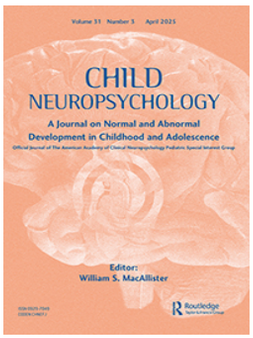

Routledge
Taylor & Francis Group

Child Neuropsychology

A Journal on Normal and Abnormal Development in Childhood and Adolescence

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/ncny20

The impact of gestational age on executive function in infancy and early-to-middle childhood following preterm birth: a systematic review

Raj Seraya Bhatoa, Simrit Nijjar, Joe Bathelt & Michelle de Haan

To cite this article: Raj Seraya Bhatoa, Simrit Nijjar, Joe Bathelt & Michelle de Haan (26 Feb 2025): The impact of gestational age on executive function in infancy and early-to-middle childhood following preterm birth: a systematic review, Child Neuropsychology, DOI: 10.1080/09297049.2025.2467950

To link to this article: https://doi.org/10.1080/09297049.2025.2467950

CHILD NEUROPSYCHOLOGY
https://doi.org/10.1080/09297049.2025.2467950

Routledge
Taylor & Francis Group

REVIEW ARTICLE

OPEN ACCESS Check for updates

The impact of gestational age on executive function in infancy and early-to-middle childhood following preterm birth: a systematic review

Raj Seraya Bhatoa[a], Simrit Nijjar[b], Joe Bathelt[a,c] and Michelle de Haan[d]

[a]Department of Psychology, Royal Holloway University London, Egham, UK; [b]EGA Institute for Women's Health, University College London, London, UK; [c]Department of Psychology, University of Amsterdam, Amsterdam, Netherlands; [d]UCL Great Ormond Street Institute of Child Health, University College London, Egham, UK

ABSTRACT

Lower gestational age (GA) is a risk factor for cognitive and developmental concerns following preterm birth. However, its impact on executive function (EF) is unclear based on conflicting conclusions across the literature. Moreover, as children below 4 years have largely been neglected from previous reviews, the impact of GA on EF within this early developmental period remains unclear. Hence, this systematic review investigated the impact of GA on EF following preterm birth in infancy and early-to-middle childhood. PubMed, Web of Science, and PsycInfo were searched for articles investigating the impact of GA on EF (inhibition, working memory, shifting) in preterm-born (<37 week gestation) and term-born participants aged 0–10 years. Eighteen studies were included. Most of the studies ($n = 10$) found no significant association between EF and GA. However, several limitations hindered conclusions to be drawn about the strength of this interpretation. Examples include inconsistencies in the theoretical underpinnings and operationalisations of EF, discrepancies in the reporting and measurement of GA, recruitment biases, and a paucity of infant or longitudinal studies available. Consequently, these issues may have contributed to inconsistent or null findings, and they must be addressed in future research to better clarify the impact of GA on EF in preterm-born infants and children.

ARTICLE HISTORY
Received 10 May 2024
Accepted 11 February 2025

KEYWORDS
Executive functions; gestational age; preterm birth; infancy; childhood

Globally, an estimated 13.4 million babies are born preterm (before 37 weeks gestation) every year (Lawn et al., 2023). Their gestational ages (GA) range from extreme preterm (EPT; <28 weeks GA), very preterm (VPT; 28–32 weeks GA) to moderate-to-late preterm birth (MLPT; 32–37 weeks GA; World Health Organization [WHO], 2022). Evidence indicates that survival rates have generally increased over time across high-income countries (Costeloe et al., 2012; Pierrat et al., 2017; Younge et al., 2017). Since the 1990s, these improvements in survival rates have also extended to babies born within lower GA ranges (<28 weeks; Victorian Infant Collaborative Study Group [VICS], 1997;

CONTACT Raj Seraya Bhatoa raj.bhatoa.2020@live.rhul.ac.uk; drbhatoa@outlook.com Department of Psychology, Royal Holloway University London, Egham Hill, Egham TW20 0EX, UK
Supplemental data for this article can be accessed at https://doi.org/10.1080/09297049.2025.2467950

WHO, 2012). Nevertheless, preterm-born babies remain at risk of experiencing a range of neurosensory, motor, and developmental concerns such as intellectual disabilities, visual and hearing impairments, cerebral palsy, and special education needs (Johnson & Marlow, 2017; Moore et al., 2012; National Institute for Health and Care Excellence [NICE], 2017). Moreover, the odds of experiencing these concerns are understood to increase as GA decreases (NICE, 2017). Consequently, studies have suggested that GA might be conceptualised as a "preterm continuum" (Baron et al., 2012; Sheehan et al., 2017), whereby EPT birth potentially represents the "severe end" of developmental adversity (Johnson & Marlow, 2017).

Importantly, cognitive-related concerns are understood to be the most prevalent developmental outcome following preterm EPT or VPT birth (Johnson & Marlow, 2017; Linsell et al., 2015). As such, preterm birth occurs at a time when the brain is still undergoing significant development and organisation during pregnancy (du Plessis & Volpe, 2018), whereby an interplay of inherent genetic characteristics, alongside environmental exposures and experiences, likely influence brain development during this period (Tau & Peterson, 2010). Nevertheless, these dynamic processes function within highly constrained and genetically organised stages of brain organisation (Stiles & Jernigan, 2010). As the greatest period of brain maturation and expansion occurs from mid-gestation onwards (Sampaio & Truwit, 2001; Yakovlev & Lecours, 1967) preterm birth likely disrupts this critical period of brain development in utero (Matthews et al., 2018; Vo Van et al., 2022). Subsequently, some evidence suggests that brain growth parameters, such as reduced regional brain volumes across the sensorimotor, temporal, and parieto-occipital regions, are potentially associated with adverse cognitive outcomes (lower IQ scores, for instance) in preterm-born children (Mulder et al., 2009). Yet structural brain injuries alone do not explain the frequency and range of developmental concerns that are prevalent in preterm-born populations (Bouyssi-Kobar et al., 2016; Miller & Ferriero, 2009). Indeed, cognitive difficulties can still be prevalent in preterm-born populations despite the absence of any apparent neurological deficits (Bhutta et al., 2002).

Of these, executive functions are commonly reported across the literature as a prominent concern within preterm-born populations (Mulder et al., 2009; van Houdt et al., 2019). *Executive functions* (EF) are higher-order cognitive and self-regulatory processes that underlie goal-directed behaviors (Cuevas & Bell, 2014). Its core components are theorised across the wider literature as inhibition, working memory, and shifting (Miyake et al., 2000). *Inhibition* represents the ability to inhibit prepotent responses and for ignoring distracting or competing information to sustain goal-directed behaviors (Nigg, 2000, 2017). *Working memory* (WM) constitutes being able to hold information in mind whilst processing or manipulating it (Baddeley, 1997; Nigg, 2017). *Shifting* (sometimes referred to as cognitive flexibility) represents the ability to switch between tasks or mental sets (Miyake et al., 2000).

EF capabilities are critical across many life domains, including academic attainment (Blair & Razza, 2007; St Clair-Thompson & Gathercole, 2006), friendships (Holmes et al., 2016; Miller et al., 2020), managing emotions (Riggs et al., 2006) and behavior (Dias et al., 2017), and supporting mental and physical health outcomes (Allan et al., 2016; Cushman et al., 2022). As such, evidence highlights how reduced EF abilities in preterm-born children are associated with academic difficulties (despite the absence of neurosensory

impairments; Allotey et al., 2018; Johnson & Marlow, 2017) and challenges with peer relationships (Reyes et al., 2019), regulating emotions (Taylor & Clark, 2016) and for managing behaviors (Loe et al., 2019; Schnider et al., 2020). Moreover, studies with preterm-born populations have highlighted that EF difficulties can also occur independently of any identifiable concerns with general cognition (Day, 2018; Mulder et al., 2009). This is especially pertinent, given that developmental instruments used to examine broad cognitive outcomes across infancy may lack specificity for detecting EF difficulties (Potharst et al., 2012). Thus, EF-related difficulties that are apparent at school age may not be predicted by current developmental surveillance assessments within the first few years of life, meaning that opportunities to implement early interventions to reduce their impact are missed (Doyle et al., 2014). Unfortunately, by remaining undetected in the meantime, EF-related difficulties may nevertheless have adverse effects on a child's socio-emotional wellbeing, academic attainment, and quality of life, as well as on the family (NICE, 2017). Subsequently, knowing that lower GA might increase a child's odds of experiencing a range of neurological and developmental concerns as previously described, it raises the question of whether lower GA might also be associated with more pronounced EF difficulties.

Several recent systematic reviews have presented conflicting conclusions about the association between GA and EF in preterm-born populations (Brydges et al., 2018; Linsell et al., 2015; Mulder et al., 2009; Reveillon et al., 2018; Sandoval et al., 2022; van Houdt et al., 2019). A potential reason is that previous systematic reviews did not consistently distinguish between the core components of EF (e.g., Brydges et al., 2018). Whilst a consensus is yet to be reached across the literature regarding a definitive theoretical model of EF, there is nevertheless general agreement that its inter-related but separable core components likely mature at different rates (Best & Miller, 2010; Nigg, 2017) across development and early adulthood (Garon et al., 2008). Understanding how GA is associated with these respective abilities is therefore crucial toward better understanding its impact on particular aspects of EF across development.

However, the degree to which each core EF component has been evaluated in previous systematic reviews has also varied considerably. With regard to inhibition, the literature provides conflicting conclusions regarding its association with GA in preterm-born children. Two recent systematic reviews have concluded there to be no association (Reveillon et al., 2018; van Houdt et al., 2019), yet at least three more reviews have found evidence to support this association (Linsell et al., 2015; Mulder et al., 2009; Sandoval et al., 2022). However, positive associations were not consistently found across all the relevant studies that were included in each respective review (Linsell et al., 2015; Sandoval et al., 2022). Moreover, associations were at times based on preterm-born participant groups with narrow GA ranges (e.g., Brumbaugh et al., 2013 [in Sandoval et al., 2022]; Orchinik et al., 2011 [in Linsell et al., 2015]), which limits the generalisability of these results in children with gestational ages outside of this range.

WM has been less frequently assessed across recent systematic reviews of EF in preterm-born children (Reveillon et al., 2018), with previous authors having reported there to be an insufficient number of studies available across the preterm literature to perform a meta-analysis and ascertain the association between GA and WM (Mulder et al., 2009). While a previous review by Mulder et al. (2009) has nevertheless reported there to be a positive association, this was based on limited data available comprising

participants aged between 8 and 17 years (Mulder et al., 2009); thus, limiting conclusions to be drawn about how these results might apply to younger preterm-born populations. Similarly, where previous reviews comprised studies of preterm-born participants with narrow GA ranges (Linsell et al., 2015; van Houdt et al., 2019), this further limits the generalisability of their conclusions to children with gestational ages outside of this range.

Finally, shifting has been the least examined component in relation to its association with GA across previous reviews (Mulder et al., 2009; Reveillon et al., 2018; Sandoval et al., 2022). While one previous review by Linsell et al. (2015) reported no association between GA and shifting, this conclusion was again based on preterm-born participants presenting with narrow GA ranges (Orchinik et al., 2011) - thus, not indicating how the impact of GA on shifting may differ across wider GA ranges. Furthermore, in a previous meta-analysis by van Houdt et al. (2019), meta-regression analyses were used to quantify EF impairments and assess the impact of GA on EF effect sizes. However, the study authors stated that they retrieved only five studies that reported data on cognitive flexibility (compared to 13 studies for inhibition and 25 studies for WM). Subsequently, this EF component was not included in the analyses – thus, implying the limited availability of studies in the preterm literature to permit a thorough analysis of its association with GA in preterm-born populations (van Houdt et al., 2019). Consequently, the relationship between GA and shifting remains undetermined and would benefit from further exploration in preterm-born populations.

Taken together, conflicting conclusions across the limited data available mean that the association between GA and EF within preterm-born children remains unclear. Importantly, none of the above reviews reviewed the core components of EF in preterm-born participants younger than 4 years (Brydges et al., 2018; Linsell et al., 2015; Reveillon et al., 2018; van Houdt et al., 2019) or 2 years of age (Mulder et al., 2009; Sandoval et al., 2022). Yet, research suggests that early EF abilities initially emerge in term-born populations within the first year of life (Diamond, 1985; Garon et al., 2008), whereby controlled, goal-directed behaviors that are posited to reflect EF capabilities have been indicated at 6 months (Diamond, 1985; Diamond & Doar, 1989; Pelphrey & Reznick, 2004; Pelphrey et al., 2004; in Miller et al., 2023), 9 months (Holmboe et al., 2018) and 12 months of age (Diamond, 1985; Marcovitch & Zelazo, 1999; Piaget, 1954; in Miller et al., 2023). Therefore, the "building blocks" of EF are potentially laid in early infancy, for which early difficulties may have lasting effects throughout development (Cuevas & Bell, 2010, 2014; Ross-Sheehy et al., 2017). Yet, it is unclear whether conclusions from previous reviews regarding the impact of GA on EF in preterm-born populations are applicable in infancy, due to their exclusion from previous reviews. Subsequently, there is a need to explore the extent of this relationship within this early developmental period.

Further clarity may have important clinical implications for how preterm-born populations are supported in the future. For example, evidence suggests that preterm-born children who demonstrate EF difficulties may continue to demonstrate persistent difficulties in adolescence (Reveillon et al., 2018) and adulthood (Eryigit Madzwamuse et al., 2015; Johnson & Marlow, 2017). Therefore, important questions emerge regarding whether EF difficulties are detectable within infancy, whether its magnitude is differentiated according to GA, and whether such difficulties persist in childhood. Investigating this further may elucidate the potentially predictive value of GA for

understanding the magnitude and evolution of EF abilities across development. In turn, this may inform the development of earlier, targeted developmental surveillance assessments (NICE, 2017) and interventions, which capitalise on improved understanding regarding whether specific GA ranges may pose more risk – or protection – toward EF within preterm-born populations (NB: A recent review by Duncan et al. (2024) provides further information regarding the characteristics and efficacy of current interventions targeting EF in infancy and early childhood). Consequently, this study aimed to build upon the current literature by conducting a systematic review of studies that investigated the impact of GA on the core components of EF following preterm birth (including comparisons of core EF abilities in preterm-born and term-born participants) in infancy and early-to-middle childhood.

Methods

The protocol was completed according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines.

Search strategy

Electronic databases (PubMed, Web of Science and PsycInfo) were searched on February 5, 2023. The current study extended van Houdt et al.'s (2019) comprehensive search terms to include terms relevant to infancy. The revised search terms were also reviewed by an academic librarian. (See Search terms in Appendix/Supplementary material).

Study selection

Inclusion criteria for study selection built on van Houdt et al.'s (2019) review, to include (1) preterm-born participants born below 37 weeks gestation, (2) a comparison group of term-born participants, (3) the "administration of WM, inhibition, and/or shifting tasks," and (4) articles "published in an English-language, peer-reviewed journal." Additional criteria included: (5) articles published from January 17, 2017 onwards (following on from van Houdt et al.'s (2019) study), (6) articles that assessed the impact of GA on inhibition, WM, and/or shifting capabilities, including articles where preterm-born participants were sub-categorised according to birthweight instead of GA (e.g., extremely low versus very low birthweight), but the average/median GA was reported to be significantly different between the respective subgroups (e.g., Nagy et al., 2022), and (7) articles with children aged 0–10 years, including articles where, despite comprising participants both *within* and *outside* the maximum age limit (10 years), the data has been sufficiently individualised in its published format to allow the analysis of information related to the intended age group (0–10 years; Silva et al., 2022).

