Received: 26 June 2020 | Revised: 28 July 2020 | Accepted: 14 August 2020

DOI: 10.1111/ejn.14946

REVIEW ARTICLE

EJN European Journal of Neuroscience FENS WILEY

A systematic review and meta-analysis of behavioural sex differences in executive control

Alexandra Gaillard[1] | **Daniel J. Fehring**[2,3] | **Susan L. Rossell**[1,4]

[1]Centre for Mental Health, Faculty of Health, Arts and Design, Swinburne University of Technology, Hawthorn, VIC, Australia

[2]Cognitive Neuroscience Laboratory, Monash Biomedicine Discovery Institute, Department of Physiology, Monash University, Clayton, VIC, Australia

[3]ARC Centre of Excellence in Integrative Brain Function, Monash University, Clayton, VIC, Australia

[4]Psychiatry, St Vincent's Hospital, Melbourne, VIC, Australia

Correspondence
Alexandra Gaillard, Centre for Mental Health, Faculty of Health, Arts and Design, Swinburne University of Technology, Hawthorn, VIC, Australia.
Email: agaillard@swin.edu.au

Funding information
ARC Centre of Excellence in Integrative Brain Function, Grant/Award Number: 2016; National Health and Medical Research Council, Grant/Award Number: GNT1154651

Abstract

Literature investigating whether an individuals' sex affects their executive control abilities and performance on cognitive tasks in a normative population has been contradictory and inconclusive. Using meta-analytic procedures (abiding by PRISMA guidelines), this study attempts to identify the magnitude of behavioural sex differences in three prominent executive control domains of cognitive set-shifting, performance monitoring, and response inhibition. PubMed, Web of Science, and Scopus were systematically searched. Across 46 included studies, a total of 1988 females and 1884 males were included in the analysis. Overall, males and females did not differ on performance in any of the three domains of performance monitoring, response inhibition, or cognitive set-shifting. Task-specific sex differences were observed in the domains of performance monitoring, in the CANTAB Spatial Working Memory task—males scored statistically higher than females (Hedges' $g = -0.60$), and response inhibition, in the Delay Discounting task—females scored statistically higher than males (Hedges' $g = 0.64$). While the meta-analysis did not detect overall behavioural sex differences in executive control, significant heterogeneity and task-specific sex differences were found. To further understand sex differences within these specific tasks and domains, future research must better control for age and sex hormone levels.

KEYWORDS

cognition, cognitive flexibility, response inhibition, sex

1 | INTRODUCTION

For many years there has been an enduring and considerable degree of interest on whether sex differences exist in cognitive processes, with executive control being one such area of study of particular importance. Some evidence suggests that an individuals' sex may affect their executive control abilities due to sex-linked neurobiological differences (Cotto et al., 2010),

Abbreviations: ANT, attention network test; CI, confidence interval; COMT, catechol-o-methyltransferase; CPT, continuous performance test; DD, delay discounting; DS, digit span; F, females; GABA, gamma-aminobutyric acid; IED, intra-/extra- dimensional set-shift; IGT, Iowa gambling task; JBI, Joanna Briggs Institute; L.L., lower limit; M, males; MSN, medium spiny neurons; N.S., not significant; PFC, prefrontal cortex; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; RT, response time; *SD*, standard deviation; SOC, stockings of Cambridge; SST, stop signal task; SWM, spatial working memory; TMT, trail making test; TOL, tower of London; U.L., upper limit; VTA, ventral tegmental area; WCST, Wisconsin Card Sorting Test.

including differences in circulating gonadal hormone concentrations, such as oestrogen (Sun, Walker, Dean, van den Buuse, & Gogos, 2016). Oestrogen has been found to effect neurotransmitters within the prefrontal cortex (PFC; Keenan, Ezzat, Ginsburg, & Moore, 2001), which is critical to the integrity of executive control functions (Mansouri, Buckley, Fehring, & Tanaka, 2020; Mansouri, Buckley, & Tanaka, 2014). Predominantly, these effects have been found to be protective in females; however, the effects of oestrogen on these neurobiological differences been found to be susceptible to fluctuations across endogenous hormone cycle (Sun et al., 2016), as well as age, losing some of the protective effects post-menopause, between 45 and 50 years of age (Shanmugan & Epperson, 2014).

Executive control refers to a set of higher order cognitive abilities that enable the flexible performance of everyday tasks (Cummings, 1993; Miller & Cohen, 2001). These abilities enable us to exclude irrelevant and distracting stimuli, as well as updating away from those which were previously relevant, cumulatively facilitating goal-directed behaviour (Bonte, Flemming, & Fagot, 2011; Mansouri, Tanaka, & Buckley, 2009; Verbruggen & Logan, 2008). While an exact set of executive control functions has not been agreed upon, it is widely thought that there are at least three prominent aspects of executive control, including performance monitoring, response inhibition, and cognitive set-shifting (Miyake et al., 2000).

Performance monitoring refers to the ability to monitor for adverse or unexpected outcomes within tasks, including conflict between possible options, and the commission of errors, and update ones working memory to be able to implement appropriate behavioural adjustments in the future (Egner & Hirsch, 2005; Miyake et al., 2000). Moreover response inhibition refers to the ability to suppress prepotent or ongoing behaviours that are no longer appropriate or relevant (Verbruggen & Logan, 2008). Response inhibition contributes to cognitive flexibility through suppression of task-irrelevant behaviours, thereby strengthening task-relevant behaviours to enable goal-directed behaviour (Verbruggen & Logan, 2008). Finally, cognitive set-shifting, also known as attention switching, refers to the ability to switch between mental sets depending on task requirements using working memory (Miyake et al., 2000).

