then compiled all primary studies included by the selected reviews. We excluded studies with age at assessment >17 years, no full text, no IQ scores, no clear GA-BW criteria or that included late preterm births (34–36 weeks GA), which were not distinguished from VPT groups. We did not apply any language restrictions in our strategy.

2.2 | Study selection and data extraction

Data were extracted in two successive stages. First, two researchers (MS and JZ) extracted data from the systematic reviews.[6,10] In a second stage, primary studies were independently reviewed by two co-authors (MS, ST, VB, AM and JZ) to extract the following information: country of origin, range of birth years, age at assessment, type of study design, eligibility criteria for VPT regarding GA and/or BW (cut-off and combination of criteria), selection and follow-up of participants, outcome measurement, statistical methods (i.e. imputation methods), sample characteristics (i.e. clinical conditions or disabilities, multiple birth, socio-economic characteristics) and results (sample sizes, mean IQ scores with SD for both groups). A pilot extraction form was developed and tested on a set of studies; disagreements were resolved by consensus.

When there was more than one study on the same cohort (i.e. with the same country, birth year(s) and research group), we selected the study with the longest follow-up. When two cohorts were included in the same study (N = 2), they were considered as two separate observations since some design features were cohort specific (i.e. control group) and results were reported separately. The final sample included studies reporting results from 103 unique cohorts of children.

The methodology of the systematic reviews was described with the domains through which bias may be introduced, as described in the ROBIS tool.[11] Quality assessment of the primary studies was not undertaken as this overlapped with our main study objective of evaluating the contribution of methodological characteristics to the results of meta-analysis. Further, no study was removed from the original systematic reviews based on quality assessments.

2.3 | Indicators of study features

We derived eight indicators describing study design features: (1) type of recruitment (single-centre, multi-centre or population-based study); (2) sample size, which refers to the total sample including the preterm and comparison groups (<100, ≥100); (3) VPT follow-up rate, computed by dividing the number of participants, with or without outcome assessments, by the number of survivors when the children were recruited for follow-up, if available, or at hospital discharge (≥80%, 50%–79%, <50%); (4) type of cognitive measurement (general intelligence based on full scale IQ, general intelligence based on short-form IQ, developmental abilities other

TABLE 1 Classification of cognitive ability tests.

Type of cognitive test	Test names/Score	Number of cohorts
General intelligence: full-scale	British Ability Scales/Differential Ability Scales	5
	Stanford–Binet Intelligence Scales – Composite standard score IQ	4
	Woodcock-Johnson III Tests – General Intellectual Ability (GIA)	1
	KABC – Mental Process Composite (MPC)	4
	WISC, WPPSI – Full scale IQ	44
	McCarthy Scales of Children's Abilities – General Cognitive Index (GCI)	3
General intelligence: short form	British Ability Scales/Differential Ability Scales – Estimated IQ	1
	Stanford-Binet Intelligence Scales – Composite IQ score	2
	KABC, KBIT – Composite IQ	4
	WASI, WISC, WPPSI – Composite IQ	19
	Revised Amsterdam Children's Intelligence Test – IQ	1
Developmental abilities	BSID, Bayley – Mental Development Index (MDI) or cognitive composite score	5
Others	Wechsler (WASI, WISC, WPPSI) – Performance IQ	1
	Raven's Coloured Progressive Matrices – Non-verbal cognition	3
	Peabody Picture Vocabulary Test – Verbal intelligence	3
	WASI – Verbal and non-verbal intelligence	1
	Combination of several tests	2

Note: The classification was supported by the Cattell–Horn–Carroll theory of cognitive abilities.

Abbreviations: BSID, Bayley Scales of Infant Development; IQ, Intelligence quotient; KABC, Kaufman Assessment Battery for Children; KBIT, Kaufman Brief Intelligence Test; WASI, Wechsler Abbreviated Scale of Intelligence; WISC, Wechsler Intelligence Scale for Children; WPPSI, Wechsler Preschool and Primary Scale of Intelligence.

than general intelligence, other measurements including cognitive subscales or different scales used between VPT and controls; Table 1); (5) management of children with severe conditions/impairment among the VPT group based on data provided in each study on exclusion criteria and/or the characteristics of the sample (children with severe impairments excluded, some exclusion criteria related to impairments, not excluded but untestable children because of severe developmental delay excluded from analyses, not excluded and scores imputed for untestable children); (6) inclusion of multiple births (excluded, not excluded); (7) management of socio-economic characteristics in both groups (matched design without differences reported between groups, independent sampling without differences, differences reported between groups without adjustment for socio-economic characteristics and differences reported between groups and adjustment); and (8) level of risk of the control group (high risk: GA <37 weeks or admitted to a neonatal unit, low risk: GA ≥37 weeks and/or no severe condition). We also included a previously constructed indicator assessing the degree of preterm birth (GA < 28 weeks or BW < 1000 g, GA < 32 weeks or BW < 1500 g, GA < 34 weeks or BW < 1800 g).[10] These indicators were derived independently by each abstractor from the information in the studies; disagreements in classifications were resolved by consensus.