Although van Houdt et al. (2019) restricted their study inclusion to papers in which participants' years of birth were "1990 or later (i.e., after the introduction of antenatal steroids and surfactant supplementation)," this was not actioned in the current review for two reasons. Firstly, whilst advancements in neonatal care have likely contributed to improved survival rates following preterm birth (Younge et al., 2017), the assumption

that neurodevelopmental outcomes of survivors have improved in parallel is not supported by the literature (Marlow et al., 2021). Rather, deteriorating health-related quality of life outcomes have since been reported (Ni et al., 2022). Second, a recent Cochrane review demonstrated that there is no increased benefit for neurodevelopmental delay after corticosteroid use (Roberts et al., 2017), with additional evidence suggesting that the absence of corticosteroid use is not predictive of poorer cognitive function in children aged below 5 years (Linsell et al., 2015). Consequently, it was interpreted that data from participants born prior to 1990 would still be informative.

Exclusion criteria included (1) systematic reviews, meta-analyses, editorials, guidelines, case studies, commentaries, book sections, and conference proceedings, (2) articles that did not compare EF abilities in preterm-born versus term-born control cohorts (including where GA or low birthweight were investigated as potential covariates within clinical conditions that were separate to preterm-born presentations), (3) articles that assessed EF abilities after the administration of medical, surgical, or therapeutic interventions, (4) articles that reported EF outcomes using animal model, magnetic resonance imaging (MRI), visual-evoked potential (VEP), physiological (e.g., skin conductance), magnetoencephalography (MEG) or electroencephalogram (EEG)-based data, and (5) EF subtests that were not located in test manuals as described (i.e., Digit Span and Colour Shape from the Neurodevelopmental Assessment [NEPSY]; Korkman et al., 1998; Stålnacke et al., 2019), or which have since been removed from superseding versions of assessment batteries to enhance their clinical utility and/or their theoretical foundations (i.e., Tower and Visual Attention subtests from the Neurodevelopmental Assessment [NEPSY-II] (Korkman et al., 2007); Animal Pegs and Arithmetic subtests from the Wechsler Preschool and Primary Scales of Intelligence [WPPSI-III] (Lichtenberger & Kaufman, 2004; Wechsler, 2002), and the Picture Arrangement subtest from the Wechsler Intelligence Scale for Children [WISC-IV] (Wechsler, 2003; Williams et al., 2003).

If multiple papers reported on overlapping cohorts of children born preterm on the same EF domain(s), the study with the most complete data for that domain was selected. When multiple papers reported on overlapping cohorts of children born preterm but the papers differed in the EF domains described, all papers were included in analyses. If it was not clear whether cohorts were overlapping, the authors of the studies were contacted where possible.

Notably, the wider literature presents conflicting operationalisations of EF subtests. For example, the digit span (forward recall) task is described as a measure of WM in some studies (Lean et al., 2021; Pérez-Pereira, Fernández, et al., 2020) and as a measure of short-term memory in others (Diamond, 2013; Gathercole & Alloway, 2006). Subsequently, measures were grouped according to the definitions of the respective executive components as described in this paper, and not according to terms used by the studies in question (Reveillon et al., 2018).

Records were screened and reviewed by one reviewer (RSB). All queries were independently reviewed in discussions with other authors (JB, MdH). Subsequent actions to resolve these queries were taken following agreement between members.

Figure 1 summarises the selection of articles obtained. Two thousand two hundred and ninety-eight records were identified across three databases (PubMed: $n = 1113$; PsycInfo: $n = 269$; Web of Science: $n = 916$). Five hundred and thirty-seven duplicate

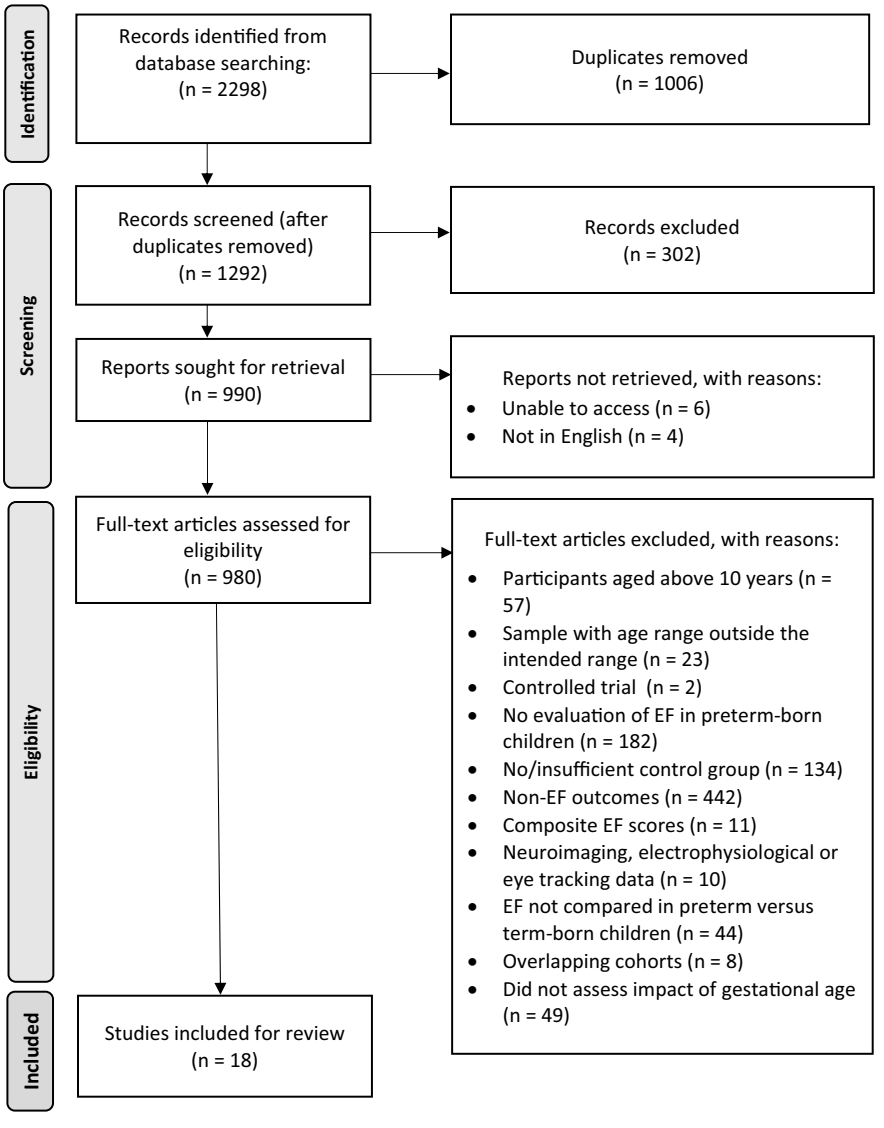

Figure 1. PRISMA flow chart of the study selection process.

records were automatically removed via Endnote, and additional 469 duplicate records were manually removed. The remaining 1292 records were screened, of which 302 records were eliminated based on the exclusion criteria. The reasons for exclusion included reviews or meta-analyses ($n = 65$), animal studies ($n = 96$), case reports ($n = 15$), adult studies ($n = 54$), theoretical perspective articles ($n = 1$), qualitative research ($n = 1$), duplicates ($n = 4$), books ($n = 2$), corrections ($n = 5$), commentaries, opinions, or panel discussions ($n = 12$), editorials ($n = 3$), executive summaries ($n = 1$), guidelines ($n = 9$), and protocols ($n = 34$). The remaining 990 records were sought for retrieval. Of these, 10 reports were unable to be retrieved. Nine hundred and eighty full-text articles were subsequently assessed based on the inclusion and inclusion criteria. The reasons for

excluding full-text articles are summarized in Figure 1. A total of 18 articles were selected in the review.

Data extraction

The following information was extracted from the final 18 articles:

- Sample characteristics – mean, standard deviations, and ranges for GA and chronological or corrected ages at assessment. Percentages of females and information regarding recruitment sources and locations were also recorded.
- Study characteristics – information regarding study design, methodology, and measures.
- Key findings – main findings related to the impact of GA on EF performance and comparisons of EF performance between preterm-born and term-born groups.
- Additional information related to study setting, analysis procedures, and psychometric properties of measures where relevant.

Study quality

The methodological quality of papers was assessed using an adapted version of Wong & Edwards' (2013) modified checklist of the Quality Assessment Tool for Quantitative Studies of the Effective Public Health Practice Project (EPHPP, 2002), as summarised in Table 1. Studies were appraised based on their study design, representativeness of the study population, quality of GA and EF outcome data obtained, and the

Table 1. Methodological quality assessment checklist criteria (adapted from Wong & Edwards, 2013).

Area assessed	Component	
Study design	Strong	Longitudinal design[a]
	Moderate	Cross-sectional design
	Weak	Not applicable
Representativeness of study population	Strong	Data from national cohort
	Moderate	Recruited from multiple centres
	Weak	Convenience sample (single-centre)
Data on gestational age	Strong	Gestational ages clearly defined (minimum reporting requirements = *M, SD*. Additional = range) and reported separately for preterm and term-born groups
	Moderate	Gestational ages clearly defined (minimum = M, SD. Additional = range) and reported for preterm group only
	Weak	Minimum requirements not reported
Data on executive function	Strong	Assessment valid and reliable
	Moderate	Either validity or reliability known
	Weak	Both validity and reliability are unknown
Statistical analyses and adjustments of confounders	Strong	Multivariate analysis with adequate reporting of relevant confounds and having controlled for these in the statistical analyses
	Moderate	Multivariate or stratification analysis. Confounds controlled via matching or stratification but no statistical controlling of confounds reported
	Weak	No adjustments for confounds reported

[a]For the purpose of this review, where studies assessed EF at additional timepoints that fell outside of the age range as defined in the inclusion criteria (i.e., 0–10 years), data from those time points were not considered in the above ratings.

appropriateness of statistical analyses undertaken (including adjustment for confounders). Each component was rated as either "strong," "moderate" or "weak," with a global quality rating assigned to each study based on the number of "weak" component ratings obtained (Wong & Edwards, 2013). Specifically, studies with no "weak" ratings were judged "strong" overall, studies with one "weak" rating were judged "moderate" overall, and studies with two or more "weak" ratings were judged "weak" overall. Any studies that were assigned a "weak" rating to every component were excluded from the review. In addition to Wong & Edwards' (2013) procedure for assigning global quality ratings, a new "total quality score" was also calculated. This was determined by assigning each "weak," "moderate" and "strong" rating a score of 0, 1, or 2, respectively. These scores were then tallied to provide a "total quality score" per study (ranging from 0 to 10).

The EPHHP tool was chosen based on its strengths for considering important factors for appropriately assessing quality and susceptibility to bias in observational studies, including study participation, measurement, and confounding factors (Sanderson et al., 2007). Two researchers (RSB and SN) independently conducted ratings to reduce bias, with discrepancies discussed and managed by a third member of the team (JB) where necessary.

Results

Study quality

The methodological quality of studies is summarised in Table 2. No studies were removed from the review based on being assigned a global "weak" rating. The average inter-rater agreement for the two raters was kappa = 0.90, indicating a satisfactory level of agreement between the two raters.

Of the 18 studies reviewed, only one study (5.6%) was assigned a "strong" global quality rating (i.e., no "weak" ratings) which had also obtained a total quality score of 9 (out of 10). The remaining 17 studies (94.4%) were assigned weak-to-moderate global quality ratings. Their total quality scores ranged from 2 to 7 (average = 4.8; median = 4.5; mode = 3).

With regard to study design, it is noted that the majority of studies (n = 16 or 88.9%) were assigned "moderate" ratings based on having utilised cross-sectional designs. The remaining two studies (11.1%) were rated as "strong," based on having comprised longitudinal designs.

With regard to study population, most studies (n = 13 or 72.2%) were assigned "weak" ratings based on having recruited preterm-born participants via single clinics. One study (5.6%) was rated as "moderate" based on having recruited participants from multiple centres, and the remaining 4 studies (22.2%) were rated as "strong" based on having recruited participants from national cohorts.

Conversely, the quality ratings varied considerably for the remaining areas assessed. Regarding GA data, seven studies (38.9%) were rated as "strong" based on having provided sufficient detail as to how GA was defined (i.e., minimum reporting requirements met for both preterm and term-born groups). Four studies (22.2%) were rated as "moderate" (i.e., minimum reporting requirements met for preterm-born group only),

Table 2. Methodological quality ratings (adapted from Wong & Edwards, 2013).

Lead author, year	Study design	Study population	GA data	EF data	Statistics and confounds	Global quality rating	Total quality score
Christians & Chow (2022)	Moderate	Strong	Weak	Strong	Strong	Moderate	7
Cunha et al. (2018)	Strong	Weak	Weak	Moderate	Moderate	Weak	4
Domellöf et al. (2020)	Moderate	Weak	Strong	Strong	Moderate	Moderate	6
Feng et al. (2018)	Moderate	Weak	Weak	Weak	Moderate	Weak	2
Fernandez-Baizan et al. (2020)	Moderate	Weak	Moderate	Weak	Weak	Weak	2
Fernandez-Baizan et al. (2021)	Strong	Weak	Moderate	Weak	Weak	Weak	3
Giordano et al. (2017)	Moderate	Weak	Strong	Strong	Moderate	Moderate	6
Hodel et al. (2017)	Moderate	Weak	Strong	Moderate	Moderate	Moderate	5
Hodel et al. (2019)	Moderate	Weak	Strong	Weak	Moderate	Weak	4
Ma et al. (2022)	Moderate	Strong	Weak	Strong	Strong	Moderate	7
Nagy et al. (2022)	Moderate	Weak	Moderate	Weak	Strong	Weak	4
O'Meagher et al. (2017)	Moderate	Weak	Weak	Strong	Strong	Weak	5
Pérez-Pereira, Fernández, et al. (2020)	Moderate	Weak	Weak	Strong	Weak	Weak	3
Pérez-Pereira, Martínez-López, et al. (2020)	Moderate	Moderate	Strong	Strong	Weak	Moderate	6
Reyes et al. (2019)	Moderate	Strong	Strong	Weak	Strong	Moderate	7
Sato et al. (2019)	Moderate	Weak	Moderate	Weak	Moderate	Weak	3
Shinya et al. (2022)	Moderate	Weak	Weak	Strong	Weak	Weak	3
Stålnacke et al. (2019)	Moderate	Strong	Strong	Strong	Strong	Strong	9

"Global quality rating" - studies with two or more "weak" ratings were judged as "weak" overall. Studies with one "weak" rating were judged as "moderate" overall. Studies with no "weak" ratings were judged as "strong" overall. "Total quality score" - each rating per domain was assigned a score of 0 ("weak"), 1 ("moderate"), and 2 ("strong"), respectively. These scores were then tallied to provide a "total quality score" for each study. Abbreviations – GA, gestational age; EF, executive function.

and the remaining 7 studies (38.9%) were rated as "weak" (minimum requirements not reported for either participant group).

Regarding EF data, nine studies (50%) were rated as "strong" based on having utilised EF assessments that were deemed valid and reliable. Two studies (11.1%) were rated as "moderate" based on having utilised EF assessments that were deemed valid or reliable. Finally, seven studies (38.9%) were rated as "weak" based on having utilised EF assessments for which their validity and reliability were unknown.

Based on statistical analyses, six studies (33.3%) were rated as "strong" based on having utilised multivariate analyses (plus adequate reporting of relevant confounds which were controlled for in the analyses). Seven studies (38.9%) were rated as "moderate" based on having utilised multivariate or stratification analysis (with matching or stratification – but no statistical controlling – of confounds having been reported). Finally, five studies (27.8%) were assigned a "weak" rating due to having reported no adjustments for confounds. Implications are discussed further in the Discussion section.

Sample characteristics

Most studies were undertaken in Western countries: Europe ($n = 9$) including Spain ($n = 4$) Sweden ($n = 2$), Vienna ($n = 1$), Hungary ($n = 1$) and Germany ($n = 1$) and North America ($n = 6$), including USA ($n = 5$) and Canada ($n = 1$). Two studies were conducted

in Asia, including China ($n = 1$) and Japan ($n = 1$). The remaining study was conducted in Australia.