The literature on sex differences in a healthy population is contradictory and inconclusive yet remains an important research topic given increasing evidence that there are sex-dependent risk factors to the development of some neuropsychiatric illnesses. These inconsistencies are observed within the literature on an array of tasks, with some studies showing sex differences are present (Geary, Saults, Liu, & Hoard, 2000; Hyde, 1981) whereas others showing none (Li et al., 2009; Mulvihill, Skilling, & Vogel-Sprott, 1997). However, to date, no meta-analysis exists on sex differences within the three domains of executive control. Therefore, using PRISMA guidelines this systematic review and meta-analyses aimed to identify the magnitude of sex differences in the three prominent executive control domains: performance monitoring, response inhibition, and cognitive set-shifting. In addition, we aimed to identify the shortcomings of the past research on sex differences which may underlie the observed variability in their findings.

2 | METHOD

This review complies with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (see Supplementary Materials I for more details; Moher et al., 2015). Methods for obtaining and handling of data and inclusion criteria were specified in advance and registered with the International Prospective Register of Systematic Reviews (PROSPERO ID: CRD42019124772).

2.1 | Search strategy

The databases PubMed, Scopus, and Web of Science were searched to obtain relevant empirical articles written in English. Search terms included "cognitive control," "executive control," "executive function," "cognitive flexibility," "working memory," "response inhibition"; and "sex," "gender." This preliminary search revealed an extensive number of empirical articles involving paediatric, senior, or non-human populations. Thus, the search was repeated to exclude titles containing "child*," "adol*," "elder*," "old*," and "rat," as these terms appeared at the highest frequency (>4,000 titles; for full search strategy see Supplementary Materials II for more details). No limits were applied for publication dates. To ensure literature saturation, the search was repeated with document type restricted to "review" to check reference and citation lists for additional relevant works. An additional 45 studies were identified for screening (Figure 1). Furthermore, after the first search strategy was completed and the tasks that were relevant to this review were identified (see Section 2.2), individual searches were conducted including the search terms "sex" and "gender" alongside each task names (e.g., "continuous performance test" or "CPT") and compared to the original search results to ensure literature saturation. While "gender" was included in the search strategy, only studies referring to the biological characteristic and not the societal/psychological construct were selected. Older studies frequently used the term "gender" rather than "sex" when referring to "sex."

2.2 | Selection criteria

One reviewer (AG) screened search results for eligibility based on title, abstract, and keywords. Two reviewers (AG

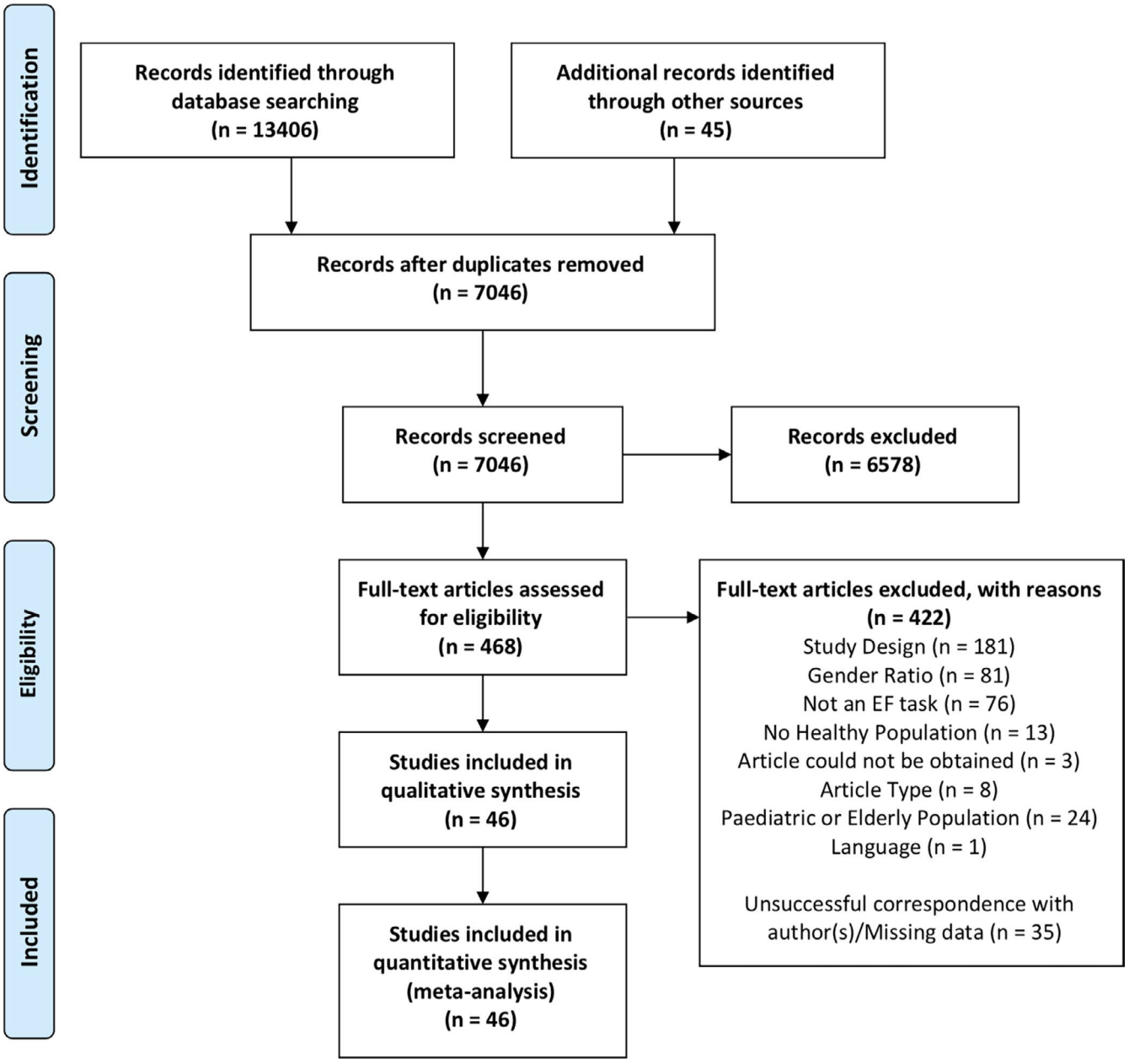

FIGURE 1 PRISMA Flow diagram of article screening and selection process

and DJF) independently screened full-text articles. Studies were included if they met the following criteria: (a) the article was written in English; (b) the basis of the article was empirical; (c) the sample included men and women within a 40% to 60% split; (d) participants were ≥18 years old, however, samples including only ≥45 year olds were excluded; (e) data from healthy/normative samples were reported; (f) the sample had no specified a priori selection factors regarding traits or behaviours; for example samples with alcoholism in first-degree relatives; (g) a robust, not modified, and commonly used executive control task measuring performance monitoring, response inhibition, or cognitive set-shifting was performed; (h) behavioural measures were presented or potentially available from which a sex difference could be calculated. If discrepancies in study selection arose between the two reviewers, these were discussed, and a consensus reached. In cases where an abstract did not provide sufficient information to establish if they met the inclusion criteria, they were included in the next stage of the selection process. Clinical studies were examined and included if data were reported on the healthy control group separately from the clinical group.