2.4 | Statistical analysis

The standardised mean difference (SMD, Hedge's formula) in IQ between VPT participants and term-born controls was computed for each study. We described the distribution of study characteristics and performed subgroup random-effect meta-analyses to generate pooled SMDs with 95% confidence intervals (95% CI) by these characteristics. The conservative Hartung–Knapp–Sidik–Jonkman method was applied to take into account the small number of studies in some groups.[12,13] The I^2 statistic was computed over the set of studies for comparison with the previous meta-analyses.[14] However, for our main analyses, we computed the between-study variance (τ^2), derived using the restricted maximum likelihood method[12] with 95% prediction intervals. This measure illustrates the predicted range of the true effect size expected for 95% of similar future studies and is more suited than the I^2 to our analysis which seeks to explain differences in the effect size.[15] Random-effects meta-regressions, without and with adjustment for the degree of preterm birth, with a restricted maximum-likelihood estimator[16] were conducted to explore the possible moderating role of study features for between-study variance in effect sizes. The proportion of variance explained (R^2_{meta}) by each study characteristic was derived from the total between-study heterogeneity (τ^2_{tot}) and the residual between-study heterogeneity (τ^2_{res}) unexplained by study characteristics. As a sensitivity analysis, an additional meta-regression model was run for the follow-up rate with adjustment for the age at assessment to take into account potential confounding due to higher expected attrition when follow-up periods

are longer. Analyses were performed using R version 4.1.2 (The R Foundation for Statistical Computing) with the 'meta' package version 4.11–0.19,20 and the 'rmeta' package version 3.0.

2.5 | Ethics approval

The study protocol was registered on PROSPERO (CRD42020176193).

3 | RESULTS

The systematic literature search identified five reviews with meta-analyses that shared the common objective of investigating differences in IQ between children born VPT and term-born controls (Figure 1). A description of study eligibility criteria, main methods of study identification and selection and risk of bias assessment carried out by the included reviews is provided in Table S1. These reviews included 199 primary studies of which 156 were eligible for our study. After removing studies reporting on the same cohort of children, by selecting the study with the longest follow-up, the sample included was 101 studies reporting on 103 unique cohorts of children.

Fewer than one-third (29%) of the studies were representative of a geographical population, with a majority (60%) having a sample of 100 children or more (Figure 2). The rate of follow-up could not be computed in 38 studies (37%) because data were not reported; one-quarter had a follow-up rate of 80% or higher, while 15 studies (15%) had a rate below 50%. The cognitive test was performed using full (59%) or short (24%) versions of an intelligence test. Children with severe conditions/impairment were excluded in 58 studies (56%), and among the other 39 studies, 16 reported that untestable children were assigned a score (usually the lowest score possible). Most studies did not provide information on inclusion or proportion of multiple births (61%), and information on socioeconomic characteristics was lacking in 12% of studies. In 54 cohorts, social characteristics (e.g. parental education, occupational status and social class) were balanced between VPT and term groups, whereas one reported IQ scores after adjustment for social factors and four others provided estimates of the association between VPT and IQ adjusted for social characteristics (Table 2).