Studies varied considerably in their recruitment locations for preterm-born participants. Most studies ($n = 8$; 44%) comprised participants recruited via hospital services including tertiary care, neonatal units, and follow-up clinics (Domellöf et al., 2020; Feng et al., 2018; Fernandez-Baizan et al., 2020; Giordano et al., 2017; O'Meagher et al., 2017; Pérez-Pereira, Fernández, et al., 2020; Sato et al., 2019; Shinya et al., 2022). Four studies recruited participants from national cohort studies (Christians & Chow, 2022; Ma et al., 2022; Reyes et al., 2019; Stålnacke et al., 2019). One study recruited families across four hospitals (Pérez-Pereira, Martínez-López, et al., 2020). Two studies recruited participants from databases of families interested in partaking in research (Hodel et al., 2017, 2019). One study recruited participants via a follow-up university programme or via advertising on a relevant website associated with preterm birth (Nagy et al., 2022). Three studies did not specify where participants were recruited from. Of these, one study stated that data collection took place in a hospital or across preschool settings with preterm or term-born participants, respectively (Fernandez-Baizan et al., 2021). Another study acknowledged that its participants were part of a longer longitudinal project (Pérez-Pereira, Fernández, et al., 2020).

The sample sizes of preterm-born participants also varied across studies, ranging from 20 to 2065 participants across studies which assessed inhibition, 15 to 2065 participants across studies that assessed WM, and 52 to 2065 across studies that assessed shifting.

Regarding participants' ages, six studies comprised participants aged below 2 years: two of which assessed inhibition (Reyes et al., 2019; Shinya et al., 2022), three studies which assessed WM (Cunha et al., 2018; Fernandez-Baizan et al., 2021; Hodel et al., 2017), and one study having assessed both inhibition and WM (Feng et al., 2018). The majority of studies that assessed inhibition comprised participants aged 4–6 years ($n = 4$). Similarly, the majority of studies that assessed WM comprised participants aged between 4 and 8 years ($n = 8$). Studies that assessed shifting comprised participants aged between 4 and 6 years ($n = 2$) or 9–10 years ($n = 2$).

Within each study, participants' gestational ages varied considerably. Four studies comprised participants whose gestational ages ranged from extremely preterm [EPT] (23–24 weeks) to very preterm [VPT] (<32 weeks; Cunha et al., 2018; Giordano et al., 2017; O'Meagher et al., 2017; Stålnacke et al., 2019). Five studies comprised participants whose GA ranges spanned from VPT (28–32 weeks) to moderate-to-late [MLPT]/late preterm [LPT] (34–36 weeks; Christians & Chow, 2022; Feng et al., 2018; Fernandez-Baizan et al., 2021; Hodel et al., 2017), though one study did not specify the exact lower bound of its respective GA range (Reyes et al., 2019). Six studies comprised samples with the largest GA ranges, spanning from ELPT (<28 weeks) to MLPT (34–36 weeks; Domellöf et al., 2020; Nagy et al., 2022; Pérez-Pereira, Martínez-López, et al., 2020; Shinya et al., 2022), though two studies did not specify the exact lower bound of their GA ranges (Ma et al., 2022; Pérez-Pereira, Fernández, et al., 2020). One study comprised participants who all fell within the MLPT GA range (32–36 weeks; Hodel et al., 2019). The final two studies only reported the mean and standard deviations of gestational ages within their preterm-born and term-born participant groups (Fernandez-Baizan et al., 2020; Sato et al., 2019).

Finally, inclusion or exclusion criteria for preterm-born participants were unclear across three studies (Cunha et al., 2018; Domellöf et al., 2020; Fernandez-Baizan et al., 2020), whereas the remaining studies made some reference to study inclusion. Based on the above, it is apparent that the preterm-born participants were a very heterogeneous group.

Study characteristics

Design

Sixteen studies (89%) employed a cross-sectional design to analyse the EF components of interest. The remaining two studies (11%) employed a longitudinal design (Cunha et al., 2018; Fernandez-Baizan et al., 2021).

Measures

In the current review, the number of EF components that were measured in each study was highly variable. Descriptions of each measure are provided in Appendix B. Nine studies assessed inhibition, 14 studies assessed WM and four studies assessed shifting. Therefore, these studies have been organised according to each of these components across Tables 3–5, respectively.

Each study varied considerably in the amount and format of tasks that they administered to assess EF. For example, four studies measured more than one EF component, which were each assessed using a different EF task (Giordano et al., 2017; Ma et al., 2022; Nagy et al., 2022; O'Meagher et al., 2017). Of these, one study used a mixture of task-based and informant-rated measures to assess inhibition and shifting, respectively (O'Meagher et al., 2017). In contrast, two studies solely utilised an informant-rated measure to assess either a single EF component (Shinya et al., 2022) or multiple EF components (Pérez-Pereira, Martínez-López, et al., 2020). Eight studies had each assessed one EF component in total, using a single relevant task (Christians & Chow, 2022; Domellöf et al., 2020; Fernandez-Baizan et al., 2020; Hodel et al., 2019; Pérez-Pereira, Fernández, et al., 2020; Reyes et al., 2019; Sato et al., 2019; Stålnacke et al., 2019). Conversely, one study assessed a specific EF component in detail (WM) using two object retrieval-based tasks over multiple timepoints (Fernandez-Baizan et al., 2021). Finally, three studies (Cunha et al., 2018; Feng et al., 2018; Hodel et al., 2017) utilised object retrieval-based tasks that were described as "problem-solving" or "planning" tasks for assessing aspects of goal-directed behavior, which were interpreted as drawing on EF capabilities in the form of developmentally appropriate WM demands (de Haan, 2013; Hodel et al., 2017; Miller et al., 2023). In addition, two of these studies used additional object retrieval tasks to concurrently measure two EF components (inhibition and WM). In one study, each EF component was represented by a specific outcome (Feng et al., 2018). In another study, the outcomes extracted were intended to simultaneously represent both inhibition and WM (Hodel et al., 2017). As the latter task in question (A-not-B) is theorised to represent inhibitory control supporting overall WM demands (Diamond, 2013; Nigg, 2000), this task was coded for the benefit of this review as representing emerging WM.

Table 3. Characteristics and main findings of studies assessing inhibition ($n = 9$).

Lead author, date	Design, Methods	Case Group (PT)	Control Group (TB)	Preterm Subgroups – if avail	Measure(s)	Main Findings
Feng et al. (2018)	Cross-sectional, between-groups comparisons	n = 88 GA = NR, NR (28–36) Age = 8 m[a] Females = 50%	n = 88 GA = NR, NR (37–42) Age = 8 m Females = 50%	n = 9 (28–31) n = 79 (32–36)	One, Two, Three-Cup Task	*PT vs. TB* = PT < TB across both task outcomes. *GA* = No significant differences in performance between PT subgroups found.
Giordano et al. (2017)	Cross-sectional, between-groups comparisons	n = 52 GA = 28.71, 2.02 (23–32) Age = 5.7y, NR (5.5–6.0)[c] Females = 48.1%	n = 52 GA = 39, 1.38 (NR) Age = 5.8y, NR (5.5–6.2) Females = 48.1%	n = 23 (23–28) n = 29 (29–32)	Distractibility; Divided Attention; Go/No-go	*PT vs. TB* = PT group showed significantly greater variability in performance and slower reaction times during Distractibility task, and significantly slower reaction times during Divided Attention task (visual condition and total time). No further significant between-group differences in task performance found. *GA* = Lower GA PT subgroup produced significantly more errors than higher GA PT subgroup during Divided Attention task. No further significant differences in task performance between PT subgroups reported.
Hodel et al. (2019)	Cross-sectional, between-groups comparisons, correlational analyses	n = 63 GA = 35.14, 1.46 (32–36) Age = NR, NR (4.5–5.0y)[c] Females = 51%	n = 66 GA = 39.80, 0.94 (37–42) Age = NR, NR (4.5–5.0y) Females = 49%	NR	Modified Directional Stroop	*PT vs. TB* = No significant differences found. *GA* = No significant associations found.
Ma et al. (2022)	Cross-sectional, between-groups comparisons	n = 2065 GA = NR, NR (≤33–39) Age = 10.0y, 0.61[↑] (NR)[c] Females = 47%	n = 9138 GA = NR, NR (≥40) Age = 9.9y, 0.63 (NR) Females = 48%	n = 457 (≤33) n = 555 (34–35) n = 591 (36) n = 578 (37–39)	Flanker	*PT vs. TB* = Based on comparisons between TB and PT subgroups, no significant effect of group found. *GA* = Based on post-hoc pairwise differences that compared the TB group (≥40 weeks GA) with each PT subgroup, significantly lower scores found for the ≤33 weeks GA subgroup only.

(*Continued*)

Table 3. (Continued).

Lead author, date	Design, Methods	Case Group (PT)	Control Group (TB)	Preterm Subgroups – if avail	Measure(s)	Main Findings
Nagy et al. (2022)	Cross-sectional, between-groups comparisons, regression analyses	n = 72 GA = 28.75, 2.21‡ (24–34) Age = 113.7 m, 3.51 (108–119)[c, ^] Females = 49.7%‡‡	n = 33 GA = NR, NR (38–41) Age = 113.7 m, 3.51 (108–119)[^] Females = NR	n = 32 (categorized as ELBW)[>] n = 40 (categorized as VLBW)[>]	Colour-Word Stroop	*PT vs. TB* = Based on comparisons between TB and PT subgroups, ELBW < TB for interference error only. No further significant differences found. *GA* = No significant effects of GA found.
O'Meagher et al. (2017)	Cross-sectional, between-group comparisons, regression analyses	n = 141 GA = 29.69, NR (23.6–32.5) Age = 49.10 m, NR (48–58)[b] Females = 50.4%	n = 77 GA = NR, NR (≥38) Age = 54.86 m, NR (48–67) Females = 41.6%	NR	Shape School; Day-Night Stroop; BRIEF-P inhibit Subscale (parent and teacher-rated versions)	*PT vs. TB* = PT < TB across all measures, except BRIEF-P (parent version). *GA* = Significant positive association based on BRIEF-P (parent version) only.
Pérez-Pereira, Martínez-López, et al. (2020)	Cross-sectional, between-group comparisons	n = 108 GA = 32.34, 0.89+ (26–36) Age = NR, NR (4y)[c] Females = NR	n = 34 GA = 39.68, 1.52(>36–42) Age = NR, NR (4y) Females = NR	n = 34 (<32) [29.62, 1.46] n = 33 (32–33) [32.58, 0.50] n = 41 (34–36) [34.81, 0.72]	CHEXI inhibition subscale	*PT vs. TB* = No significant differences found. *GA* = Significantly lower scores in 34–36 weeks GA subgroup compared to 32–33 weeks GA subgroup, only.

(*Continued*)

Table 3. (Continued).

Lead author, date	Design, Methods	Case Group (PT)	Control Group (TB)	Preterm Subgroups – if avail	Measure(s)	Main Findings
Reyes et al. (2019)	Cross-sectional, between-group comparisons, regression analyses	n = 639 GA = 33.68, 0.83[+] (<32–38) Age = NR, NR (20 m)[a] Females = 48.7%	n = 542 GA = 39.9, 0.70 (39–41) Age = NR, NR (20 m) Females = 50.7%	VPT: n = 199 (<32) [29.6, 1.50] MPT: n = 79 (32–33) [32.50, 0.50] LPT: n = 186 (34–36) [35.10, 0.80] Early term: n = 175 (37–38) [37.50, 0.50]	Snack Delay	*PT vs. TB* = Lower GA groups < higher GA groups [post-hoc information NR]. *GA* = Significant positive association between GA and total waiting time.
Shinya et al. (2022)	Cross-sectional, between-group comparisons	n = 20 GA = 31.82, 1.54 (25.57–36)[Δ] Age = 18.81 m, 0.7 (18.17–20.47)[a, ΔΔ] Females = 37% [Δ]	n = 20 GA = 39.29, 1.22 (37.14–41.43) [Δ] Age = 18.48 m, 0.38 (17.68–19.12) Females = 48%[Δ]	VPT: n = 9 (25.57–31.86) [28.96, 2.20] [Δ] MLPT: n = 11 (32.86–36) [34.67, 0.88] [Δ]	Early Childhood Behaviour Questionnaire	*PT vs. TB* = Based on comparisons between TB and PT subgroups, no significant differences found. *GA* = VPT < MLPT only.

Groups – Number (n) of participants; Gestational age (GA) in weeks, *M, SD* (Range); Age at assessment in months (m) or years (y), *M, SD* (Range) [[a], corrected age; [b], chronological (uncorrected) age; [c], not specified]; % of Females in group. Preterm subgroups – Number (n) of participants per subgroup (if available – GA in weeks, *M, SD*). Main findings – *PT vs. TB*, comparisons of executive function [EF] performance between preterm-born (PT) and term-born (TB) groups; *GA*, findings related to gestational age (GA) and EF performance; <, significantly lower EF performance found in former group compared to latter group. Abbreviations – PT, preterm-born; TB, term-born; if avail, if available; GA, gestational age; NR = information was not reported; ELBW, extremely low birthweight; VLBW, very low birthweight; BRIEF-P, Behavior Rating Inventory of Executive Function – Preschool Version (parent and teacher-rated); CHEXI, Childhood Executive Function Inventory; VPT, very preterm; MPT, moderately preterm; LPT, late preterm; MLPT, moderate-to-late preterm.

For the purpose of this review, where studies did not provide sufficient detail regarding participant characteristics according to preterm-born versus term-born group, this information was approximated based on the available data provided in each study. The steps undertaken are described per study in the subsequent notes below.

[†], average *SD* calculated from each *SD* per preterm subgroup; [‡], average *M* and *SD* calculated from those reported per gender and birthweight category; [^], information was reported for entire sample only; [‡‡], average female ratio calculated from data reported per birthweight subgroup; [>], although the preterm subgroups were categorised according to birthweight (extremely low [ELBW] versus very low birthweight [VLBW]), GA was reported to be significantly different between these subgroups (Nagy et al., 2022); [+], average *M* and *SD* calculated from those reported per preterm subgroup; [Δ], information obtained from data at 12-month timepoint (for Case group – GA (M, SD) and female ratio (%) averaged from preterm subgroups at 12-month timepoint); [ΔΔ], GA range plus average *M* and *SD* calculated from data at 18-month timepoint.

Table 4. Characteristics and main findings of studies assessing WM ($n = 14$).

Lead author, date	Design, Methods	Case Group (PT)	Control Group (TB)	Preterm Subgroups – if avail	Measure(s)	Main Findings
Christians & Chow (2022)	Cross-sectional, between-group comparisons	n = 615 GA = 40.03, 0.01 (28–36) ~ Age = NR, NR, (7y)[c] Females = 49%‴	n = 3777 GA = 40.03, 0.01 (37–43)~ Age = NR, NR, (7y) Females = 50%‴	VPT: n = 140 (28–32) MPT: n = 475 (33–36)	Digit Span	*PT vs. TB* = Based on comparisons between TB and PT subgroups, VPT < TB only. *GA* = VPT < MPT.
Cunha et al. (2018)	Longitudinal, hierarchical modeling, regression analyses	n = 30 GA = 26.7, 1.82, (23.7–29.7) Age ! = [6 m, 9 m, 12 m, 18 m, 24 m][a] Females = 67%	n = 23 GA = 39.4, 1.1 (NR-41.1) Age ! = [6 m, 9 m, 12 m, 18 m, 24 m] Females = 43%	NR	Towel Task; Turntable Task	*PT vs. TB* = PT < TB across both tasks. *GA* = Significantly positive association between GA and success rate for Towel task only, when entire sample analysed. No significant association found when assessed in PT group only.
Domellöf et al. (2020)	Cross-sectional, between-group comparisons, regression analyses	n = 51 GA = 31.1, 3.5 (23–35) Age = 7.8y, 0.6 (7.0–8.7)[c] Females = 41%	n = 57 GA = 40.3, 0.9 (38–42) Age = 7.9y, 0.6 (6.2–8.7) Females = 42%	EPT: n = 13 (<28) VPT: n = 16 (28–32) MPT: n = 22 (33–35)	WMI	*PT vs. TB* = Based on PT group as a whole, PT < TB. Based on comparisons between TB and PT subgroups, EPT < TB only. *GA* = No significant association found.
Feng et al. (2018)	Cross-sectional, between-groups comparisons	n = 88 GA = NR, NR (28–36) Age = 8 m[a] Females = 50%	n = 88 GA = NR, NR (37–42) Age = 8 m Females = 50%	n = 9 (28–31) n = 79 (32–36)	One, Two, Three-Cup Task; Means-End Planning Task	*PT vs. TB* = PT < TB across both tasks. *GA* = No significant differences in performance between PT subgroups found.