2.3 | Risk of bias within studies

Two reviewers (AG and DJF) independently assessed the quality and risk of bias of the included studies. The quality

criteria were adapted from the guidelines set by the Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Case Control Studies (Moola et al., 2017; see Supplementary Materials VII for more details). Discrepancies were resolved by discussion.

2.4 | Data extraction

Data extraction was conducted by two reviewers (AG and DJF) independently during the screening of full-text articles. For each study the following data was extracted: (a) study information (authors, title, date, journal); (b) sample characteristics for each sex, including sample size, age (mean, standard deviation (*SD*), range); (c) executive control task type and relevant variables; (d) all statistics relevant to the magnitude of the sex difference (means, *SD*, correlations, *t* and *F* tests). Where studies did not include adequate data on task variables or statistics, corresponding authors were contacted to request additional data. To ensure data was homogeneous, included studies were organised into three groups based on task domain: (a) performance monitoring; (b) response inhibition; (c) cognitive set-shifting (Table 1; Martin, Clare, Altgassen, Cameron, & Zehnder, 2011). Studies reporting on multiple tasks and tasks containing aspects from multiple domains were considered in more than one group.

2.5 | Statistical analysis

Analyses were conducted with RevMan (The Nordic Cochrane Centre, The Cochrane Collaboration, 2011, Copenhagen, Denmark). Estimates of effect size (Hedges' *g*) were calculated for each eligible study using the available statistics (*F* values, *t* values, *P* values, *r* values, mean ± *SD*, *n*) for the difference in scores between males and female groups within each task domain. Hedges' *g* was calculated by dividing the difference between male and female means by the pooled and weighted standard deviation:

$$\text{Hedges}'\ g = \frac{M_{\text{male}} - M_{\text{female}}}{SD^{*}_{\text{pooled}}}$$

Hedges' *g* was calculated from *F* (4 effect sizes) and *t* (1 effect size). The standardised mean difference (Hedges' *g*) for all studies was aggregated and interpreted as: small = 0.20, medium = 0.50, large = 0.80 (Hedges & Olkin, 1985). A random effects model was used to calculate the standardised weighted mean difference and their 95% confidence interval (CI). This approach was used rather than a fixed-effects model to prevent undue weight being given to small studies with low variance. Heterogeneity was assessed by computing the *Q*-statistic. The *Q* statistic has a chi-squared distribution with $(k - 1)$ degrees of freedom, where *k* is the number of effect sizes being combined (Hedges & Olkin, 1985). The critical alpha for the *Q* statistic was set at 0.05. The overall effect size was assessed through the inspection of forest plots. The risk of publication bias was assessed through funnel plots and Egger's test to determine any asymmetry (Egger, Davey Smith, Schneider, & Minder, 1997); however, Egger's test was not conducted on individual tasks due to the small (<10) number of studies in each group resulting in insufficient power (Sterne et al., 2011). Moreover post hoc analyses, using the random effects model described above, were conducted when necessary to identify task related effects.

3 | RESULTS

3.1 | Included studies

After the removal of duplicates, 7,046 studies were inspected for eligibility, and 6,578 studies were excluded based on the title and abstract. In total, 468 full-text articles were then assessed, with 46 case control studies being eligible for inclusion in the meta-analysis (Figure 1). Ultimately, the sample consisted of 1,988 females (51.34%), and 1,884 males (48.66%). Due to some tasks being included in multiple domains, 31 studies were classed as performance monitoring, 30 as response inhibition, and 15 as cognitive set-shifting (Table 2).

3.2 | Characteristics of included studies

Tables 3–5 report sample characteristics for each study by task domain, as well as each individual task.

3.3 | Performance monitoring

Twenty-six studies reported measures of performance monitoring, producing 32 effect sizes. Across this domain of executive control, the combined effect size was small and statistically not significant (Hedges' $g = -0.08$; 95% CI: −0.21 to 0.05; $Z = 1.17$; $p = 0.24$; Figure 2). There was heterogeneity across these findings ($Q = 63.08$; $I^2 = 51\%$; $p < 0.001$). Inspection of funnel plot did not show asymmetry (Egger's intercept = −0.70; $p = 0.49$; see Supplementary Materials III for more detail).

Post hoc observation revealed that domain scores were not concrete across tasks, and there were clear task related effects. To assess whether there was task specificity between sexes post hoc analyses were conducted. Results of the mean effect analysis for individual performance monitoring tasks are presented in Table 6. Results revealed moderate and