Overall the 103 cohorts, children born VPT scored lower on IQ measures compared to term-born children (SMD −0.81, 95% CI −0.87, −0.75; equivalent to 12.2 IQ points) with $I^2 = 64.9\%$. There was no strong evidence of effect size differences according to the study design features, except for the type of cognitive assessment. The forest plot (Figure 2) showed a larger pooled SMD for the 61 studies using full scale IQ (SMD −0.89, 95% CI −0.96, −0.82; equivalent to a deficit of 13.4 IQ points) compared to the 27 studies based on a short-form IQ test (SMD −0.68, 95% CI −0.79, −0.57; 10.2 IQ points). According to the meta-regression with adjustment for degree of preterm birth (Table 2), the pooled SMD for studies using a full-scale IQ test was 0.17 (95% CI 0.03, 0.30), equivalent to 2.6
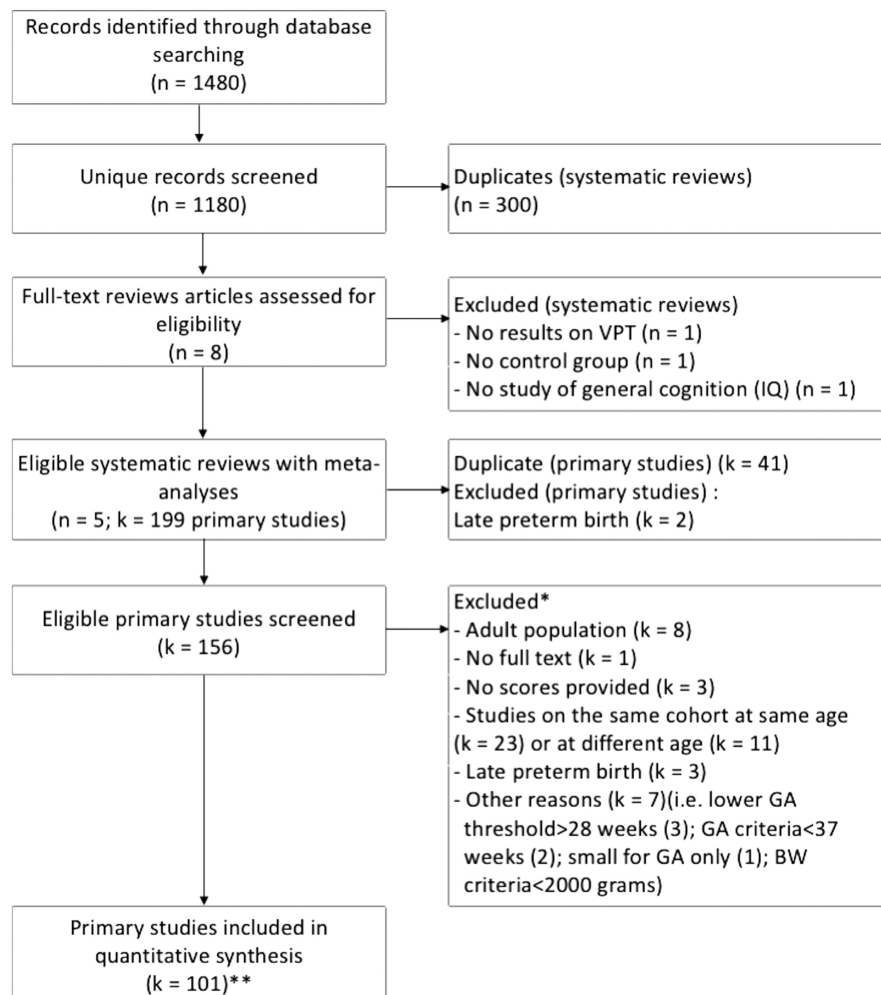
FIGURE 1 Flowchart of the selection of systematic reviews and primary studies investigating VPT birth and cognition in childhood.

IQ points, higher than for studies using a short version. Study characteristics explained a low proportion (<1%) of the between-study variance in effect sizes, with the exception of the type of test (14%) (Table 2). Sensitivity analyses adjusting for age at follow-up confirmed the stability of our findings for the attrition rate.

4 | COMMENT

4.1 | Principal findings

This study reveals substantial methodological diversity in studies investigating the consequences of VPT birth on cognitive outcomes in childhood and adolescence with respect to definitions of the eligible study population (e.g. severity of impairment), follow-up rates, the tests and procedures for the measurement of cognitive abilities and adjustment for socioeconomic factors. However, these study features did not explain between-study heterogeneity in outcomes with the exception of the type of cognitive test (14%). A larger difference in SMDs was found for studies using full scale IQ compared to studies based on a short-form IQ test. Some analyses were limited

by poor reporting, notably of the rate of follow-up, missing in 37% of studies and of adjustment for social characteristics, missing in 12% of studies.

4.2 | Strengths of the study

The strengths of this study comprise the identification and inclusion of all studies used in previous meta-analyses of IQ for children born VPT compared to term-born controls and the extraction by two independent reviewers of study design features.