(*Continued*)

Table 4. (Continued).

Lead author, date	Design, Methods	Case Group (PT)	Control Group (TB)	Preterm Subgroups – if avail	Measure(s)	Main Findings
Fernandez-Baizan et al. (2020)	Cross-sectional, between-group comparisons, correlational analyses	n = 88 GA = 30.12, 2.83 (NR) Age = NR, NR (5-7y)[c] Females = 45%	n = 59 GA = NR Age = NR, NR (5-7y) Females = 48%	NR	Egocentric Spatial Memory Task (part a)	*PT vs. TB* = PT < TB. *GA* = No significant association found.
Fernandez-Baizan et al. (2021)	Longitudinal, between-group comparisons, correlational analyses	n = 15 GA = 32.90, 1.91 (30–36) Age[!] = [12 m, 15, 18, 22 m][c] Females = 50%	n = 65 GA = NR Age [!] = [12 m, 15, 18, 22 m] Females = 51%	NR	Baby Mnemo Task; Hide-and-Seek Task	*PT vs. TB* = For Baby Mnemo task, PT < TB at 22 m only (on location recall and update outcomes only). For Hide-and-Seek Task, PT < TB at 12 m only (TB group displayed significantly fewer errors and greater correct answers). No further significant differences found on additional task outcomes or ages at assessments. *GA* = No associations found on either task.
Hodel et al. (2017)	Cross-sectional, between-group comparisons, correlational analyses	n = 71 GA = 35.01, 1.72 (30–36) Age = 9.13 m, 0.15 (NR);[a] 10.30, 0.43 (NR)[b] Females = 46.5%	n = 67 GA = 39.94, 1.05 (37–42) Age = 9.14 m, 0.14 (NR) Females = 50.7%	NR	A-not-B; One, Two, Three-Step Problem-Solving Task	*PT vs. TB* = No significant differences found on either task. *GA* = For A-not-B task, significantly positive associations found between GA and task outcomes (when assessed in relation to both the entire sample and PT-only group). For Problem-Solving task, no significant associations found based on entire sample, but significant association found between lower GA and slower latencies within PT-only group for Two-Step task, specifically.
Ma et al. (2022)	Cross-sectional, between-groups comparisons	n = 2065 GA = NR, NR (≤33–39) Age = 10.0y, 0.61[†] (NR)[c] Females = 47%	n = 9138 GA = NR, NR (≥40) Age = 9.9y, 0.63 (NR) Females = 48%	n = 457 (≤33) n = 555 (34–35) n = 591 (36) n = 578 (37–39)	List-sorting	*PT vs. TB* = Based on comparing T-scores across PT and TB subgroups, a main effect of group found. *GA* = Based on post-hoc pairwise differences that compared the TB group (≥40 weeks GA) with each PT subgroup, significantly lower scores found for the ≤33 weeks GA and 34–35 weeks GA subgroups only.

(*Continued*)

Table 4. (Continued).

Lead author, date	Design, Methods	Case Group (PT)	Control Group (TB)	Preterm Subgroups – if avail	Measure(s)	Main Findings
Nagy et al. (2022)	Cross-sectional, between-groups comparisons, regression analyses	n = 72 GA = 28.75, 2.21‡ (24–34) Age = 113.7 m, 3.51 (108–119)[c, ^] Females = 49.7%‡‡	n = 33 GA = NR, NR (38–41) Age = 113.7 m, 3.51 (108–119) ^ Females = NR	n = 32 (categorised as ELBW)[>] n = 40 (categorised as VLBW)[>]	Corsi Block (backward recall); WMI	*PT vs. TB* = Based on comparisons between TB and PT subgroups, ELBW < TB and VLBW < TB only, for Corsi Block task. For WMI, ELBW < TB only. *GA* = No significant effects of GA found based on either task.
O'Meagher et al. (2017)	Cross-sectional, between-group comparisons, regression analyses	n = 141 GA = 29.69, NR (23.6–32.5) Age = 49.10 m, NR (48–58)[b] Females = 50.4%	n = 77 GA = NR, NR (≥38) Age = 54.86 m, NR (48–67) Females = 41.6%	NR	BRIEF-P WM subscale (parent and teacher-rated versions)	*PT vs. TB* = PT < TB for teacher-rated version only. *GA* = Significant associations between GA and WM based on parent-rated version (in the expected direction) and teacher-rated version (not in the expected direction).
Pérez-Pereira, Martínez-López, et al. (2020)	Cross-sectional, between-group comparisons	n = 108 GA = 32.34, 0.89 [+] (26–36) Age = NR, NR (4y)[c] Females = NR	n = 34 GA = 39.68, 1.52 (>36–42) Age = NR, NR (4y) Females = NR	n = 34 (<32) [29.62, 1.46] n = 33 (32–33) [32.58, 0.50] n = 41 (34–36) [34.81, 0.72]	CHEXI WM subscale	*PT vs. TB* = No significant differences found. *GA* = No significant differences found between PT subgroups.
Pérez-Pereira, Fernández, et al. (2020)	Cross-sectional, between-groups comparisons	n = 138[≠] GA = 32.62, 2.41 [≠] (<31–36) Age = NR, NR (60 m)[c] Females = 47%	n = 43[≠] GA = 39.70, 1.48[≠] (≥37) Age = NR, NR (60 m) Females = 47%	EPT and VPT: n = 43 [≠] (<31) MPT: n = 36 [≠] (32–33) LPT: n = 58 [≠] (34–36)	Corsi Block (forward recall)	*PT vs. TB* = No significant differences found. *GA* = No significant differences found between PT subgroups.

(*Continued*)

Table 4. (Continued).

Lead author, date	Design, Methods	Case Group (PT)	Control Group (TB)	Preterm Subgroups – if avail	Measure(s)	Main Findings
Sato et al. (2019)	Cross-sectional, between-groups comparisons, correlational analyses	n = 15 GA = 29.0, 1.3 (NR) Age = 6.48y, 0.4 (NR)[c] Females = 46.7%	n = 20 GA = NR Age = 6.65y, 0.34 (NR) Females = 50%	NR	Visual MEG Task	*PT vs. TB* = No significant differences found. *GA* = No significant association found.
Stålnacke et al. (2019)	Cross-sectional, between-groups, multivariate analyses	n = 115 GA = 27.0, 2.0 (24–31) Age = NR, NR (5.5y)[c] Females = 53.9%	n = 94 GA = 39.8, 1.2 (37–42) Age = NR, NR (5.5y) Females = 53.2%	EPT: n = 73 (≤27) VPT: n = 42 (28–31)	Knox Cubes	*PT vs. TB* = PT < TB. *GA* = No significant association between GA and task outcome (based on WM as a latent variable).

Groups – Number (n) of participants; Gestational age (GA) in weeks, *M, SD* (Range); Age at assessment in months (m) or years (y), *M, SD* (Range) [[a], corrected age; [b], chronological (uncorrected) age; [c], not specified]; % of Females in group. Preterm subgroups – Number (n) of participants per subgroup (gestational age range in weeks) [if available – GA in weeks, *M, SD*]. Main findings – *PT vs. TB*, comparisons of executive function [EF] performance between preterm-born (PT) and term-born (TB) groups; *GA*, findings related to gestational age (GA) and EF; <, significantly lower EF performance in former group compared to latter group. Abbreviations – PT, preterm-born; TB, term-born; if avail, if available; GA, gestational age; NR = information was not reported; VPT, very preterm; MPT, moderately preterm; EPT, extremely preterm; WMI; Working Memory Index; ELBW, extremely low birthweight; VLBW, very low birthweight; BRIEF-P, Behavior Rating Inventory of Executive Function – Preschool Version (parent and teacher-rated); WM, working memory; CHEXI, Childhood Executive Function Inventory; LPT, late preterm; MEG, magnetoencephalography.

For the purpose of this review, where studies did not provide sufficient detail regarding participant characteristics according to preterm-born versus term-born group, this information was approximated based on the available data provided in each study. The steps undertaken are described per study in the subsequent notes below.

[~], information was reported for entire sample only (who had complete data for all study variables ($n = 2824$); [‴], average female ratio calculated from data reported per gender and preterm group; [¥], details reported separately across study (*M* = Methods; *SD* and range = Results); [!], Age data (*M, SD*, range) not reported in study, therefore ages reported per timepoint are presented; [ↄ], *SD* reported in original study according to "1.1 weeks;" [ϯ], average *SD* calculated from each *SD* per preterm subgroup; [‡], average *M* and *SD* calculated from those reported per gender and birthweight category; [^], information was reported for entire sample only; [‡‡], average female ratio calculated from data reported per birthweight subgroup; [>], although the preterm subgroups were categorised according to birthweight (extremely low [ELBW] versus very low birthweight [VLBW]), GA was reported to be significantly different between these subgroups (Nagy et al., 2022); [+], average *M* and *SD* calculated from those reported per preterm subgroup; [≠], sample sizes and GA (*M, SD*) pertaining to 60-month timepoint were not reported in original study, therefore information reported here pertains to data from 22-month timepoint.

Table 5. Characteristics and main findings of studies assessing shifting ($n = 4$).

Lead author, date	Design, Methods	Case Group (PT)	Control Group (TB)	Preterm Subgroups – if avail	Measure(s)	Main Findings
Giordano et al. (2017)	Cross-sectional, between-groups comparisons	n = 52 GA = 28.71, 2.02 (23–32) Age = 5.7y, NR (5.5–6.0)[c] Females = 48.1%	n = 52 GA = 39, 1.38 (NR) Age = 5.8y, NR (5.5–6.2) Females = 48.1%	n = 23 (23–28) n = 29 (29–32)	Flexibility	*PT vs. TB* = PT group showed significantly greater variability in performance and significantly slower reaction times than TB group. No further significant differences reported in relation to errors made. *GA* = No significant differences between PT subgroups reported.
Ma et al. (2022)	Cross-sectional, between-groups comparisons	n = 2065 GA = NR, NR (≤33–39) Age = 10.0y, 0.61[†] (NR)[c] Females = 47%	n = 9138 GA = NR, NR (≥40) Age = 9.9y, 0.63 (NR) Females = 48%	n = 457 (≤33) n = 555 (34–35) n = 591 (36) n = 578 (37–39)	Dimensional Change Card Sort	*PT vs. TB* = Based on comparing T-scores across TB and PT subgroups, no significant effect of group found. *GA* = Based on post-hoc pairwise differences that compared the TB group (≥40 weeks GA) with each PT subgroup, significantly lower scores found for the ≤33 weeks GA subgroup only.
Nagy et al. (2022)	Cross-sectional, between-groups comparisons, regression analyses	n = 72 GA = 28.75, 2.21[‡] (24–34) Age = 113.7 m, 3.51 (108–119)[c,^] Females = 49.7%[‡‡]	n = 33 GA = NR, NR (38–41) Age = 113.7 m, 3.51 (108–119)[^] Females = NR	n = 32 (categorised as ELBW)[>] n = 40 (categorised as VLBW)[>]	Wisconsin Card Sorting Test	*PT vs. TB* = No significant differences found. *GA* = No significant effects of GA found.

(*Continued*)

Table 5. (Continued).

Lead author, date	Design, Methods	Case Group (PT)	Control Group (TB)	Preterm Subgroups – if avail	Measure(s)	Main Findings
O'Meagher et al. (2017)	Cross-sectional, between-group comparisons, regression analyses	n = 141 GA = 29.69, NR (23.6–32.5) Age = 49.10 m, NR (48–58)[b] Females = 50.4%	n = 77 GA = NR, NR (≥38) Age = 54.86 m, NR (48–67) Females = 41.6%	NR	Shape School; BRIEF-P Shift subscale (parent and teacher-rated versions)	*PT vs. TB* = For Shape school, PT < TB across both raw and age-adjusted scores. No significant differences found for BRIEF-P subscales. *GA* = Significant positive association between GA and BRIEF-P (parent version) only.

Groups – Number (n) of participants; Gestational age (GA) in weeks, *M, SD* (Range); Age at assessment in months (m) or years (y), *M, SD* (Range) [[a], corrected age; [b], chronological (uncorrected) age; [c], not specified]; % of Females in group. Preterm subgroups – Number (n) of participants per subgroup (gestational age range in weeks) [if available – GA in weeks, *M, SD*]. Main findings – *PT vs. TB*, comparisons of executive function [EF] performance between preterm-born (PT) and term-born (TB) groups; *GA*, findings related to gestational age (GA) and EF; <, significantly lower EF performance in former group compared to latter group. Abbreviations – PT, preterm-born; TB, term-born; if avail, if available; GA, gestational age; NR = information was not reported; ELBW, extremely low birthweight; VLBW, very low birthweight; BRIEF-P, Behavior Rating Inventory of Executive Function – Preschool Version (parent and teacher-rated).

For the purpose of this review, where studies did not provide sufficient detail regarding participant characteristics according to preterm-born versus term-born group, this information was approximated based on the available data provided in each study. The steps undertaken are described per study in the subsequent notes below.

†, average *SD* calculated from each *SD* per preterm subgroup; ‡, average *M* and *SD* calculated from those reported per gender and birthweight category; ^, information was reported for entire sample only; ‡‡, average female ratio calculated from data reported per birthweight subgroup; >, although the preterm subgroups were categorised according to birthweight (extremely low [ELBW] versus very low birthweight [VLBW]), GA was reported to be significantly different between these subgroups (Nagy et al., 2022).

Methodological and statistical approaches

As illustrated in Tables 3–5, each paper differed in its respective methodological approach for assessing the impact of GA on EF performance. Some studies initially used between-group designs to compare EF abilities between preterm-born and term-born groups and then completed correlational or regression analyses to assess associations between GA and respective EF domains across the whole sample and/or within preterm-born participants, specifically. Conversely, other studies subdivided participants according to specified GA ranges and then compared EF abilities between the respective subgroups.

Main findings

Inhibition

Fourteen measures of inhibition were used across nine studies (as illustrated in Table 3). Of these, ten were performance-based assessments (71%) and four (29%) were rating scales. The most commonly used task was the Stroop task, for which three variations featured across three studies (Hodel et al., 2019; Nagy et al., 2022; O'Meagher et al., 2017).

Five out of nine studies (55.5%) reported a positive relationship between GA and inhibition. This was based on data retrieved from five out of 14 measures (35.7%) used across the studies to assess this EF component, of which three (60%) were performance-based assessments (Giordano et al., 2017; Ma et al., 2022; Reyes et al., 2019) and two (40%) were rating scales (O'Meagher et al., 2017; Shinya et al., 2022). Within infancy, 18- to 20-month-old infants born VPT (25–31 weeks) demonstrated lower inhibitory control compared to those born MLPT (32–36 weeks; Shinya et al., 2022). Another study reported that infants of higher GA groups at 20 months demonstrated higher proportions of inhibitory control (Reyes et al., 2019). However, since post-hoc comparisons between its five respective GA subgroups were not reported, it was not possible to determine at which GA range(s) that this effect was significant. Association analyses demonstrated that lower gestational ages were associated with lower inhibitory control within children aged 4–5 years, who were born EPT-to-MLPT (23.6–32.5 weeks; O'Meagher et al., 2017). Similarly, 5- to 6-year-old children born EPT (23–28 weeks) displayed reduced inhibitory control compared to children born VPT (29–32 weeks GA; Giordano et al., 2017). In later childhood, 10-year-old children born below 33 weeks GA, specifically, showed reduced inhibitory control compared to term-born peers (Ma et al., 2022).