TABLE 1 Summary of behavioural tasks by executive control domain

Performance monitoring	
N-back Task (Kirchner, 1958)	Participants are presented with a sequence of stimuli, and for each stimulus, need to identify if it matches the stimulus presented a certain number of trials back (*N*). *N* can be 0-, 1-, 2-, or 3- trials back. The percent correct responses (accuracy %) was used as a measure of performance monitoring.
Continuous Performance Test [a] (CPT; Conners *et al*., 2000)	A series of stimuli are presented, and for each stimulus, either a response must be made or withheld depending on the instructions. Conner's CPT: a sequence of letters is presented, and participants must respond to each letter except the letter X. CPT-IP: a sequence of number series (2, 3, or 4 digits in length) are presented, and participants must respond when the same number series are presented consecutively. A score for the discrimination between target and non-target trials was used, d' (d prime), as a measure of performance monitoring and response inhibition.
Digit Span (Richardson, 2007)	A sequence of digits is presented, and participants are required to repeat the sequence of digits in the identical order (Forward), or in the reverse order as presented (Backward) as presented. Sequences start as 2 digits and difficulty increases to 9 digits. Total completion time in seconds was used as a total score as a measure of performance monitoring.
Oddball Task (Squires *et al*., 1975)	Participants are presented with a sequence of standard stimuli. Participants must respond to or mentally count and report the infrequent appearance of deviant stimuli. The response time to deviant stimuli is a measure of performance monitoring.
Stockings of Cambridge (Barrett, Kelly, Bell, & King, 2008)	Participants are shown two beam displays with three suspended stockings, each containing three coloured balls. Between the two beams the balls are arranged differently, and participants must move the balls in the bottom beam display one at a time until it is identical to the pattern in the top beam display. Each trial has a set minimum number of moves, and the total trials that are achieved in the minimum number of moves are recorded.
Tower of London [a] (Shallice, 1982)	A sample of three vertical pegs containing arrangements of coloured disks or beads is presented to participants. The arrangement of these coloured disks/beads changes and the participant has to replicate the sample. Preplanning time, errors on the first move, average move time, trials solved in the minimum number of moves, and excess moves are recorded.
CANTAB Spatial Working Memory (Barrett *et al*., 2008)	An array of boxes is presented, and the participant is required to locate a blue token. A blue token will only be found under a particular box once, and that box will not be used again. The number of boxes present in each array increases as the task progresses, starting with four 4-box trials, followed by four 6-box trials, and lastly four 8-box trials. A strategy score is calculated for each trial. A strategy score indicates that the participant began each new search for the blue token from the same box.
Response inhibition	
Attention Network Test (Fan, McCandliss, Fossella, Flombaum, & Posner, 2005)	A combination of the Eriksen Flanker Task, and a spatial cueing task. Participants focus on a fixation point in the centre of the screen and wait for a sequence of five horizontal arrows to which they must respond based on the direction of the centre arrow's direction. The arrows surrounding will either point in the same direction (congruent trial) or opposite direction (incongruent trial). The sequence of arrows may occur above or below the fixation point and may be a cued condition (preceded by an asterisk indicating the correct position of the upcoming arrows) or a no-cue condition. A cue can appear at the centre (centre cue), above or below (spatial cue), or both locations (double cue). Executive control response time (RT) is used as a measure of response inhibition and is measured as the difference between the incongruent and congruent trials, where in incongruent trials participants must implement one response while inhibiting another competing response.
Stroop Task (Stroop, 1935)	Participants are required to name the ink colour of a shape as quickly as possible (e.g., a blue, red, yellow, or green circle) as a control condition. Following this, participants are required to name the colour of the ink used to write the name of a different colour (e.g., the word "yellow" is written in blue ink) as the interference condition. The times taken for each condition are compared. The colour word interference time is a measure of response inhibition as participants must inhibit a dominant and verbal response.
Stop Signal Task (Logan, Cowan, & Davis, 1984)	Evaluates the ability to stop an action after it is initiated. A target stimulus appears requiring participants to make a speeded response. In a subset of randomly intermingled trials, a stop signal occurs at a brief delay after the target stimulus, and participants must inhibit their response. The delay interval is varied to achieve 50% successful inhibition of responses, allowing calculation of stop signal reaction time (SSRT), which is a robust index of participants stopping efficacy (response inhibition).

(Continues)

TABLE 1 (Continued)

Go/No-Go Task (Lappin & Eriksen, 1966)	Involves the performance of two tasks: a frequently occurring go task and an infrequently occurring no-go task. In go trials, participants must make reaction time (RT) responses to go-stimuli. While in no-go trials, participants must not respond. Accuracy on no-go trials is used as a measure of response inhibition, with higher accuracy indicating better inhibitory ability.
Delay Discounting Task (Richards, Zhang, Mitchell, & de Wit, 1999)	Consists of 27 questions that require a choice between a small but immediate reward or a more substantial but delayed reward. The 27 questions are divided into three groups, small, medium, or large, depending on the size of the delayed reward. A discounting rate is then calculated depending on whether there was a bias in selecting the immediate or delayed option and the sizes of the selected rewards.
Eriksen Flanker Task (Eriksen & Eriksen, 1974)	Participants are presented with sequences (i.e., numbers or letters) where they are instructed to focus on the central item while ignoring the surrounding items (the flankers). Participants are informed of the possible central items and the relevant response to each (e.g., a left button press to left arrow stimuli). The flankers are either identical to the central item (congruent trial) or different (incongruent trial). Accuracy on incongruent trials is used as a measure of response inhibition, as participants must implement one response while inhibiting another competing response.
Cognitive set-shifting	
Trail Making Test Part B (Reitan, 1955)	Participants receive a sheet of paper with 25 circles, including both numbers (1–13) and letters (A-L). The participant must connect the circles by alternating between numbers and letters (1-A-2-B, etc.) as quickly as they can while not lifting the pen from the paper. Participants are notified when an error occurs, and correction is allowed, although increases the time to completion. Total completion time in seconds is used as a total score as a measure of cognitive set-shifting.
Intra-/Extra-Dimensional Set-shift (Downes *et al*., 1989)	Two stimuli that represent one stimulus dimension (shape) each are presented initially, then progress to include two stimulus dimensions (shape and lines) each later. Participants identify the correct dimension based on feedback on their selections as to which dimension is correct. After a number of correct responses, the rule for correct stimuli is changed without warning. Changes can be intra-dimensional (i.e., one shape to another) or extra-dimensional (i.e., shape to colour). The total errors committed are recorded as a measure of cognitive set-shifting.
Wisconsin Card Sorting Test [b] (WCST; Berg, 1948)	Participants are required to sort/match cards that contain different numbers (1–4) of shapes (crosses, circles, triangles, or stars) in different colours (red, blue, yellow, or green) to a changing rule. During the task, the required rule changes without informing the participant (i.e., from colour to shape, or shape to number). Participants need to identify and shift to the new rule using trial and error—failure to switch to the new rule results in perseverative errors. Within the WCST, perseverative errors are a valuable index of executive function in multiple aspects. Perseverative errors refer to when the participant selects the previously relevant rule and has been described as 'the inability to suppress ongoing activity despite environmental feedback that it is no longer appropriate' (Everett *et al*., 2001). Thus, in order to avoid the commission of perseverative errors, the participant must: (a) actively monitor response feedback to adjust behaviour, (b) successfully inhibit the previously relevant rule, and (c) shift from the previously relevant rule when error commission feedback is received. Thus perseverative errors may be used as an index of performance monitoring, response inhibition, and cognitive set-shifting, respectively.
Iowa Gambling Task [b] (Bechara, Damasio, Damasio, & Anderson, 1994)	The participant can choose cards from four decks (A–D). Each card choice results in a win or a simultaneous win and loss. There are two "disadvantageous" decks that result in a high monetary reward but occasionally occur with a higher monetary loss (A and B). The other two decks are "advantageous" as they result in a smaller monetary reward but lower monetary losses (C and D), leading to long-term advantage. Participants net score is calculated on how many draws are chosen from disadvantageous decks subtracted from advantageous decks. Higher net scores indicate better task performance, requiring the participant to (a) learn the contingencies of each deck and (b) shift their strategy accordingly to choose from the advantageous deck, despite receiving a smaller reward (Verdejo-Garcia *et al*., 2006). Thus net score may be used as an index of performance monitoring, cognitive set-shifting, and response inhibition, respectively.