4.3 | Limitations of the data

Limitations result from the heterogeneity across multiple measures and poor reporting of key items, which reduces the study's power to detect relationships between individual methodological characteristics and effect sizes. For instance, we had to rely on heterogeneous information provided in the studies for our classification concerning the management of severe conditions. We did not
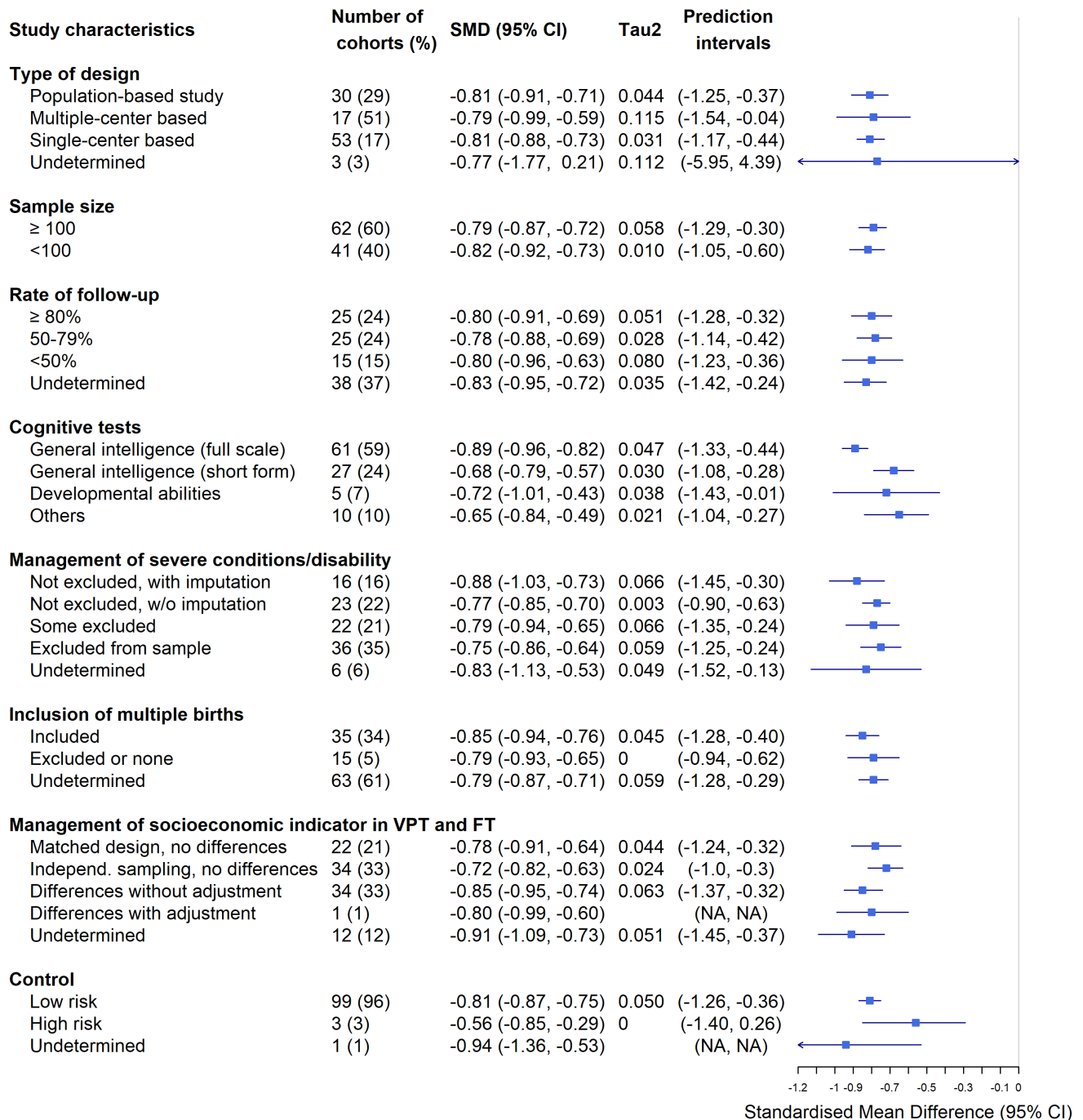
Study characteristics	Number of cohorts (%)	SMD (95% CI)	Tau2	Prediction intervals
Type of design				
Population-based study	30 (29)	-0.81 (-0.91, -0.71)	0.044	(-1.25, -0.37)
Multiple-center based	17 (51)	-0.79 (-0.99, -0.59)	0.115	(-1.54, -0.04)
Single-center based	53 (17)	-0.81 (-0.88, -0.73)	0.031	(-1.17, -0.44)
Undetermined	3 (3)	-0.77 (-1.77, 0.21)	0.112	(-5.95, 4.39)
Sample size				
≥ 100	62 (60)	-0.79 (-0.87, -0.72)	0.058	(-1.29, -0.30)
<100	41 (40)	-0.82 (-0.92, -0.73)	0.010	(-1.05, -0.60)
Rate of follow-up				
≥ 80%	25 (24)	-0.80 (-0.91, -0.69)	0.051	(-1.28, -0.32)
50-79%	25 (24)	-0.78 (-0.88, -0.69)	0.028	(-1.14, -0.42)
<50%	15 (15)	-0.80 (-0.96, -0.63)	0.080	(-1.23, -0.36)
Undetermined	38 (37)	-0.83 (-0.95, -0.72)	0.035	(-1.42, -0.24)
Cognitive tests				
General intelligence (full scale)	61 (59)	-0.89 (-0.96, -0.82)	0.047	(-1.33, -0.44)
General intelligence (short form)	27 (24)	-0.68 (-0.79, -0.57)	0.030	(-1.08, -0.28)