Three studies (33.3%) did not report reduced inhibitory control in preterm-born participants as a function of lower GA. This included comparisons between 8-month-old infants born VPT (28–31 weeks) versus MLPT (32–36 weeks; Feng et al., 2018), and for association analyses within 4.5 to 5-year children born MLPT (32–36 weeks; Hodel et al., 2019) and 9-year-olds born EPT-to-MLPT (26–34 weeks; Nagy et al., 2022). Furthermore, although one study demonstrated a positive association between parent-rated inhibitory control (BRIEF-P) and GA at 4–5 years, this association was not found across task-based assessments that were administered in parallel (O'Meagher et al., 2017). Similarly, another study reported significantly reduced inhibitory control in 5- to 6-year-olds born EPT (23–28 weeks GA) compared to those born VPT (29–32 weeks GA) based

on a single task (Giordano et al., 2017). However, no further significant results were reported in relation to EPT vs. VPT performance for additional inhibition tasks that were administered in parallel (Giordano et al., 2017).

Unexpectedly, one study reported a negative association between GA and inhibition, whereby 4-year-old children born within a higher GA range (34–36 weeks) were reported via informant ratings as having displayed reduced inhibition capabilities compared to children of a lower GA (32–33 weeks; Pérez-Pereira, Martínez-López, et al., 2020). However, it is important to consider that this effect was small and that preterm-born children did not differ from their full-term counterparts (Sandoval et al., 2022). Therefore, it is possible that this was a spurious result.

Working memory

Twenty measures of WM were used across 14 studies (as illustrated in Table 4), of which 17 were performance-based assessments (85%) and the remaining three (15%) were rating scales. The most commonly used tasks were object retrieval-based tasks, for which eight variations featured across four studies (Cunha et al., 2018; Feng et al., 2018; Fernandez-Baizan et al., 2021; Hodel et al., 2017).

Most studies ($n = 9$) did not report an association between GA and WM (Domellöf et al., 2020; Feng et al., 2018; Fernandez-Baizan et al., 2020, 2021; Nagy et al., 2022; Pérez-Pereira, Fernández, et al., 2020; Pérez-Pereira, Martínez-López, et al., 2020; Sato et al., 2019; Stålnacke et al., 2019). The remaining five studies (35.7%) reported a positive association between GA and WM. This was based on data retrieved from six out of 20 measures (30%) used across the studies to assess this EF component, of which five of these measures (83.3%) were performance-based assessments (Christians & Chow, 2022; Cunha et al., 2018; Hodel et al., 2017; Ma et al., 2022) and one (16.7%) was a rating scale (O'Meagher et al., 2017).

Two out of the five studies demonstrated a positive association between GA and WM within infancy: at 6–24 months for infants born ELPT-to-VPT (23.7–29.7 weeks; Cunha et al., 2018), and at 9 months within both MLPT-only (30–36 weeks) and whole sample (30–42 weeks) groups (Hodel et al., 2017). However, no significant association was found between GA and WM based on a second task in the former paper (i.e., turntable task; Cunha et al., 2018). The remaining three studies demonstrated a positive association between GA and WM in early-to-middle childhood. This included at 4–5 years within children born EPT-MLPT (23.6–32.5 weeks; O'Meagher et al., 2017), at 7 years between children born VPT (28–32 weeks) compared to those born MLPT (33–36 weeks; Christians & Chow, 2022), and at 10 years based on comparisons between term-born children and those born VPT (≤33 weeks GA) or MLPT (34–35 weeks), respectively (Ma et al., 2022).

Unexpectedly, O'Meagher et al. (2017) also reported a negative association between GA and WM in parallel, based on teacher ratings (O'Meagher et al., 2017). That is, that higher scores – pertaining to higher WM difficulties – were positively associated with GA. However, WM was the only EF domain of interest that reached statistical significance based on teacher ratings (whereas all three EF components of interest reached statistical significance based on parent ratings). Thus, it is possible that this was a spurious result.

Shifting

Six measures assessing shifting abilities were used across four studies (as illustrated in Table 5), of which four were performance-based assessments (66.7%), and the remaining two were rating scales (33.3%). The most commonly used tasks were variations of the card sorting task, for which one study utilised the Dimensional Change Card Sort version (Ma et al., 2022) and another study utilised the Wisconsin Card Sorting version (Nagy et al., 2022).

Half of the studies ($n = 2$) found an association between GA and shifting. This was based on data retrieved from two out of six measures (33.3%) used across the studies to assess this EF component, of which one measure was a performance-based assessment (Ma et al., 2022) and the other was a parent rating scale (O'Meagher et al., 2017). Namely, gestational age was positively associated with shifting in children aged 4–5 years who were born EPT-to-MLPT (23.6–32.5 weeks; O'Meagher et al., 2017). The remaining study reported reduced shifting at 10 years in children born VPT (≤33 weeks GA) compared to term-born controls (Ma et al., 2022).

Discussion

Summary

The current systematic review investigated the impact of GA on EF following preterm birth, in infancy and early-to-middle childhood. Across the 18 studies included in this review, each EF component of interest was predominantly assessed using performance-based assessments rather than rating scales. The results of this systematic review suggested that GA was most prominently associated with inhibition. However, only five out of nine studies (55.5%) that assessed this component supported this conclusion. There were five out of the total 14 measures (35.7%) that assessed inhibition (i.e., three out of the total 10 performance-based assessments (30%) and two out of the total four rating scales (50%) utilised across the studies).

Moreover, most studies ($n = 9$) that assessed WM did not report finding a significant positive association with GA. Of the five studies (35.7%) that did report a positive association between GA and WM, the results stemmed from only six out of the total 20 measures (30%) that assessed WM (i.e., five out of the total 17 performance-based assessments (29.4%) and one out of the total three rating scales (33.3%) utilised across the studies).

Finally, only half ($n = 2$) of the limited number of studies available that assessed shifting reported a significant positive association with GA. Moreover, these results were based on only two out of the total six measures (33.3%) that assessed shifting (i.e., one out of the total four performance-based assessments (25%) and one out of the total two rating scales (50%) utilised across the studies).

Taken together, the overall results suggest that there is limited evidence to support that GA is associated with EF in preterm-born populations, in infancy and early-to-middle childhood. These results align with previous conclusions that have suggested no relationship between GA and EF within preterm-born children (Reveillon et al., 2018; van Houdt et al., 2019). However, the undertaking of this review has highlighted several important limitations which potentially affect the strength of this interpretation.

The first limitation relates to how participants' gestational ages were verified, classified, and reported across studies. Namely, some studies relied on retrospective recall, which may have compromised the accuracy of gestational ages obtained. This is particularly notable for studies where GA-defined subgroups were compared, as it would be unclear whether the allocation of participants within specific cutoffs was wholly accurate (Ma et al., 2022).

Furthermore, studies varied in the overall GA range that participants presented with, which limited the ability to extrapolate how the association of GA might impact EF across different GA ranges. In particular, studies with narrower GA ranges may have been likely less to have detected an association between GA and EF, due to possible limited variance. Different studies had also varied in the cut-points used to define respective GA subgroups, or they used inconsistent spacing between the cut-points of their designated subgroups (i.e., where GA ranges spanned either one, two or three weeks, respectively; Ma et al., 2022). Consequently, this limits the ability to cross-compare results according to specific gestational weeks or days. Moreover, although preterm births are broadly defined as occurring before 37 completed weeks gestation (WHO, 2022), the studies used different gestational weeks to define term-born statuses. Consequently, this resulted in variable and inconsistent margins between preterm and term-born birth statuses; thus, hampering the ability to cross-compare results based on GA with precision and consistency.

Some authors stated the reasoning behind the chosen cutoffs. For example, Stålnacke et al. (2019) excluded participants born at 32–26 week gestation because they all presented with intrauterine growth restriction and cognitive deficits that were distinct from the rest of the cohort. Moreover, due to lower numbers of participants who were born at particular GA ranges, certain gestational weeks were subsequently grouped together to ensure that each subgroup had similar numbers of participants for analyses purposes (Ma et al., 2022; Pérez-Pereira, Fernández, et al., 2020; Pérez-Pereira, Martínez-López, et al., 2020). Whilst these approaches for managing potential confounding factors are completely reasonable for individual analyses, the resultant inconsistencies in how GA subgroups and prematurity were categorised limit the ability to definitively ascertain the significance or magnitude of the impact of GA on EF in preterm-born populations, across the entire GA spectrum.

Studies also varied considerably in the level of detail and specificity that they provided when reporting GA ranges of participants. Consequently, it is difficult to ascertain the central tendency or spread of GA that was assessed within studies that have not reported this information, thus further limiting cross-comparisons of results regarding the impact of GA on EF.

A future recommendation to ensure that GA is more accurately captured would be for future studies to obtain this via independently verified records. To improve specificity and precision when reporting gestational ages, studies should include the exact ranges in addition to the mean and median GA of groups (American Psychological Association [APA], 2020). In agreement with Sandoval et al. (2022), standardised definitions of prematurity and stratifications of preterm subgroups according to GA (such as by WHO, 2022) are also recommended, to enable smoother cross-comparisons of results across studies. Moreover, utilising sufficiently detailed, standardised units of

measurement of GA (such as by O'Meagher et al., 2017, who reported both weeks and days) would also allow for more precise investigations and cross-comparisons of the impact of GA on EF.

The second limitation that hinders definitive conclusions to be drawn about the impact of GA on EF in preterm-born populations concerns participants' clinical and demographic characteristics, both in terms of their potential confounding effects on study results and their limited representativeness of wider, preterm-born populations.

Regarding confounding effects, it is important to acknowledge the complexity and heterogeneity in the wide range of potential clinical outcomes that may follow preterm birth, from neurosensory, motor, and developmental concerns (Johnson & Marlow, 2017; Moore et al., 2012; NICE, 2017), through to additional perinatal characteristics associated with lower cognitive outcomes (i.e., bronchopulmonary dysplasia (Short et al., 2003; Sriram et al., 2018), intraventricular haemorrhage and periventricular leukomalacia (Mukerji et al., 2015; Shipley et al., 2019; Volpe, 2009), retinopathy-of-prematurity (Diggikar et al., 2023) intrauterine growth restriction and being small-for-gestational age (Sacchi et al., 2020)). Moreover, systemic risk factors such as socioeconomic disadvantage not only increase the likelihood of preterm birth (Taylor et al., 2023), but they also increase risks of adverse EF outcomes across wider, term-born populations (Lean et al., 2021), which places preterm-born children from lower socio-economic backgrounds at "dual-risk" of adverse EF outcomes (Lean et al., 2021). It is therefore likely that a range of biological and environmental factors lead to a complex pattern of developmental outcomes – including EF – following preterm birth (Mulder et al., 2009). While it was beyond the scope of this review to assess the associations between these additional respective factors and EF following preterm birth, it was nevertheless important to understand how studies controlled for these confounding effects to better understand the specific impact of GA on EF.

However, at least five studies did not report having adequately controlled or adjusted for these additional factors, or they reported being unable to do so because of having undertaken non-parametric analyses (see Table 2). Therefore, it cannot be ruled out whether these factors had potentially confounded results, which might also partially account for the discrepancies in results obtained across the studies.

In parallel, studies were undertaken in predominantly Western, high-income countries, and study samples reflected homogenous demographic characteristics that were not representative of wider populations. This included participants from predominantly White ethnic backgrounds with relatively low socioeconomic risk, such as high-income (Shinya et al., 2022), accessible healthcare and education (O'Meagher et al., 2017) and high maternal educational level (Nagy et al., 2022). As these factors have been associated with higher EF capabilities within the general literature (Allan et al., 2016; Ardila et al., 2005; Cushman et al., 2022; Diamond et al., 2007; Rea-Sandin et al., 2021; Ursache et al., 2016), this means that their study results regarding the impact of GA on EF abilities may not generalise to broader groups.

Furthermore, studies were inconsistent in whether they corrected participants' ages for prematurity, with many studies not reporting their decision-making at all (see Tables 3–5). Corrected ages are advocated in clinical assessments of preterm-born infants' functional and development skills (NICE, 2017), which might help to avoid overestimating potential difficulties when otherwise based on their unadjusted ages (Aylward, 2020).

Therefore, it is possible that interpretations of results in studies that did not correct for participants' ages are biased. This also hinders cross-comparisons of results with studies that did correct for participants' ages, since their respective interpretations of participants' abilities are based on inconsistent developmental levels of attainment.

Nevertheless, the premise of correction also raises the question of whether "partial corrections" are more appropriate at higher gestational ages (Aylward, 2020). However, this would infer that the magnitude of EF capabilities across the GA spectrum are linear (which, based on this review, is yet to be determined). To facilitate the investigation of this hypothesis further, studies would ideally report both corrected and uncorrected ages of preterm-born participants. However, to the best of the author's knowledge, only one study did this (Hodel et al., 2017). Therefore, based on the studies reviewed, the impact of correction for appropriately estimating preterm-born participants' EF abilities is unclear.

Subsequent recommendations for future research include the explicit reporting and measurement of possible confounding effects, which are also accounted for in the study design and analyses. Further recommendations include stratifying key clinical and demographic characteristics to assess the relationship between GA and EF more thoroughly across broader populations. This action is crucial, not least because of medical and socio-economic risk factors which may increase a preterm-born infant or child's risks for developing poorer cognitive outcomes. Nevertheless, this action is especially pertinent, given that the latest UK birth statistics highlight significant ethnic disparities in overall proportions of preterm birth rates, with the largest increases in preterm birth rates being amongst racially minoritised – or global majority – groups (Office for National Statistics [ONS], 2023). Therefore, actioning these recommendations will enable future research to better target and address those who are most likely impacted by adverse preterm and EF outcomes, as a consequence of these intersecting systemic risk factors. It is also hoped that such actions will contribute to better understanding and support for these minoritised groups in future. Finally, studies are encouraged to report both corrected and uncorrected ages (mean, standard deviation, range) to enable smoother cross-comparisons of study results, and to facilitate a wider understanding of the impact of correction for appropriately estimating preterm-born participants' EF abilities.

The third limitation that challenges definitive conclusions to be drawn about the relationship between GA and EF in preterm-born populations concerns the extent to which aspects of study design enabled or hindered understanding about this relationship across development. Firstly, whilst many studies analysed data stemming from longitudinal cohorts, the analyses only considered a single timepoint, which therefore limited understanding of how the relationship between GA and EF in preterm-born populations may present or change across development. Moreover, it was particularly crucial to clarify the nature of this relationship within infancy as this period had been largely neglected in previous reviews. However, given the limited number of infant studies available, this hinders the ability to draw firm conclusions about the significance or magnitude of associations between GA and EF within this early developmental period. As the majority of infant studies were also cross-sectional, this further limits the ability to understand how these associations evolve over time. Thus, the scarcity of infant studies exploring this association signifies an important gap in developmental research. Consequently, a key recommendation for future research is for more studies to explore the longitudinal relationship between GA and EF from early infancy.

Nevertheless, whilst longitudinal designs are considered advantageous to examine the associations between GA and EF across development, they often face high attrition over time (Song & Chung, 2010). Yet, infant studies are particularly susceptible to higher attrition rates than studies with children or adults (Baek et al., 2023). Indeed, the three highest-powered studies in this review had all reported significant effects or associations of GA with EF following preterm birth, within infancy and early-to-middle childhood (Christians & Chow, 2022; Ma et al., 2022; Reyes et al., 2019). Subsequently, this might suggest that one possible contributing factor toward the lack of significant effects or associations that were reported by small-sample studies may have been due to Type II errors (e.g., Cunha et al., 2018) or low power in study cohorts (Linsell et al., 2015), meaning that genuine effects or associations were potentially missed.

Subsequent recommendations for future research include more consistent reporting of effect sizes and using Bayesian statistical approaches to quantify the evidence for or against null effects (Harms & Lakens, 2018; Masson, 2011). Moreover, prioritising the attainment of appropriate sample sizes could provide studies with sufficient statistical power to detect genuine and clinically relevant effects. Utilising datasets from population-based cohort studies may provide a cheaper, workable solution for investigating the relationship between GA and EF in preterm-born populations (Song & Chung, 2010). Furthermore, if sample definitions and measurements were more consistent across research, together with researchers embracing open-science practices such as sharing de-identified data (Segerstrom et al., 2023), this could also enable cohorts to be combined to attain larger sample sizes than is possible for individual studies. There are, of course, circumstances where sufficiently powered studies are particularly challenging to obtain due to the nature of the research, as discussed by Baek et al. (2023). One proposal from the wider preterm literature is to not correct for multiple comparisons, in order to (a) avoid Type II error as a result of having used restrictive significance thresholds, and (b) to account for the fact that infants do not typically respond in consistent ways across measures of higher-order cognitive skills (Hodel et al., 2017, 2019). However, this would require interpreting any subsequent results with caution. Nevertheless, taking steps to increase participants' comfort (Baek et al., 2023) and capacity for engagement during the data collection phase may also support efforts toward minimising data loss and attrition, particularly for younger-aged participants.