[a]Also in response inhibition domain.

[b]Across all three domains.

statistically significant higher scores for male participants on the CANTAB Spatial Working Memory (SWM) Test (Hedges' $g = -0.60$; 95% CI: -0.96 to -0.24; $Z = 3.26$; $p = 0.001$). The remaining 6 performance monitoring tasks; the 2-back task, Continuous Performance Test, Digit Span Backward, Iowa Gambling Task, Stockings of Cambridge, and Wisconsin Card Sorting Test (WCST), yielded non-significant sex differences (see Supplementary Materials IV for more detail). Heterogeneity was observed on the Continuous Performance, and Iowa Gambling tasks (Table 6).

3.4 | Response inhibition

Twenty-seven studies reported measures of response inhibition, producing 30 effect sizes. The combined effect size was

TABLE 2 Number of studies and participants included in the analysis of each executive control task domain and task

Task domain	Task	Number of studies[a]	Total number of participants	Total number M: F
Performance monitoring		26	1,819	879:940
	2-Back	8	523	263:260
	CPT	4	181	84:97
	DS Backward	6	467	229:238
	IGT	4	387	173:214
	Oddball	1	30	15:15
	SOC	3	195	96:99
	SWM	3	127	60:67
	TOL	1	18	9:9
	WCST	2	86	41:45
Response inhibition		27	2,628	1,262:1,366
	ANT	1	73	38:35
	CPT	4	181	84:97
	DD	2	97	47:50
	Flanker	3	358	176:182
	Go/No-Go	3	174	87:87
	IGT	4	387	173:214
	SST	**6**	866	400:466
	Stroop	4	468	247:221
	TOL	1	18	9:9
	WCST	2	86	41:45
Cognitive set-shifting		14	936	446:490
	IED	2	136	72:64
	IGT	4	387	173:214
	TMT B	7	379	183:194
	WCST	2	86	41:45

Abbreviations: ANT, Attention Network Test; CPT, Continuous Performance Task; DD, Delay Discounting; DS, Digit Span; IED, Intra-/Extra-Dimensional Set-shift; IGT, Iowa Gambling Task; SOC, Stockings of Cambridge; SST, Stop Signal Task; SWM, Spatial Working Memory; TMT, Trail Making Test; TOL, Tower of London; WCST, Wisconsin Card Sorting Test.

[a]Each study is included once in the total number of studies per domain, whereas if a study included two tasks, it is included in the number of studies for each individual task.

small and statistically not significant (Hedges' $g = -0.01$; 95% CI: −0.15 to 0.13; $Z = 0.12$; $p = 0.90$; Figure 3), indicating that across all response inhibition-based tasks, males and females performance did not differ. There was heterogeneity across these findings ($Q = 83.02$; $I^2 = 65\%$; $p < 0.001$). Inspection of funnel plot did not show asymmetry (Egger's intercept = 0.35; $p = 0.73$; see Supplementary Materials III for more detail).

Post hoc observation revealed that domain scores were not concrete across tasks, and there were clear task related effects. To assess whether there was task specificity between sexes post hoc analyses were conducted. Results of the mean effect analysis for individual response inhibition tasks are presented in Table 7. Results revealed females scored moderately higher on the Delay Discounting Task, and this was statistically significant (Hedges' $g = 0.64$; 95% CI: 0.24–1.05; $Z = 3.12$; $p = 0.002$). The remaining 6 response inhibition tasks; the Eriksen Flanker Task, Go/No-Go Task, Iowa Gambling Task, Stop Signal Task, Stroop Task, and WCST, yielded non-significant sex differences (see Supplementary Materials IV for more detail). Heterogeneity was observed on the Continuous Performance, Go/No-Go, Iowa Gambling and Stroop tasks (Table 7).