Developmental abilities	5 (7)	-0.72 (-1.01, -0.43)	0.038	(-1.43, -0.01)
Others	10 (10)	-0.65 (-0.84, -0.49)	0.021	(-1.04, -0.27)
Management of severe conditions/disability				
Not excluded, with imputation	16 (16)	-0.88 (-1.03, -0.73)	0.066	(-1.45, -0.30)
Not excluded, w/o imputation	23 (22)	-0.77 (-0.85, -0.70)	0.003	(-0.90, -0.63)
Some excluded	22 (21)	-0.79 (-0.94, -0.65)	0.066	(-1.35, -0.24)
Excluded from sample	36 (35)	-0.75 (-0.86, -0.64)	0.059	(-1.25, -0.24)
Undetermined	6 (6)	-0.83 (-1.13, -0.53)	0.049	(-1.52, -0.13)
Inclusion of multiple births				
Included	35 (34)	-0.85 (-0.94, -0.76)	0.045	(-1.28, -0.40)
Excluded or none	15 (5)	-0.79 (-0.93, -0.65)	0	(-0.94, -0.62)
Undetermined	63 (61)	-0.79 (-0.87, -0.71)	0.059	(-1.28, -0.29)
Management of socioeconomic indicator in VPT and FT				
Matched design, no differences	22 (21)	-0.78 (-0.91, -0.64)	0.044	(-1.24, -0.32)
Independ. sampling, no differences	34 (33)	-0.72 (-0.82, -0.63)	0.024	(-1.0, -0.3)
Differences without adjustment	34 (33)	-0.85 (-0.95, -0.74)	0.063	(-1.37, -0.32)
Differences with adjustment	1 (1)	-0.80 (-0.99, -0.60)		(NA, NA)
Undetermined	12 (12)	-0.91 (-1.09, -0.73)	0.051	(-1.45, -0.37)
Control				
Low risk	99 (96)	-0.81 (-0.87, -0.75)	0.050	(-1.26, -0.36)
High risk	3 (3)	-0.56 (-0.85, -0.29)	0	(-1.40, 0.26)
Undetermined	1 (1)	-0.94 (-1.36, -0.53)		(NA, NA)