The fourth limitation that hinders definitive conclusions to be drawn about the relationship between GA and EF concerns the measurement and analyses of EF. For example, many studies used single tasks to assess specific EF processes. However, this approach is potentially problematic because EF components cannot be adequately captured with a single task due to task impurity (i.e., any single task requires a combination of EF and other cognitive abilities; Anderson, 2002; Miyake et al., 2000; Mulder et al., 2009). This represents a wider issue in infant research, whereby single task designs are utilised within the context of infants' limited linguistic and attentional abilities (Miller et al., 2023). Consequently, this issue of task impurity threatens the validity of any interpretations made regarding the association between GA and EF in preterm-born populations.

Second, studies varied considerably in their operationalisations and choices of measures used to assess each executive component (Stålnacke et al., 2019). As such, while some EF measures stemmed from validated test batteries, other tasks (particularly

within the infancy range) were either novel or had been adapted from existing experimental measures in the literature. As the latter tests were not standardised or validated for use with preterm-born and/or wider populations, their reliability and validity for assessing executive components accurately and consistently is unknown.

Thirdly, whilst some experimental-based measures are considered to be established EF paradigms in the literature, aspects of the original task administration and scoring procedures were amended across studies (e.g., Feng et al., 2018; Reyes et al., 2019). This issue also extended to validated subtests that stemmed from different test batteries (e.g., for further details, see Arce & McMullen, 2021), and where studies varied in their administration of tasks across different formats (i.e., digitally (Nagy et al., 2022) or testing conditions (e.g., in a MEG scanner (Sato et al., 2019) or during EEG testing (Hodel et al., 2019))). Taken together, these various aspects of methodological inconsistencies hamper efforts to cross-compare results from studies, since the test norms cannot be equated from one task to the next (Russell et al., 2005).

Additional aspects of measurement that further limit cross-comparisons to be made include differences in test scores that were used across studies (e.g., individual subtest, index, or latent variable scores to represent EF abilities). Although latent variable analyses are considered to be a useful method for overcoming such issues related to task impurity, only one study included in this review (Stålnacke et al., 2019) utilised this statistical approach. Moreover, most studies solely assessed EF using either informant-rated measures or performance-based assessments. Whilst informant-rated measures provide valuable information regarding EF across wider contexts, the incorporation of validated performance-based measures would have enabled more objective measurements of these abilities within a controlled setting. Conversely, the nature of controlled, performance-based assessments may have scaffolded children's performances and therefore masked genuine difficulties that may otherwise be seen in naturalistic, everyday environments (Anderson, 2002). Each type of assessment therefore provides "non-redundant," vital information about EF capabilities and neither type should be used interchangeably to measure EF (Mareva et al., 2024). Consequently, it is possible that these issues may have partially contributed to instances where nonsignificant results were obtained for differential EF performance based on GA, because the scores may not have provided an accurate reflection of children's genuine EF abilities.

Taken together, it is possible that inconsistent or null findings may be partially due to measurement issues, which need to be clarified in typical infant development more generally. Subsequent recommendations for future research for managing variance due to measurement error (i.e., from non-executive demands) include utilising a standardised battery of tasks and using latent variable approaches to extract "purer" latent measurements of EF abilities (Miller et al., 2023; Miyake et al., 2000) at different ages. However, this approach's requirement for large samples ($n > 100$) may limit its application to studies with smaller sample sizes (Spurk et al., 2020; Weller et al., 2020). In this instance, the formation of composite aggregate scores based on latent constructs of correlated indicators may offer reliable, alternative measures of EF (Cuevas et al., 2014). It is hoped that utilising these approaches in future research will support better understanding of EF's structural and functional organisations across development within preterm-born populations, and to enable further clarity regarding how the nature of its organisation differs as a function of GA (Miller et al., 2023). Furthermore,

methodological consensus regarding the most appropriate operationalisations and choices of tasks used to measure variables of interest, including for task administration and scoring procedures, are imperative for enabling adequate cross-comparison of results across the literature. Finally, utilising a mixture of objective and real-world assessments (both participant- and informant-based) would allow for more fine-grained analyses of EF abilities across wider contexts (Sherman & Brooks, 2010), thus potentially enhancing their clinical and research utility for detecting differential performances in EF as a function of GA.

Limitations of this review

A main limitation of this review concerns the relatively small number of studies included. It is acknowledged that having restricted the selection criteria to publications in English-language, peer-reviewed journals may have overlooked eligible studies published in grey literature or alternative languages. This approach also potentially limits the generalisability of interpretations made in this review to diverse populations. It is also acknowledged that the use of a second reviewer in the screening process might have increased the number of studies identified for inclusion (Stoll et al., 2019).

Additionally, the decision to only include case-control studies might have also resulted in papers being missed, which assessed EF within clinical presentations known to overlap with preterm-born populations (such as attention deficit/hyperactivity disorder (Franz et al., 2018)) or autism spectrum disorder (Guinchat et al., 2012; Larsson et al., 2005). Nevertheless, these papers were excluded to reduce further possible confounding effects of these clinical presentations on the study results.

Furthermore, the focus on shifting abilities may have potentially overlooked other forms of cognitive flexibility such as verbal or semantic fluency (Diamond, 2013). Nevertheless, its requirements of verbal/lexical access and EF abilities for completing fluency tasks might also imply that verbal fluency scores do not provide as pure a measure of EF compared to those covered in this review (Shao et al., 2014).

Moreover, by excluding studies which assessed EF using eye-tracking methods, it is possible that additional findings were missed that could have further clarified understanding regarding the impact of GA on EF abilities. This is particularly pertinent when considering the infancy period, in which this methodology has been utilised more prominently (de Jong et al., 2018; Downes et al., 2018; Ross-Sheehy et al., 2017). However, it is also possible that different neural circuits mediate the shifting of visual attention and voluntary eye movements compared to mechanisms involved in more executive-oriented shifting behaviors (Miyake et al., 2000; Posner & Raichle, 1994). Given that such methods still require further research for ascertaining their validity and reliability for appropriately measuring EF (de Jong et al., 2018), it was therefore not clear whether the mechanisms involved in controlling eye movements equated to those as typically involved in cognitive control (Mulder et al., 2009). Therefore, they were excluded from this review. Nevertheless, outcomes from the means-end problem-solving or planning tasks had conflated correct eye fixations and successful object reaching to produce a single EF score (Cunha et al., 2018; Feng et al., 2018; Hodel et al., 2017). Therefore, it is possible that the validity of conclusions drawn from these studies are limited, given that these scores may not reflect purely cognitive-based capabilities.

It is also acknowledged that a number of subtests were described by authors as stemming from notable cognitive batteries (such as the NEPSY), but they were subsequently excluded from this review as per the study exclusion criteria (e.g., where named subtests did not feature in any official editions of a test battery; where they had since been deleted from newer editions of test batteries due to their limited clinical utility and/or theoretical foundation; in concurrence with Reveillon et al. (2018), where measures did not fit the operational definitions of the respective EF components as described in this review).

From a scientist-practitioner perspective, these steps were intended to enhance the clinical relevance of this review based on having prioritised data from more up-to-date, clinically validated measures where possible. However, it also acknowledged that, due to the lack of a definitive developmental model of EF within infancy and childhood, conflicting perspectives remain across the literature as to how each EF component is defined, operationalised, and measured – which might also impact test selection(s). Consequently, it is acknowledged that the above exclusion criteria may be perceived by others to have resulted in some studies or subtests having been missed.

By having also excluded data within eligible studies from preterm-born participants aged 11 years and above (Ma et al., 2022; Stålnacke et al., 2019), this limited the opportunity to investigate the longitudinal association of GA with EF across a more extensive range of development. Yet, it was observed that the majority of infant studies included in this review did not provide any further longitudinal data regarding EF and GA beyond infancy. This only further highlights the paucity of longitudinal research available across the wider preterm literature, which examines the impact of GA on EF from infancy through to adolescence. It is hoped that further research will be generated in future to address this crucial limitation. This would also enable the utilisation of meta-analytic techniques to provide more thorough estimates of this relationship across infancy, childhood, and adolescence.

Several additional limitations are also considered regarding the eligible studies that were included in this review. First, by including studies that used latent variable approaches, this included subtests that had since been removed from successive test batteries in their analyses (Stålnacke et al., 2019). Consequently, this limits the clinical relevance of interpretations made based on this latent WM variable, whilst also hindering cross-comparisons of results with studies which have not incorporated this subtest.

Importantly, it was beyond the scope of this review to identify and resolve discrepancies between possible underlying rationales for different measures within a specific EF domain (Cahill, 2018). It is acknowledged that the current review focused on the three most common EF components that have featured across previous systematic reviews of EF within preterm-born infants and children, and it built upon these reviews by extending this investigation across a wider range of the infancy period. Whilst there are many theoretical models of EF across the literature, there is general agreement that the core components of EF comprise inhibition, working memory and shifting, from which higher-order skills such as planning and problem-solving are built (Diamond, 2013; Johnson & de Haan, 2015). However, it is acknowledged that there remains a lack of consensus as to how these core EF components (Miyake & Friedman, 2012; Miyake et al., 2000) are organised in developmental populations (i.e., whether they constitute dissociable components or a unitary entity; Garon et al., 2008). One consequent limitation may

be that the grouping of subtests may have conflated instruments which had operationalised EF components based on different theoretical underpinnings. Nevertheless, in concurrence with Silva et al. (2022), the inclusion of different instruments used to assess EF processes can be justified by the lack of methodological agreement in the wider literature regarding the assessments of EF abilities, particularly within the infancy and preschool periods for which no definitive developmental model of EF exists (Best & Miller, 2010).

It is also acknowledged that having focused on inhibition, WM, and shifting might have overlooked additional potential EF components of interest purported in the developmental literature, such as elements of the "hot/cool EF" model (Zelazo & Carlson, 2012). This might also include more affective aspects operating within contexts where EF or "controlled behaviour" are exercised, such as emotional regulation, but are not necessarily studied as "EF" (Miller et al., 2023). Nevertheless, it is acknowledged that these aspects might therefore represent future directions of research.

Conclusion

The current systematic review did not find compelling evidence to conclude that GA was associated with EF following preterm birth, in infancy and early-to-middle childhood. Although some studies reported a positive association or effect (mainly in relation to inhibition), overall, the vast majority of studies did not. However, this review highlighted several critical issues which hindered the ability to draw definitive conclusions about the strength of this interpretation. These issues included methodological inconsistencies in how EF was operationalised and discrepancies in the reporting and measurement of GA, which hindered the ability to cross-compare results effectively and draw definitive conclusions about their relationship. Moreover, biases in sample recruitment procedures, together with the range of heterogeneous perinatal characteristics that may typically occur following preterm birth, had potentially confounded results; thus, further limiting definitive conclusions. In addition, a key objective of this review was to build and expand on the current literature to improve understanding about the relationship between GA and EF from infancy. Yet, the paucity of infant or longitudinal studies available, together with the lack of consensus regarding the theoretical underpinnings of EF, meant that understanding this relationship across development was considerably limited. Consequently, these findings align with conclusions in the wider field regarding how methodological limitations, as well as the limited availability of studies across the literature, may have likely contributed to inconsistent or null findings (Linsell et al., 2015; Mulder et al., 2009; van Houdt et al., 2019).

The field would benefit from addressing these multiple limitations in future, to better clarify the impact of GA on EF abilities in preterm-born populations. This may also have important clinical implications regarding the understanding, identification, and management of these outcomes across infancy, childhood, and adolescence. Future research should also consider the interplay of different factors that can impact preterm development, including (but not limited to) birthweight, socio-economic status, and perinatal characteristics associated with lower cognitive outcomes (as previously described). It is hoped that future meta-analytic techniques may aid future understanding of the relationship between GA and EF, whilst also accounting for these diverse factors.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported that there is no funding associated with the work featured in this article.

ORCID

Raj Seraya Bhatoa http://orcid.org/0000-0001-9268-6836
Joe Bathelt http://orcid.org/0000-0001-5195-956X
Michelle de Haan http://orcid.org/0000-0001-5063-5475

References

Allan, J. L., McMinn, D., & Daly, M. (2016). A bidirectional relationship between executive function and health behavior: Evidence, implications, and future directions. *Frontiers in Neuroscience*, *10*, 386. https://doi.org/10.3389/fnins.2016.00386

Allotey, J., Zamora, J., Cheong-See, F., Kalidindi, M., Arroyo-Manzano, D., Asztalos, E., van der Post, J. A. M., Mol, B. W., Moore, D., Birtles, D., Khan, K. S., & Thangaratinam, S. (2018). Cognitive, motor, behavioural and academic performances of children born preterm: A meta-analysis and systematic review involving 64 061 children. *Bjog*, *125*(1), 16–25. https://doi.org/10.1111/1471-0528.14832

American Psychological Association. (2020). *Publication manual of the American psychological association* (7th ed.). https://doi.org/10.1037/0000165-000

Anderson, P. (2002). Assessment and development of executive function (EF) during childhood. *Child Neuropsychology*, *8*(2), 71–82. https://doi.org/10.1076/chin.8.2.71.8724

Arce, T., & McMullen, K. (2021). The corsi block-tapping test: Evaluating methodological practices with an eye towards modern digital frameworks. *Computers in Human Behavior Reports*, *4*, 100099. https://doi.org/10.1016/j.chbr.2021.100099

Ardila, A., Rosselli, M., Matute, E., & Guajardo, S. (2005). The influence of the parents' educational level on the development of executive functions. *Developmental Neuropsychology*, *28*(1), 539–560. https://doi.org/10.1207/s15326942dn2801_5

Aylward, G. P. (2020). Is it correct to correct for prematurity? Theoretic analysis of the Bayley-4 normative data. *Journal of Developmental & Behavioral Pediatrics*, *41*(2), 128–133. https://doi.org/10.1097/dbp.0000000000000739

Baddeley, A. D. (1997). *Human memory: Theory and practice*. Psychology Press.