3.5 | Cognitive set-shifting

Fourteen studies reported measures of cognitive set-shifting, producing 15 effect sizes. Within this domain of executive control, there was evidence indicating that males scored

TABLE 3 Performance monitoring task characteristics

Study (year)	*N*	M:F	Mean age (*SD*)	Age range	Measure	Main result[a]
2-Back Task						
Dumais, Chernyak, Nickerson, & Janes (2018)	190	95:95	29.6 (0.30)	22–36	Accuracy (%)	N.S.
Halari *et al*. (2005)	84	42:42		19–35	Accuracy (%)	N.S.
Haut and Barch (2006)	49	23:26	36.602 (11.19)		Accuracy (%)	N.S.
Hsu *et al*. (2015)	30	15:15	33.95 (9.39)	21–57	Accuracy (%)	N.S.
Kalmady *et al*. (2013)	25	14:11	25.456 (3.81)		Accuracy (%)	N.S.
Li, Luo, & Gong, (2010)	50	26:24		18–23	Accuracy (%)	N.S.
Schmidt *et al*. (2009)	50	25:21	33.80 (12.70)		Accuracy (%)	N.S.
Valera *et al*. (2010)	49	23:26	32.5 (10.10)	18–53	Accuracy (%)	N.S.
Continuous Performance Test						
Burton *et al*. (2010)	91	36:55	19.68 (2.49)		d'	**F > M**
Labad *et al*. (2016)	50	28:22	23.79 (4.74)	18–35	d'	N.S.
Mataix-Cols et al. (2006)	40	20:20	28 (7.57)		d'	N.S.
Digit Span Backward						
Duff and Hampson (2001)	92	46:46		18–34	Total Score	N.S.
Hsu *et al*. (2015)	30	15:15	33.95 (9.39)	21–57	Total Score	N.S.
Mataix-Cols et al. (2006)	40	20:20	28 (7.57)		Total Score	N.S.
Robert and Savoie (2006)	100	50:50		19–25	Total Score	N.S.
Torniainen *et al*. (2011)	123	61:62		27.1–69.9	Total Score	N.S.
Zandara *et al*. (2016)	82	37:45		20–39	Total Score	N.S.
Iowa Gambling Task						
Bolla, Eldreth, Matochik, & Cadet, (2004)	20	10:10		21–42	Net Score	**M > F**
Fridberg, Gerst, & Finn, (2013)	152	68:84		18–30	Net Score	N.S.
Icellioglu (2015)	90	45:45	47.9 (15.40)	20–86	Net Score	N.S.
Lage et al. (2013)	125	50:75	24.28 (4.12)		Net Score	N.S.
Oddball Task						
Yuan, He, Qinglin, Chen, & Li, (2008)	30	15:15		18–22	Deviant RT	**F > M**
Stockings of Cambridge						
Barrett *et al*. (2008)	26	12:14	44.54 (11.93)		Total in Minimum Moves	N.S.
Saylik, Raman, & Szameitat, (2018)	110	60:50		18–40	Total in Minimum Moves	N.S.
Suwalska & Lojko (2014)	59	24:35	52.7 (11.70)	26–75	Total in Minimum Moves	N.S.
Spatial Working Memory						
Barrett *et al*. (2008)	26	12:14	44.54 (11.93)		Strategy	**M > F**
Martoni, Salgari, Galimberti, Cavallini, & O'Neill, (2015)	42	24:18	33.09 (12.81)		Strategy	N.S.
Suwalska & Lojko (2014)	59	24:35	52.7 (11.70)	26–75	Strategy	N.S.
Tower of London						
Boghi *et al*. (2006)	18	9:9		27–43	Performance	N.S.

(Continues)

TABLE 3 (Continued)

Study (year)	*N*	M:F	Mean age (*SD*)	Age range	Measure	Main result[a]
Wisconsin Card Sorting Test						
Carrus *et al*. (2010)	46	21:25	42.65 (11.30)	18–70	Perseverative Errors	N.S.
Mataix-Cols et al. (2006)	40	20:20	28 (7.57)		Perseverative Errors	N.S.

Abbreviations: F, female; M, male; N.S., not significant; RT, response time.

[a]Bold significant *P* values are based on between group *t* tests (*t* value) or ANOVA (*F* value).

slightly higher than females across all tasks as the combined effect size was small and statistically significant (Hedges' $g = -0.06$; 95% CI: −0.26 to 0.13; $Z = 0.63$; $p = 0.53$; Figure 4). There was heterogeneity across these findings ($Q = 28.72$; $I^2 = 51\%$; $p = 0.01$). Inspection of funnel plot did not show asymmetry (Egger's intercept = −0.86; $p = 0.40$; see Supplementary Materials III for more detail).

Moreover to assess whether there was task specificity between sexes post hoc analyses were conducted. Results of the mean effect analysis for individual cognitive set-shifting tasks are presented in Table 8. Across all cognitive set-shifting tasks, no significant sex differences were observed. Heterogeneity was observed on the Iowa Gambling tasks (Table 8).

4 | DISCUSSION

To the best of our knowledge, this was the first meta-analysis to assess the magnitude of sex differences in the three executive control domains of performance monitoring, response inhibition, and cognitive set-shifting. We aimed to examine in which executive control domains sex differences were present. We found no sex differences present in the domain of performance monitoring, response inhibition, or cognitive set-shifting. Thus, we explored post hoc whether sex differences were dependent on task within each of the executive control domains. At a task level, we found that males on average scored higher in the performance monitoring domain, specifically on the CANTAB Spatial Working Memory Test. Whereas, females scored higher compared to males on the Delay Discounting Task in the response inhibition domain.

The overall lack of sex differences present across the three domains of executive control indicates that the substantial amount of methodological inconsistencies present (discussed in section 4.4) across the literature need to be addressed to allow for further statistically rigorous analyses. Due to the small sample size within each domain or task and the methodological inconsistencies, it is difficult to make strong claims about the lack of significant differences. However, in accordance with a recently published review (Grissom & Reyes, 2019), it might be time to consider the possibility that there are no sex differences in the majority of behavioural performance on executive control tasks. Grissom and Reyes (2019) postulate that when sex differences are observed in executive functions, sex differences in strategy and outcome assessment may drive apparent effects on executive function, rather than an innate sex difference in executive function directly. In line with this, sex differences in strategy may be facilitated by sex-dependent differences in neural circuitry and/or molecular mechanisms utilised to complete the same cognitive tasks. Behavioural methods alone do not take into consideration macroscopic and microscopic levels of sex differences that have been observed in studies on structure, connectivity, and neurochemistry of the brain (Cahill, 2006). These factors should not be discounted as there is substantial evidence indicating that males and females engage distinct networks and show differences in neurophysiology and distinct regional brain activity during task-performance (Chen, Sachdev, Wen, & Anstey, 2007; Gur et al., 1999; Hidalgo-Lopez et al., 2020; Hjelmervik, Hausmann, Osnes, Westerhausen, & Specht, 2014; Li, Huang, Constable, & Sinha, 2006; Ruigrok et al., 2014; Sacher, Neumann, Okon-Singer, Gotowiec, & Villringer, 2013; Takeuchi et al., 2017).