Standardised Mean Difference (95% CI)

FIGURE 2 Design and methodology of 101 primary studies included in five meta-analyses of cognition and very preterm birth and standardised mean differences (SMD) in IQ between cases and controls by these characteristics. Abbreviations: FT, full term; NA, not available; SMD, standardised mean difference; VPT, very preterm.

perform formal assessment of risk of bias in primary studies using scales and quality scores, as this issue was addressed by our main study objective. Further, such scales, applied to observational studies, may introduce a certain degree of subjectivity and add heterogeneity.[17,18] We also wished to represent the full set of studies used in the five published meta-analyses and none of these systematic reviews, despite carrying out formal quality analyses, most often using the Newcastle Ottawa Scale,[19] removed studies from their final models.

4.4 | Interpretation

Systematic reviews and meta-analyses of observational studies, as opposed to randomised trials, face specific methodological challenges related to the well-known limitations of observational designs (e.g. selection of participants and confounding), leading to high levels of heterogeneity.[17,20] Most design features in our study did not measurably affect the heterogeneity in results. We did not reveal meaningful differences in SMD related to management of

TABLE 2 Proportion of between-study heterogeneity explained by study characteristics estimated from random-effects meta-regressions.

Study characteristics	Number of cohorts	Change in standardised mean difference in IQ (95% CI)				
		Models 1[a]	$\tau^2_{residual}$	Models 2[b]	$\tau^2_{residual}$	R^2_{meta} [c]
Type of design						
Population-based study	30	1.00 (Reference)	0.052	1.00 (Reference)	0.046	0%
Multiple-centre based	17	0.02 (−0.16, 0.19)		−0.03 (−0.21, 0.15)		
Single-centre based	53	0.00 (−0.13, 0.13)		−0.06 (−0.19, 0.07)		
Undetermined	3	−0.07 (−0.29, 0.44)		−0.08 (−0.45, 0.29)		
Sample size						
≥100	62	1.00 (Reference)	0.049	1.00 (Reference)	0.044	0%
<100	41	−0.03 (−0.16, 0.09)		−0.06 (−0.19, 0.06)		
Rate of follow-up						
≥80%	25	1.00 (Reference)	0.051	1.00 (Reference)	0.044	0%
50–79%	25	0.02 (−0.14, 0.17)		−0.02 (−0.17, 0.13)		
<50%	15	−0.01 (−0.20, 0.18)		−0.08 (−0.27, 0.11)		
Undetermined	38	−0.04 (−0.19, 0.12)		−0.12 (−0.27, 0.03)		
Cognitive tests						
General intelligence (full scale)	61	1.00 (Reference)	0.042	1.00 (Reference)	0.040	14.0%
General intelligence (short form)	25	0.21 (0.08, 0.34)		0.17 (0.03, 0.30)		
Developmental abilities	5	0.17 (−0.06, 0.39)		0.15 (−0.07, 0.38)		
Others	12	0.21 (−0.02, 0.40)		0.18 (−0.01, 0.37)		
Management of severe conditions/disability						
No exclusion with imputation	16	1.00 (Reference)	0.050	1.00 (Reference)	0.046	0%
No exclusion, no imputation	23	0.06 (−0.12, 0.23)		0.01 (−0.17, 0.18)		
Some excluded	22	0.08 (−0.10, 0.26)		−0.02 (−0.21, 0.18)		
All excluded	36	0.13 (−0.03, 0.29)		0.03 (−0.15, 0.20)		
Undetermined	6	0.05 (−0.22, 0.32)		−0.08 (−0.36, 0.19)		
Inclusion of multiple births						
Included	35	1.00 (Reference)	0.050	1.00 (Reference)	0.045	0%
Excluded or none	5	0.07 (−0.21, 0.34)		0.06 (−0.21, 0.32)		
Undetermined	63	0.06 (−0.06, 0.18)		0.05 (−0.06, 0.17)		
Management of socioeconomic indicator in VPT and FT						
Matched design, no differences reported	22	1.00 (Reference)	0.049	1.00 (Reference)	0.043	0.6%
Independent sampling, no difference reported	34	0.05 (−0.11, 0.22)		0.02 (−0.14, 0.18)		
Differences reported, no adjustment[d]	34	−0.07 (−0.23, 0.09)		−0.10 (−0.25, 0.06)		
Differences reported and adjusted IQ score reported	1	−0.02 (−0.51, 0.48)		−0.07 (−0.54, 0.40)		
Undetermined	12	−0.13 (−0.34, 0.08)		−0.16 (−0.36, 0.05)		
Level of risk of the control group						
Low risk	99	1.00 (Reference)	0.049	1.00 (Reference)	0.044	0%
High risk	3	0.23 (−0.14, 0.60)		0.21 (−0.15, 0.57)		
Undetermined	1	−0.14 (−0.74, 0.47)		−0.16 (−0.75, 0.43)		

Note: p-values of test for linear trend (after exclusion of the studies with undetermined information): distribution of severe condition/disability: $p = .112$; distribution of SES indicators in both groups: $p = .262$.

Abbreviations: FT, full term; IQ, Intelligence quotient; VPT, very preterm.

[a]Models 1: univariable models adjusted for each study characteristics separately.

[b]Models 2: models 1 adjusted for the degree of preterm birth.

[c]R^2_{meta}: Proportion of variance explained by each study characteristic derived from Model 1; Total between-study variance = 0.0493.

[d]Among these studies, four studies provided estimates of the association adjusted for socioeconomic characteristics.

VPT children with severe conditions/impairment or the absence of controls for socio-economic characteristics between VPT and term groups, as expected.[21] The heterogeneity of the classifications of impairment and the low number of studies (16%) imputing missing data for children unable to complete standardised cognitive tests because of severe impairment, the latter of which has been shown to increase the magnitude of differences in IQ between very preterm and term-born comparison groups,[22] may partially explain our result concerning exclusions.