Baek, S., Marques, S., Casey, K., Testerman, M., McGill, F., & Emberson, L. (2023). Attrition rate in infant fNIRS research: A meta-analysis. *Infancy*, *28*(3), 507–531. https://doi.org/10.1111/infa.12521

Baron, I. S., Litman, F. R., Ahronovich, M. D., & Baker, R. (2012). Late preterm birth: A review of medical and neuropsychological childhood outcomes. *Neuropsychology Review*, *22*(4), 438–450. https://doi.org/10.1007/s11065-012-9210-5

Best, J. R., & Miller, P. H. (2010). A developmental perspective on executive function. *Child Development*, *81*(6), 1641–1660. https://doi.org/10.1111/j.1467-8624.2010.01499.x

Bhutta, A. T., Cleves, M. A., Casey, P. H., Cradock, M. M., & Anand, K. J. (2002). Cognitive and behavioral outcomes of school-aged children who were born preterm: A meta-analysis. *JAMA*, *288*(6), 728–737. https://doi.org/10.1001/jama.288.6.728

Blair, C., & Razza, R. P. (2007). Relating effortful control, executive function, and false belief understanding to emerging math and literacy ability in kindergarten. *Child Development*, *78*(2), 647–663. https://doi.org/10.1111/j.1467-8624.2007.01019.x

Bouyssi-Kobar, M., du Plessis, A. J., McCarter, R., Brossard-Racine, M., Murnick, J., Tinkleman, L., Robertson, R. L., & Limperopoulos, C. (2016). Third trimester brain growth in preterm infants compared with in utero healthy fetuses. *Pediatrics*, *138*(5), e20161640. https://doi.org/10.1542/peds.2016-1640

Brumbaugh, J. E., Hodel, A. S., & Thomas, K. M. (2013). The impact of late preterm birth on executive function at preschool age. *American Journal of Perinatology*, *31*(04), 305–314. https://doi.org/10.1055/s-0033-1348950

Brydges, C. R., Landes, J. K., Reid, C. L., Campbell, C., French, N., & Anderson, M. (2018). Cognitive outcomes in children and adolescents born very preterm: A meta-analysis. *Developmental Medicine & Child Neurology*, *60*(5), 452–468. https://doi.org/10.1111/dmcn.13685

Cahill, J. (2018). *The relation of cognition to cerebellar function in relapsing-remitting multiple sclerosis* [Doctoral Dissertation, Royal Holloway University London]. Royal Holloway University London. https://core.ac.uk/download/pdf/195281924.pdf

Christians, J. K., & Chow, N. A. (2022). Are there sex differences in fetal growth strategies and in the long-term effects of pregnancy complications on cognitive functioning? *Journal of Developmental Origins of Health and Disease*, *13*(6), 766–778. https://doi.org/10.1017/s2040174422000204

Costeloe, K. L., Hennessy, E. M., Haider, S., Stacey, F., Marlow, N., & Draper, E. S. (2012). Short term outcomes after extreme preterm birth in England: Comparison of two birth cohorts in 1995 and 2006 (the EPICure studies). *BMJ: British Medical Journal*, *345*(dec04 3), e7976. https://doi.org/10.1136/bmj.e7976

Cuevas, K., & Bell, M. A. (2010). Developmental progression of looking and reaching performance on the A-not-B task. *Developmental Psychology*, *46*(5), 1363. https://doi.org/10.1037/a0020185

Cuevas, K., & Bell, M. A. (2014). Infant attention and early childhood executive function. *Child Development*, *85*(2), 397–404. https://doi.org/10.1111/cdev.12126

Cuevas, K., Deater-Deckard, K., Kim-Spoon, J., Wang, Z., Morasch, K. C., & Bell, M. A. (2014). A longitudinal intergenerational analysis of executive functions during early childhood. *British Journal of Developmental Psychology*, *32*(1), 50–64. https://doi.org/10.1111/bjdp.12021

Cunha, A. B., Babik, I., Ross, S. M., Logan, S. W., Galloway, J. C., Clary, E., & Lobo, M. A. (2018). Prematurity may negatively impact means-end problem solving across the first two years of life. *Research in Developmental Disabilities*, *81*, 24–36. https://doi.org/10.1016/j.ridd.2018.03.007

Cushman, G. K., West, K. B., Davis, M., LaMotte, J., Eaton, C. K., Gutierrez-Colina, A. M., & Blount, R. L. (2022). The role of executive functioning, healthcare management, and self-efficacy in college students' health-related quality of life. *Journal of American College Health*, *70*(8), 2356–2364. https://doi.org/10.1080/07448481.2020.1862128

Day, K. L. (2018). *The development of executive functions and information processing speeds in toddlers born preterm* [Doctoral Dissertation, University College London]. University College London. https://discovery.ucl.ac.uk/id/eprint/10044924/1/Kayleigh_Day_approved_PhD_e-thesis.pdf

de Haan, M. (2013). Attention and executive control. In D. Mareschal, B. Butterworth, & A. Tolmie (Eds.), *Educational neuroscience* (1st ed., pp. 529–566). John Wiley & Sons.

de Jong, M., Verhoeven, M., Hooge, I. T. C., Maingay-Visser, A., Spanjerberg, L., & van Baar, A. L. (2018). Cognitive functioning in toddlerhood: The role of gestational age, attention capacities, and maternal stimulation. *Developmental Psychology*, *54*(4), 648–662. https://doi.org/10.1037/dev0000446

Diamond, A. (1985). Development of the ability to use recall to guide action, as indicated by infants' performance on AB. *Child Development*, *56*(4), 868–883. https://doi.org/10.2307/1130099

Diamond, A. (2013). Executive functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science*, *318*(5855), 1387–1388. https://doi.org/10.1126/science.1151148

Diamond, A., & Doar, B. (1989). The performance of human infants on a measure of frontal cortex function, the delayed response task. *Developmental Psychobiology*, *22*(3), 271–294. https://doi.org/10.1002/dev.420220307

Dias, N. M., Trevisan, B. T., León, C. B. R., Prust, A. P., & Seabra, A. G. (2017). Can executive functions predict behavior in preschool children? *Psychology and Neuroscience*, *10*(4), 383. https://doi.org/10.1037/pne0000104

Diggikar, S., Gurumoorthy, P., Trif, P., Mudura, D., Nagesh, N. K., Galis, R., & Kramer, B. W. (2023). Retinopathy of prematurity and neurodevelopmental outcomes in preterm infants: A systematic review and meta-analysis. *Frontiers in Pediatrics*, *11*, 1055813. https://doi.org/10.3389/fped.2023.1055813

Domellöf, E., Johansson, A. M., Farooqi, A., Domellöf, M., & Rönnqvist, L. (2020). Risk for behavioral problems independent of cognitive functioning in children born at low gestational ages. *Frontiers in Pediatrics*, *8*, 311. https://doi.org/10.3389/fped.2020.00311

Downes, M., Kelly, D., Day, K., Marlow, N., & de Haan, M. (2018). Visual attention control differences in 12-month-old preterm infants. *Infant Behavior and Development*, *50*, 180–188. https://doi.org/10.1016/j.infbeh.2018.01.002

Doyle, L. W., Anderson, P. J., Battin, M., Bowen, J. R., Brown, N., Callanan, C., & Woodward, L. J. (2014). Long term follow up of high risk children: Who, why and how? *BMC Pediatrics*, *14*(1), 279. https://doi.org/10.1186/1471-2431-14-279

du Plessis, A. J., & Volpe, J. J. (2018). Chapter 1 - neural tube development. In J. J. Volpe, T. E. Inder, B. T. Darras, L. S. de Vries, A. J. du Plessis, J. J. Neil, & J. M. Perlman (Eds.), *Volpe's neurology of the newborn* (6th ed., pp. 3–33.e38). Elsevier. https://doi.org/10.1016/B978-0-323-42876-7.00001-6

Duncan, A. F., Gerner, G. J., Neel, M. L., Burton, V. J., Byrne, R., & Warschausky, S. (2024). Interventions to improve executive functions in children aged 3 years and under: A systematic review. *Child: Care, Health and Development*, *50*(4), e13298. https://doi.org/10.1111/cch.13298

Effective Public Health Practice Project. (2002). *Quality assessment tool for quantitative studies*. https://www.ephpp.ca/PDF/Quality%20Assessment%20Tool_2010_2.pdf

Eryigit Madzwamuse, S., Baumann, N., Jaekel, J., Bartmann, P., & Wolke, D. (2015). Neuro-cognitive performance of very preterm or very low birth weight adults at 26 years. *Journal of Child Psychology and Psychiatry*, *56*(8), 857–864. https://doi.org/10.1111/jcpp.12358

Feng, Y., Zhou, H., Zhang, Y., Perkins, A., Wang, Y., & Sun, J. (2018). Comparison in executive function in Chinese preterm and full-term infants at eight months. *Frontiers of Medicine*, *12*(2), 164–173. https://doi.org/10.1007/s11684-017-0540-9

Fernandez-Baizan, C., Alcántara-Canabal, L., Solis, G., & Mendez, M. (2020). Development of egocentric and allocentric spatial orientation abilities in children born preterm with very low birth weight. *Early Human Development*, *141*, 104947. https://doi.org/10.1016/j.earlhumdev.2019.104947

Fernandez-Baizan, C., Caunedo-Jimenez, M., Martinez, J., Arias, J., Mendez, M., & Solis, G. (2021). Development of visuospatial memory in preterm infants: A new paradigm to assess short-term and working memory. *Child Neuropsychology*, *27*(3), 296–316. https://doi.org/10.1080/09297049.2020.1847264

Franz, A. P., Bolat, G. U., Bolat, H., Matijasevich, A., Santos, I. S., Silveira, R. C., & Moreira-Maia, C. R. (2018). Attention-deficit/hyperactivity disorder and very preterm/very low birth weight: A meta-analysis. *Pediatrics*, *141*(1). https://doi.org/10.1542/peds.2017-1645

Garon, N., Bryson, S. E., & Smith, I. M. (2008). Executive function in preschoolers: A review using an integrative framework. *Psychological Bulletin*, *134*(1), 31. https://doi.org/10.1037/0033-2909.134.1.31

Gathercole, S. E., & Alloway, T. P. (2006). Practitioner review: Short-term and working memory impairments in neurodevelopmental disorders: Diagnosis and remedial support. *Journal of Child Psychology and Psychiatry*, *47*(1), 4–15. https://doi.org/10.1111/j.1469-7610.2005.01446.x

Giordano, V., Fuiko, R., Leiss, U., Brandstetter, S., Hayde, M., Bartha-Doering, E., & Weiler, L. J. (2017). Differences in attentional functioning between preterm and full-term children underline

the importance of new neuropsychological detection techniques. *Acta Paediatrica*, *106*(4), 601–611. https://doi.org/10.1111/apa.13723

Guinchat, V., Thorsen, P., Laurent, C., Cans, C., Bodeau, N., & Cohen, D. (2012). Pre-, peri-and neonatal risk factors for autism. *Acta Obstetricia Et Gynecologica Scandinavica*, *91*(3), 287–300. https://doi.org/10.1111/j.1600-0412.2011.01325.x

Harms, C., & Lakens, D. (2018). Making'null effects' informative: Statistical techniques and inferential frameworks. *Journal of Clinical and Translational Research*, *3*(Suppl 2), 382. https://doi.org/10.18053/jctres.03.2017S2.007

Hodel, A. S., Brumbaugh, J. E., Hunt, R. H., Van den Heuvel, S. E., Wiltgen, A. M., & Thomas, K. M. (2019). Individual differences in ERP measures of executive function in early childhood: Relation to low-risk preterm birth and parent-reported behavior. *Child Neuropsychology*, *25*(7), 914–942. https://doi.org/10.1080/09297049.2018.1540690

Hodel, A. S., Senich, K. L., Jokinen, C., Sasson, O., Morris, A. R., & Thomas, K. M. (2017). Early executive function differences in infants born moderate-to-late preterm. *Early Human Development*, *113*, 23–30. https://doi.org/10.1016/j.earlhumdev.2017.07.007

Holmboe, K., Bonneville-Roussy, A., Csibra, G., & Johnson, M. H. (2018). Longitudinal development of attention and inhibitory control during the first year of life. *Developmental Science*, *21* (6), e12690. https://doi.org/10.1111/desc.12690

Holmes, C. J., Kim-Spoon, J., & Deater-Deckard, K. (2016). Linking executive function and peer problems from early childhood through middle adolescence. *Journal of Abnormal Child Psychology*, *44*(1), 31–42. https://doi.org/10.1007/s10802-015-0044-5

Johnson, M. H., & de Haan, M. D. (2015). *Developmental cognitive neuroscience: An introduction*. John Wiley & Sons.

Johnson, S., & Marlow, N. (2017). Early and long-term outcome of infants born extremely preterm. *Archives of Disease in Childhood*, *102*(1), 97–102. https://doi.org/10.1136/archdischild-2015-309581

Korkman, M., Kirk, U., & Kemp, S. (1998). *NESPY: A developmental neuropsychological assessment*. The Psychological Corporation.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-Second edition (NEPSY-II)*. https://doi.org/10.1037/t15125-000

Larsson, H. J., Eaton, W. W., Madsen, K. M., Vestergaard, M., Olesen, A. V., Agerbo, E., & Mortensen, P. B. (2005). Risk factors for autism: Perinatal factors, parental psychiatric history, and socioeconomic status. *American Journal of Epidemiology*, *161*(10), 916–925. https://doi.org/10.1093/aje/kwi123

Lawn, J. E., Ohuma, E. O., Bradley, E., Idueta, L. S., Hazel, E., Okwaraji, Y. B., & Lee, A. C. (2023). Small babies, big risks: Global estimates of prevalence and mortality for vulnerable newborns to accelerate change and improve counting. *Lancet*, *401*(10389), 1707–1719. https://doi.org/10.1016/s0140-6736(23)00522-6

Lean, R. E., Gerstein, E. D., Smyser, T. A., Smyser, C. D., & Rogers, C. E. (2021). Socioeconomic disadvantage and parental mood/affective problems links negative parenting and executive dysfunction in children born very preterm. *Development & Psychopathology*, *35*(3), 1–16. https://doi.org/10.1017/s0954579421000961

Lichtenberger, E. O., & Kaufman, A. S. (2004). *Essentials of WPPSI-III assessment*. John Wiley & Sons Inc.

Linsell, L., Malouf, R., Morris, J., Kurinczuk, J. J., & Marlow, N. (2015). Prognostic factors for poor cognitive development in children born very preterm or with very low birth weight: A systematic review. *JAMA Pediatrics*, *169*(12), 1162–1172. https://doi.org/10.1001/jamapediatrics.2015.2175

Loe, I. M., Heller, N. A., & Chatav, M. (2019). Behavior problems and executive function impairments in preterm compared to full term preschoolers. *Early Human Development*, *130*, 87–95. https://doi.org/10.1016/j.earlhumdev.2019.01.014

Ma, Q., Wang, H., Rolls, E. T., Xiang, S., Li, J., Li, Y., & Li, F. (2022). Lower gestational age is associated with lower cortical volume and cognitive and educational performance in adolescence. *BMC Medicine*, *20*(1), 424. https://doi.org/10.1186/s12916-022-02627-3

Marcovitch, S., & Zelazo, P. D. (1999). The A-not-B error: Results from a logistic meta-analysis. *Child Development*, *70*(6), 1297–1313. https://doi.org/10.1111/1467-8624.00095

Mareva, S., Team, C. A. L. M., & Holmes, J. (2024). Mapping neurodevelopmental diversity in executive function. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *172*, 204–221. https://doi.org/10.1016/j.cortex.2023.11.021

Masson, M. E. (2011). A tutorial on a practical bayesian alternative to null-hypothesis significance testing. *Behavior Research Methods*, *43*(3), 679–690. https://doi.org/10.3758/s13428-010-0049-5

Matthews, L. G., Walsh, B. H., Knutsen, C., Neil, J. J., Smyser, C. D., Rogers, C. E., & Inder, T. E. (2018). Brain growth in the NICU: Critical periods of tissue-specific expansion. *Pediatric Research*, *83*(5), 976–981. https://doi.org/10.1038/pr.2018.4

Miller, S. E., Avila, B. N., & Reavis, R. D. (2020). Thoughtful friends: Executive function relates to social problem solving and friendship quality in middle childhood. *Journal of Genetic Psychology*, *181*(2–3), 78–94. https://doi.org/10.1080/00221325.2020.1719024

Miller, S. E., Galvagno, L. G., & Elgier, Á. (2023). Universality and context-specificity in early executive function development. *Infant Behavior and Development*, *71*, 101841. https://doi.org/10.1016/j.infbeh.2023.101841

Miller, S. P., & Ferriero, D. M. (2009). From selective vulnerability to connectivity: Insights from newborn brain imaging. *Trends in Neurosciences*, *32*(9), 496–505. https://doi.org/10.1016/j.tins.2009.05.010

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. *The Current Directions in Psychological Science*, *21*(1), 8–14. https://doi.org/10.1177/0963721411429458

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Moore, T., Hennessy, E. M., Myles, J., Johnson, S. J., Draper, E. S., Costeloe, K. L., & Marlow, N. (2012). Neurological and developmental outcome in extremely preterm children born in England in 1995 and 2006: The EPICure studies. *BMJ-British Medical Journal*, *345*, (3), e7961–e7961. https://doi.org/10.1136/bmj.e7961

Mukerji, A., Shah, V., & Shah, P. S. (2015). Periventricular/Intraventricular hemorrhage and neurodevelopmental outcomes: A meta-analysis. *Pediatrics*, *136*(6), 1132–1143. https://doi.org/10.1542/peds.2015-0944