4.1 | Performance monitoring and cognitive set-shifting

Our meta-analytic results did not support sex-specific findings within the performance monitoring and cognitive set-shifting domains. However, a male advantage was observed in tasks which required spatial working memory, specifically the CANTAB Spatial Working Memory Task.

Previous studies examining performance monitoring and cognitive set-shifting had not reached a consensus on the presence and influence of behavioural sex differences across tasks. In some tasks such as the Digit Span, WCST, and Iowa Gambling Task inconsistent or no behavioural sex differences have been observed (Barel & Tzischinsky, 2018; Lage, Albuquerque, Fuentes, Correa, & Malloy-Diniz, 2013; Mataix-Cols et al., 2006; Niemeier, Marwitz, Lesher, Walker, & Bushnik, 2007). However, in tasks which require spatial working memory, it has emerged that males may possess better spatial abilities than females (Maeda & Yoon, 2013; Voyer, Voyer, & Bryden, 1995).

TABLE 4 Response inhibition task characteristics

Study (year)	*N*	M:F	Mean age (*SD*)	Age range	Measure	Main result
Attention Network Task						
Liu, Hu, Fan, & Wang (2013)	73	38:35	22.6 (1.30)	18–26	Executive Control RT (ms)	N.S.
Continuous Performance Test						
Burton *et al*. (2010)	91	36:55	19.68 (2.49)		d'	**F > M**
Labad *et al*. (2016)	50	28:22	23.79 (4.74)	18–35	d'	N.S.
Mataix-Cols et al. (2006)	40	20:20	28 (7.57)		d'	N.S.
Delay Discounting Task						
Doi, Nishitani, & Shinohara, (2015)	57	27:30	21.97 (2.01)		Log (DD rate)	**F > M**
Peper *et al*. (2013)	40	20:20	21.3 (2.00)	18–25	DD rate	**F > M**
Flanker Task						
Archer, Lee, Qiu, & Annabel Chen, (2018)	46	24:22		21–65	Accuracy (%) Incongruent	N.S.
Clayson, Clawson, & Larson, (2011)	114	54:60	20.96 (2.51)	18–30	Accuracy (%) Incongruent	N.S.
Larson, South, & Clayson (2011)	198	98:100		18–52	Accuracy (%) Incongruent	N.S.
Go/No-Go Task						
Melynyte, Ruksenas, & Griskova-Bulanova, (2017)	79	39:40	22.44	18–30	Accuracy (%) No-Go Trials	N.S.
Ramos-Loyo *et al*. (2016)	30	15:15	27.73 (2.48)		Accuracy (%) No-Go Trials	N.S.
Sjoberg & Cole (2018)	65	33:32	24.89		Accuracy (%) No-Go Trials	**F > M**
Iowa Gambling Task						
Bolla *et al*. (2004)	20	10:10		21–42	Net Score	**M > F**
Fridberg *et al*. (2013)	152	68:84		18–30	Net Score	N.S.
Icellioglu (2015)	90	45:45	47.9 (15.40)	20–86	Net Score	N.S.
Lage et al. (2013)	125	50:75	24.28 (4.12)		Net Score	N.S.
Stop Signal Task						
Gaillard *et al*. (2020)	38	15:23	26.63 (7.49)	18–45	SSRT	**M > F**
Li et al. (2006a)	40	20:20		22–42	SSRT	N.S.
Li et al. (2009)	60	30:30		22–42	SSRT	N.S.
Mei, Tian, Xue, & Li, (2017)	60	30:30	22.21 (2.26)		SSRT	N.S.
Smith, Iredale, & Mattick, (2016)	37	20:17		18–21	SSRT	N.S.
Thakkar *et al*. (2014)	631	285:346		21–50	SSRT	N.S.
Stroop Test						
Fein, Torres, Price, & Di Sclafani, (2006)	48	25:23	45.6	34–59	Colour Word Interference	N.S.
Mataix-Cols et al. (2006)	40	20:20	28 (7.57)		Colour Word Interference	N.S.

(Continues)

TABLE 4 (Continued)

Study (year)	*N*	M:F	Mean age (*SD*)	Age range	Measure	Main result
Tschernegg *et al*. (2017)	40	20:20	26.5 (4.73)	18–36	Colour Word Interference	N.S.
Vaskinn *et al*. (2011)	340	182:158	33.7 (9.70)	18–60	Colour Word Interference	
Tower of London						
Boghi *et al*. (2006)	18	9:9		27–43	Performance	N.S.
Wisconsin Card Sorting Test						
Carrus *et al*. (2010)	46	21:25	42.65 (11.30)	18–70	Perseverative Errors	N.S.
Mataix-Cols et al. (2006)	40	20:20	28 (7.57)		Perseverative Errors	N.S.

Note: Bold significant *P* values are based on between group *t*-tests (*t* value) or ANOVA (*F* value).

Abbreviations: DD, delay discounting; F, female; M, male; N.S., not significant; SSRT, stop signal reaction time.

Similarly, to past literature, this current meta-analysis established that a male advantage only arose in the task requiring spatial working memory.