Effects sizes were also similar regardless of rates of attrition, even though high loss to follow-up, which occurs more often in families with lower social status[23,24] and can be related to the presence of impairments,[25] could be expected to underestimate effect size between groups. These findings are concordant with recent studies suggesting that loss to follow-up may have no or only modest effects. An individual participant data meta-analysis of cognitive outcomes among adults born preterm did not find an association between rates of attrition and IQ despite attrition rates ranging between studies from 19.1% to 59.5%,[26] while another study observed only a small, albeit systematic, effect of loss to follow-up on estimates of neurodevelopmental impairment at 2 years of age in two European cohorts.[23]

The one methodological characteristic with a substantial effect on study heterogeneity was the choice of a full scale versus short-form IQ test. Short-form tests have been shown to have sufficient psychometric quality for use in research settings in children with neurological disorders, including children born preterm.[27] However, while the subtests constituting the short forms vary, they often exclude working memory and processing speed, functions that are particularly affected in children born VPT.[28] Further, the longer time needed to administer full-scale tests may negatively impact the performance of VPT children. Our results provide support for considering this study feature in future meta-analyses as a potential source of between-study heterogeneity, but do not make it possible to conclude which approach best reflects the differences in cognition between VPT and term-born children.

Poor reporting of key information was common. This is a well-known challenge for meta-analyses, despite existing reporting guidelines (e.g. STROBE).[29] Further, for investigations of VPT birth, reporting requirements do not cover all relevant dimensions, such as methods for assessing and classifying children with severe developmental impairments. Harmonisation and better reporting of study features would improve the assessment of study quality for evidence synthesis and potentially the precision of estimates. While our study focused on studies of cognition, our results are relevant to primary studies and meta-analyses of other health and neurodevelopmental outcomes after VPT birth, as the same cohorts are used to investigate multiple outcomes.[30,31] Initiatives to harmonise reporting can benefit from those that bring together individual participant data from multiple cohorts and common schemas for cataloguing data from VPT cohorts to promote collaborative research, such as the European RECAP Preterm platform (https://platform.recap-preterm.eu/pub/).[21,26,32]

5 | CONCLUSIONS

Our findings illustrate the variability of design features in studies of VPT birth and cognition and inadequate reporting of important methodological information. Recommendations for harmonising the reporting of study design features in individual studies on VPT birth as well as their integration into systematic reviews with meta-analysis could improve the quality of evidence synthesis needed to inform clinical and policy decisions on the consequences of VPT birth.

AUTHOR CONTRIBUTIONS

Mariane Sentenac, Sabrina Twilhaar and Jennifer Zeitlin contributed to the study design. Material preparation and data collection were performed by Mariane Sentenac, Sabrina Twilhaar, Valérie Benhammou, Andrei Morgan and Jennifer Zeitlin, and the analysis was conducted by Mariane Sentenac. The first draft of the manuscript was written by Mariane Sentenac, Sabrina Twilhaar and Jennifer Zeitlin, and all authors commented on subsequent versions of the manuscript. All authors read and approved the final manuscript.

CONFLICT OF INTEREST

The authors have no relevant financial or non-financial interests to disclose.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Mariane Sentenac https://orcid.org/0000-0002-2156-154X
Sabrina Twilhaar https://orcid.org/0000-0001-7912-9883
Valérie Benhammou https://orcid.org/0000-0001-6568-0324
Andrei S. Morgan https://orcid.org/0000-0003-4143-1977
Samantha Johnson https://orcid.org/0000-0001-8963-7881
Anna Chaimani https://orcid.org/0000-0003-2828-6832
Jennifer Zeitlin https://orcid.org/0000-0002-9568-2969

REFERENCES

1. 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.* 2018;125:16-25.

2. Brydges CR, Landes JK, Reid CL, Campbell C, French N, Anderson M. Cognitive outcomes in children and adolescents born very preterm: a meta-analysis. *Dev Med Child Neurol.* 2018;60:452-468.

3. 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;172:361-367.

4. 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 Paediatr.* 2019;108:1567-1579.

5. 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 Pediatr.* 2015;169:1162-1172.

6. Sentenac M, Boutron I, Draper ES, et al. Defining very preterm populations for systematic reviews with meta-analyses. *JAMA Pediatr.* 2020;174:997-999.

7. Kerr-Wilson CO, Mackay DF, Smith GC, Pell JP. Meta-analysis of the association between preterm delivery and intelligence. *J Public Health.* 2012;34:209-216.