Mulder, H., Pitchford, N. J., Hagger, M. S., & Marlow, N. (2009). Development of executive function and attention in preterm children: A systematic review. *Developmental Neuropsychology*, *34*(4), 393–421. https://doi.org/10.1080/87565640902964524

Nagy, A., Kalmár, M., Beke, A. M., Gráf, R., & Horváth, E. (2022). Intelligence and executive function of school-age preterm children in function of birth weight and perinatal complication. *Applied Neuropsychology: Child*, *11*(3), 400–411. https://doi.org/10.1080/21622965.2020.1866571

National Institute for Health and Care Excellence. (2017). *Developmental follow-up of children and young people born preterm* [NICE Guideline No. 72]. https://www.nice.org.uk/guidance/ng72

Ni, Y., Johnson, S., Marlow, N., & Wolke, D. (2022). Reduced health-related quality of life in children born extremely preterm in 2006 compared with 1995: The EPICure studies. *Archives of Disease in Childhood - Fetal & Neonatal Edition*, *107*(4), 408. https://doi.org/10.1136/archdischild-2021-322888

Nigg, J. T. (2000). On inhibition/disinhibition in developmental psychopathology: Views from cognitive and personality psychology and a working inhibition taxonomy. *Psychological Bulletin*, *126*(2), 220. https://doi.org/10.1037/0033-2909.126.2.220

Nigg, J. T. (2017). Annual research review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *58*(4), 361–383. https://doi.org/10.1111/jcpp.12675

O'Meagher, S., Kemp, N., Norris, K., Anderson, P., & Skilbeck, C. (2017). Risk factors for executive function difficulties in preschool and early school-age preterm children. *Acta Paediatrica*, *106* (9), 1468–1473. https://doi.org/10.1111/apa.13915

Office for National Statistics. (2023). *Statistical bulletin, birth characteristics in England and Wales: 2021*. https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/livebirths/bulletins/birthcharacteristicsinenglandandwales/2021

Orchinik, L. J., Taylor, H. G., Espy, K. A., Minich, N., Klein, N., Sheffield, T., & Hack, M. (2011). Cognitive outcomes for extremely preterm/extremely low birth weight children in kindergarten. *Journal of the International Neuropsychological Society*, *17*(6), 1067–1079. https://doi.org/10.1017/s135561771100107x

Pelphrey, K. A., & Reznick, J. S. (2004). Working memory in infancy. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 31, pp. 173–227). Academic Press. https://doi.org/10.1016/S0065-2407(03)31005-5

Pelphrey, K. A., Reznick, J. S., Davis Goldman, B., Sasson, N., Morrow, J., Donahoe, A., & Hodgson, K. (2004). Development of visuospatial short-term memory in the second half of the 1st year. *Developmental Psychology*, *40*(5), 836–851. https://doi.org/10.1037/0012-1649.40.5.836

Pérez-Pereira, M., Fernández, M. P., Gómez-Taibo, M. L., Martínez-López, Z., & Arce, C. (2020). A follow-up study of cognitive development in low risk preterm children. *International Journal of Environmental Research and Public Health*, *17*(7), 2380. https://doi.org/10.3390/ijerph17072380

Pérez-Pereira, M., Martínez-López, Z., & Maneiro, L. (2020). Longitudinal relationships between reading abilities, phonological awareness, language abilities and executive functions: Comparison of low risk preterm and full-term children. *Frontiers in Psychology*, *11*, 468. https://doi.org/10.3389/fpsyg.2020.00468

Piaget, J. (1954). *The construction of reality in the child* (M. Cook, Trans.). Basic Books. https://doi.org/10.1037/11168-000

Pierrat, V., Marchand-Martin, L., Arnaud, C., Kaminski, M., Resche-Rigon, M., Lebeaux, C., & Marret, S. (2017). Neurodevelopmental outcome at 2 years for preterm children born at 22 to 34 weeks' gestation in France in 2011: EPIPAGE-2 cohort study. *BMJ*, *358*, j3448. https://doi.org/10.1136/bmj.j3448

Posner, M., & Raichle, M. (1994). *Images of mind*. Scientific American Library.

Potharst, E. S., Houtzager, B. A., Van Sonderen, L., Tamminga, P., Kok, J. H., Last, B. F., & Van Wassenaer, A. G. (2012). Prediction of cognitive abilities at the age of 5 years using developmental follow-up assessments at the age of 2 and 3 years in very preterm children. *Developmental Medicine & Child Neurology*, *54*(3), 240–246. https://doi.org/10.1111/j.1469-8749.2011.04181.x

Rea-Sandin, G., Korous, K. M., & Causadias, J. M. (2021). A systematic review and meta-analysis of racial/ethnic differences and similarities in executive function performance in the United States. *Neuropsychology*, *35*(2), 141. https://doi.org/10.1037/neu0000715

Reveillon, M., Huppi, P. S., & Barisnikov, K. (2018). Inhibition difficulties in preterm children: Developmental delay or persistent deficit? *Child Neuropsychology*, *24*(6), 734–762. https://doi.org/10.1080/09297049.2017.1294665

Reyes, L. M., Jaekel, J., Heuser, K. M., & Wolke, D. (2019). Developmental cascades of social inhibition and friendships in preterm and full-term children. *Infant and Child Development*, *28* (6), Article e2165. https://doi.org/10.1002/icd.2165

Riggs, N. R., Jahromi, L. B., Razza, R. P., Dillworth-Bart, J. E., & Mueller, U. (2006). Executive function and the promotion of social-emotional competence. *Journal of Applied Developmental Psychology*, *27*(4), 300–309. https://doi.org/10.1016/j.appdev.2006.04.002

Roberts, D., Brown, J., Medley, N., & Dalziel, S. R. (2017). Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth. *Cochrane Database of Systematic Reviews*, (3). https://doi.org/10.1002/14651858.cd004454.pub3

Ross-Sheehy, S., Perone, S., Macek, K. L., & Eschman, B. (2017). Visual orienting and attention deficits in 5-and 10-month-old preterm infants. *Infant Behavior and Development*, *46*, 80–90. https://doi.org/10.1016/j.infbeh.2016.12.004

Russell, E. W., Russell, S. L. K., & Hill, B. D. (2005). The fundamental psychometric status of neuropsychological batteries. *Archives of Clinical Neuropsychology*, *20*(6), 785–794. https://doi.org/10.1016/j.acn.2005.05.001

Sacchi, C., Marino, C., Nosarti, C., Vieno, A., Visentin, S., & Simonelli, A. (2020). Association of intrauterine growth restriction and small for gestational age status with childhood cognitive outcomes: A systematic review and meta-analysis. *JAMA Pediatrics*, *174*(8), 772–781. https://doi.org/10.1001/jamapediatrics.2020.1097

Sampaio, R., & Truwit, C. (2001). Myelination in the developing human brain. In C. A. Nelson & M. Luciana (Eds.), *Handbook of developmental cognitive neuroscience* (pp. 35–44). MIT Press.

Sanderson, S., Tatt, I. D., & Higgins, J. P. (2007). Tools for assessing quality and susceptibility to bias in observational studies in epidemiology: A systematic review and annotated bibliography. *International Journal of Epidemiology*, *36*(3), 666–676. https://doi.org/10.1093/ije/dym018

Sandoval, C. C., Gaspardo, C. M., & Linhares, M. B. M. (2022). The impact of preterm birth on the executive functioning of preschool children: A systematic review. *Applied Neuropsychology: Child*, *11*(4), 873–890. https://doi.org/10.1080/21622965.2021.1915145

Sato, J., Mossad, S. I., Wong, S. M., Hunt, B. A. E., Dunkley, B. T., Urbain, C., & Taylor, M. J. (2019). Spectral slowing is associated with working memory performance in children born very preterm. *Scientific Reports*, *9*(1), Article 15757 https://doi.org/10.1038/s41598-019-52219-0

Schnider, B., Disselhoff, V., Held, U., Latal, B., Hagmann, C. F., & Wehrle, F. M. (2020). Executive function deficits mediate the association between very preterm birth and behavioral problems at school-age. *Early Human Development*, *146*, 105076. https://doi.org/10.1016/j.earlhumdev.2020.105076

Segerstrom, S. C., Diefenbach, M. A., Hamilton, K., O'Connor, D. B., & Tomiyama, J. A. (2023). Open science in health psychology and behavioral medicine: A statement from the behavioral medicine research council. *Health Psychology*, *42*(5), 287–298. https://doi.org/10.1037/hea0001236

Shao, Z., Janse, E., Visser, K., & Meyer, A. S. (2014). What do verbal fluency tasks measure? Predictors of verbal fluency performance in older adults. *Frontiers in Psychology*, *5*, 772. https://doi.org/10.3389/fpsyg.2014.00772

Sheehan, J. C., Kerns, K. A., & Müller, U. (2017). The effect of task complexity on planning in preterm-born children. *The Clinical Neuropsychologist*, *31*(2), 438–458. https://doi.org/10.1080/13854046.2016.1244248

Sherman, E. M., & Brooks, B. L. (2010). Behavior rating inventory of executive function-preschool version (BRIEF-P): Test review and clinical guidelines for use. *Child Neuropsychology*, *16*(5), 503–519. https://doi.org/10.1080/09297041003679344

Shinya, Y., Kawai, M., Niwa, F., Kanakogi, Y., Imafuku, M., & Myowa, M. (2022). Cognitive flexibility in 12-month-old preterm and term infants is associated with neurobehavioural development in 18-month-olds. *Scientific Reports*, *12*(1), Article 3. https://doi.org/10.1038/s41598-021-04194-8

Shipley, L., Gyorkos, T., Dorling, J., Tata, L. J., Szatkowski, L., & Sharkey, D. (2019). Risk of severe intraventricular hemorrhage in the first week of life in preterm infants transported before 72 hours of age. *Pediatric Critical Care Medicine*, *20*(7), 638–644. https://doi.org/10.1097/pcc.0000000000001937

Short, E. J., Klein, N. K., Lewis, B. A., Fulton, S., Eisengart, S., Kercsmar, C., & Singer, L. T. (2003). Cognitive and academic consequences of bronchopulmonary dysplasia and very low birth weight: 8-year-old outcomes. *Pediatrics*, *112*(5), e359–e359. https://doi.org/10.1542/peds.112.5.e359

Silva, C., Sousa-Gomes, V., Fávero, M., Oliveira-Lopes, S., Merendeiro, C. S., Oliveira, J., & Moreira, D. (2022). Assessment of preschool-age executive functions: A systematic review. *Clinical Psychology & Psychotherapy*, *29*(4), 1374–1391. https://doi.org/10.1002/cpp.2718

Song, J. W., & Chung, K. C. (2010). Observational studies: Cohort and case-control studies. *Plastic and Reconstructive Surgery*, *126*(6), 2234. https://doi.org/10.1097/PRS.0b013e3181f44abc

Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. *Journal of Vocational Behavior*, *120*, 103445. https://doi.org/10.1016/j.jvb.2020.103445

Sriram, S., Schreiber, M. D., Msall, M. E., Kuban, K. C. K., Joseph, R. M., Tm, O. S., & Leviton, A. (2018). Cognitive development and quality of life associated with BPD in 10-year-olds born preterm. *Pediatrics*, *141*(6). https://doi.org/10.1542/peds.2017-2719

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology*, *59*(4), 745–759. https://doi.org/10.1080/17470210500162854

Stålnacke, J., Lundequist, A., Böhm, B., Forssberg, H., & Smedler, A. C. (2019). A longitudinal model of executive function development from birth through adolescence in children born very or extremely preterm. *Child Neuropsychology*, *25*(3), 318–335. https://doi.org/10.1080/09297049.2018.1477928

Stiles, J., & Jernigan, T. L. (2010). The basics of brain development. *Neuropsychology Review*, *20*(4), 327–348. https://doi.org/10.1007/s11065-010-9148-4

Stoll, C. R., Izadi, S., Fowler, S., Green, P., Suls, J., & Colditz, G. A. (2019). The value of a second reviewer for study selection in systematic reviews. *Research Synthesis Methods*, *10*(4), 539–545. https://doi.org/10.1002/jrsm.1369

Tau, G. Z., & Peterson, B. S. (2010). Normal development of brain circuits. *Neuropsychopharmacology Reviews*, *35*(1), 147–168. https://doi.org/10.1038/npp.2009.115

Taylor, H. G., & Clark, C. A. (2016). Executive function in children born preterm: Risk factors and implications for outcome. *Seminars in Perinatology*, *40*(8), 520–529. https://doi.org/10.1053/j.semperi.2016.09.004

Taylor, R. L., Rogers, C. E., Smyser, C. D., & Barch, D. M. (2023). Associations between preterm birth, inhibitory control-implicated brain regions and tracts, and inhibitory control task performance in children: Consideration of socioeconomic context. *Child Psychiatry and Human Development*, 1–15. https://doi.org/10.1007/s10578-023-01531-y

Ursache, A., Noble, K. G., Pediatric Imaging, N., & Study, G. (2016). Socioeconomic status, white matter, and executive function in children. *Brain and Behavior*, *6*(10), e00531. https://doi.org/10.1002/brb3.531

van Houdt, C. A., Oosterlaan, J., van Wassenaer-Leemhuis, A. G., van Kaam, A. H., & Aarnoudse-Moens, C. S. H. (2019). Executive function deficits in children born preterm or at low birthweight: A meta-analysis. *Developmental Medicine & Child Neurology*, *61*(9), 1015–1024. https://doi.org/10.1111/dmcn.14213

Victorian Infant Collaborative Study Group. (1997). Outcome at 2 years of children 23-27 weeks' gestation born in Victoria in 1991–92. *Journal of Paediatrics and Child Health*, *33*(2), 161–165. https://doi.org/10.1111/j.1440-1754.1997.tb01021.x

Vo Van, P., Alison, M., Morel, B., Beck, J., Bednarek, N., Hertz-Pannier, L., & Loron, G. (2022). Advanced brain imaging in preterm infants: A narrative review of microstructural and connectomic disruption. *Children (Basel)*, *9*(3), 356. https://doi.org/10.3390/children9030356

Volpe, J. J. (2009). Brain injury in premature infants: A complex amalgam of destructive and developmental disturbances. *Lancet Neurology*, *8*(1), 110–124. https://doi.org/10.1016/S1474-4422(08)70294-1

Wechsler, D. (2002). *The Wechsler preschool and primary scale of intelligence* (3rd ed.). The Psychological Corporation.

Wechsler, D. (2003). *Wechsler intelligence scale for children* (4th ed.). The Psychological Corporation.

Weller, B. E., Bowen, N. K., & Faubert, S. J. (2020). Latent class analysis: A guide to best practice. *The Journal of Black Psychology*, *46*(4), 287–311. https://doi.org/10.1177/0095798420930932

Williams, P. E., Weiss, L., & Rolfhus, P. D. E. (2003). *WISC-IV technical report #1 theoretical model and test blueprint*. The Psychological Corporation.

Wong, H. S., & Edwards, P. (2013). Nature or nurture: A systematic review of the effect of socio-economic status on the developmental and cognitive outcomes of children born preterm. *Maternal and Child Health Journal*, *17*(9), 1689–1700. https://doi.org/10.1007/s10995-012-1183-8

World Health Organization. (2022). *Preterm birth* [Fact sheet]. https://www.who.int/news-room/fact-sheets/detail/preterm-birth

World Health Organization [WHO]. (2012). *Born Too Soon. The Global Action Report on Preterm Birth*. https://www.who.int/publications/i/item/9789241503433

Yakovlev, P. I., & Lecours, A. R. (1967). The myelogenetic cycles of regional maturation of the brain. In A. Minowski (Ed.), *Regional development of the brain in early life* (pp. 3–70). Blackwell.

Younge, N., Goldstein, R. F., Bann, C. M., Hintz, S. R., Patel, R. M., Smith, P. B., & Cotten, C. M. (2017). Survival and neurodevelopmental outcomes among periviable infants. *New England Journal of Medicine*, *376*(7), 617–628. https://doi.org/10.1056/NEJMoa1605566

Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child Development Perspectives*, *6*(4), 354–360. https://doi.org/10.1111/j.1750-8606.2012.00246.x