8. Fellman V, Hellstrom-Westas L, Norman M, et al. One-year survival of extremely preterm infants after active perinatal care in Sweden. *JAMA.* 2009;301:2225-2233.

9. Lyman GH, Kuderer NM. The strengths and limitations of meta-analyses based on aggregate data. *BMC Med Res Methodol.* 2005;5:14.

10. Sentenac M, Chaimani A, Twilhaar S, et al. 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. *Paediatr Perinat Epidemiol.* 2022;36:717-725.

11. Whiting P, Savović J, Higgins JPT, et al. ROBIS: a new tool to assess risk of bias in systematic reviews was developed. *J Clin Epidemiol.* 2016;69:225-234.

12. Langan D, Higgins JPT, Jackson D, et al. A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses. *Res Synth Methods.* 2019;10:83-98.

13. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. *BMC Med Res Methodol.* 2014;14:25.

14. Borenstein M, Hedges LV, Higgins JP, Rothstein HR. *Introduction to Meta-Analysis.* John Wiley & Sons; 2021.

15. IntHout J, Ioannidis JP, Rovers MM, Goeman JJ. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open.* 2016;6:e010247.

16. Thompson SG, Sharp SJ. Explaining heterogeneity in meta-analysis: a comparison of methods. *Stat Med.* 1999;18:2693-2708.

17. Mueller M, D'Addario M, Egger M, et al. Methods to systematically review and meta-analyse observational studies: a systematic scoping review of recommendations. *BMC Med Res Methodol.* 2018;18:44.

18. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *Eur J Epidemiol.* 2010;25:603-605.

19. Wells GA, Shea B, O'Connell D, et al. *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses.* 2000. https://web.archive.org/web/2021071612 1605id_/http://www3.med.unipmn.it/dispense_ebm/2009-2010/Corso%20Perfezionamento%20EBM_Faggiano/NOS_oxford.pdf

20. Metelli S, Chaimani A. Challenges in meta-analyses with observational studies. *Evid Based Ment Health.* 2020;23:83-87.

21. Sentenac M, Benhammou V, Aden U, et al. Maternal education and cognitive development in 15 European very-preterm birth cohorts from the RECAP preterm platform. *Int J Epidemiol.* 2022;50:1824-1839.

22. Johnson S, Wolke D, Marlow N. Outcome monitoring in preterm populations: measures and methods. *Z für Psychol.* 2008;216:135-146.

23. Piedvache A, van Buuren S, Barros H, Ribeiro AI, Draper E, Zeitlin J. Strategies for assessing the impact of loss to follow-up on estimates of neurodevelopmental impairment in a very preterm cohort at 2 years of age. *BMC Med Res Methodol.* 2021;21:118.

24. Teixeira R, Queiroga AC, Freitas AI, et al. Completeness of retention data and determinants of attrition in birth cohorts of very preterm infants: a systematic review. *Front Pediatr.* 2021;9:529733.

25. Wolke D, Söhne B, Ohrt B, Riegel K. Follow-up of preterm children: important to document dropouts. *Lancet.* 1995;345:447.

26. Eves R, Mendonça M, Baumann N, et al. Association of Very Preterm Birth or very low birth weight with intelligence in adulthood: an individual participant data meta-analysis. *JAMA Pediatr.* 2021;175:e211058.

27. van Ool JS, Hurks PPM, Snoeijen-Schouwenaars FM, et al. Accuracy of WISC-III and WAIS-IV short forms in patients with neurological disorders. *Dev Neurorehabil.* 2018;21:101-107.

28. Hutchinson EA, De Luca CR, Doyle LW, Roberts G, Anderson PJ. School-age outcomes of extremely preterm or extremely low birth weight children. *Pediatrics.* 2013;131:e1053-e1061.

29. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *J Clin Epidemiol.* 2008;61:344-349.

30. Kajantie E, Strang-Karlsson S, Evensen KAI, Haaramo P. Adult outcomes of being born late preterm or early term–what do we know? *Semin Fetal Neonatal Med.* 2019;24:66-83.

31. Lambert PC, Sutton AJ, Abrams KR, Jones DR. A comparison of summary patient-level covariates in meta-regression with individual patient data meta-analysis. *J Clin Epidemiol.* 2002;55:86-94.

32. Bamber D, Collins HE, Powell C, et al. Development of a data classification system for preterm birth cohort studies: the RECAP preterm project. *BMC Med Res Methodol.* 2022;22:8.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.