Past literature has suggested that this male advantage in spatial working memory may emerge from the influence of differences in gonadal steroid hormonal exposure, particularly testosterone; suggesting a strong correlation between testosterone levels and spatial working memory ability (Beauchet, 2006). With some studies have suggested that this relationship is curvilinear (inverted U-shape; Moffat & Hampson, 1996), while others propose a linear interaction (Christiansen & Knussmann, 1987; Silverman, Kastuk, Choi, & Phillips, 1999). Hausmann, Slabbekoorn, Van Goozen, Cohen-Kettenis, and Gunturkun (2000) found that testosterone had a positive effect on mental rotation performance, whereas oestrogen had the opposite effect resulting in decreased performance. In parallel, imaging studies have observed a significant relationship between BOLD activation and testosterone levels occurring alongside better task performance in men (Schoning et al., 2007). The link between testosterone levels and spatial working memory ability is further supported by studies conducted in older males, who have lowered testosterone levels, where substitute testosterone administration evoked positive effect in some cognitive domains, particularly an enhancement in spatial working memory (Beauchet, 2006; Cherrier et al., 2001; Janowsky, Oviatt, & Orwoll, 1994). Although the biological mechanism underlying how testosterone modulates spatial working memory remains poorly characterised, it is clear that increased levels of testosterone have the capacity to enhance spatial cognition. Thus males, who have higher baseline testosterone level than females, may have higher performance in tasks requiring spatial working memory. This framework is supported by our results where males had a significant behavioural advantage within CANTAB Spatial Working Memory Test.

4.2 | Response inhibition

Our meta-analytic data did not support sex-specific findings within response inhibition. Most studies investigating sex differences in response inhibition have hypothesised that there will be a female advantage on such tasks (Labouvie & McGee, 1986). Female superiority has been predicted due to the influence of oestrogen, with oestrogen having facilitating effects on dopaminergic transmission, on which tasks of response inhibition are dependent. One of oestrogens main effects on dopaminergic transmission is through degradation, specifically by targeting catechol-o-methyltransferase (COMT), the primary regulator of dopamine in the PFC (Bilder et al., 2002). Oestrogen has been found to downregulate COMT function, thereby increasing dopamine availability within the PFC (White et al., 2014). However, the results of this meta-analysis indicate no significant differences in behavioural measures of response inhibition between sexes, except for the Delay Discounting task where females had superior performance compared to males. This finding is of note as majority of the response inhibition tasks analysed involved the conscious suppression of a prepotent or dominant response, including the Stop Signal, Go/No-Go, WCST, Continuous Performance, and Stroop tasks; however, the Delay Discounting task relies on different inhibition mechanisms (Cross, Copping, & Campbell, 2011; which we will return to later in this section).

The ability to consciously suppress a prepotent or dominant response has been linked to the Motor Theory of response inhibition in each of these tasks, excluding the Delay Discounting task. Basal ganglia circuits have been linked to successful response inhibition. Yet, at present, there is a lack of clarity to how or whether reported sex differences, both in studies on neural pathway activation (Li et al., 2006; Rubia et al., 2013), and studies investigating the neurobiology influence behaviour on these response inhibition tasks

TABLE 5 Cognitive set-shifting task characteristics

Study (year)	*N*	M:F	Mean age (*SD*)	Age range	Measure	Main result
IED set-shift						
Barrett *et al.* (2008)	26	12:14	44.54 (11.93)		Total Errors	N.S.
Saylik *et al.* (2018)	110	60:50		18–40	Total Errors	N.S.
Iowa Gambling Task						
Bolla *et al.* (2004)	20	10:10		21–42	Net Score	**M > F**
Fridberg *et al.* (2013)	152	68:84		18–30	Net Score	
Icellioglu (2015)	90	45:45	47.9 (15.40)	20–86	Net Score	N.S.
Lage et al. (2013)	125	50:75	24.28 (4.12)		Net Score	N.S.
Trail Making Test B						
Fein *et al.* (2006)	48	25:23	45.6	34–59	Seconds	N.S.
Mataix-Cols et al. (2006)	40	20:20	28 (7.57)		Seconds	N.S.
Misteli *et al.* (2011)	30	15:!5	31.25 (7.39)		Seconds	N.S.
Rahman and Clarke (2005)	39	19:20	26.10 (3.35)	20–34	Seconds	N.S.
Suwalska & Lojko (2014)	59	24:35	52.7 (11.70)	26–75	Seconds	N.S.
Torniainen *et al.* (2011)	111	57:54		27.1–69.9	Seconds	**F > M**
Tschernegg *et al.* (2017)	40	20:20	26.5 (4.73)	18–36	Seconds	N.S.
Wisconsin Card Sorting Test						
Carrus *et al.* (2010)	46	21:25	42.65 (11.30)	18–70	Perseverative Errors	N.S.
Mataix-Cols et al. (2006)	40	20:20	28 (7.57)		Perseverative Errors	N.S.

Note: Bold significant *P* values are based on between group *t*-tests (*t* value) or ANOVA (*F* value).

Abbreviations: F, female; IED, intra-/extra-dimensional set-shift; M, male; N.S., not significant.

(Becker, Perry, & Westenbroek, 2012; Yoest, Cummings, & Becker, 2014). It has been suggested that dopaminergic neurons of the nigrostriatal pathway synapse onto gamma-aminobutyric acid-ergic (GABAergic) medium spiny neurons (MSNs) in the dorsal striatum in which sex differences have been identified which may contradict the facilitating effects of oestrogen on response inhibition. This is due to GABAergic MSNs being situated on either the striatonigral pathway, necessary for the initiation of a response, or the striatopallidal pathway, necessary for the suppression of a prepotent or dominant response (Alexander & Crutcher, 1990). The interaction of oestrogen and these GABAergic MSNs evoke a cascade of effects resulting in a bias toward the striatonigral pathway in females, alongside inhibition of dopamine release from the ventral tegmental area (VTA) which has dopamine efferents to the PFC which attenuates inhibitory control (Chartoff & Mavrikaki, 2015; Margolis, Hjelmstad, Bonci, & Fields, 2003). The results of this literature suggest that oestrogen may not be as beneficial as previously thought and result in females performing on par with males, despite possible differences in the underlying mechanisms, in tasks involving the motor theory of response inhibition, in line with the findings of this meta-analysis.

The Delay Discounting task has been linked to the Evolutionary Theory of response inhibition (Weafer & de Wit, 2014). This theory dictates that greater risk taking is a

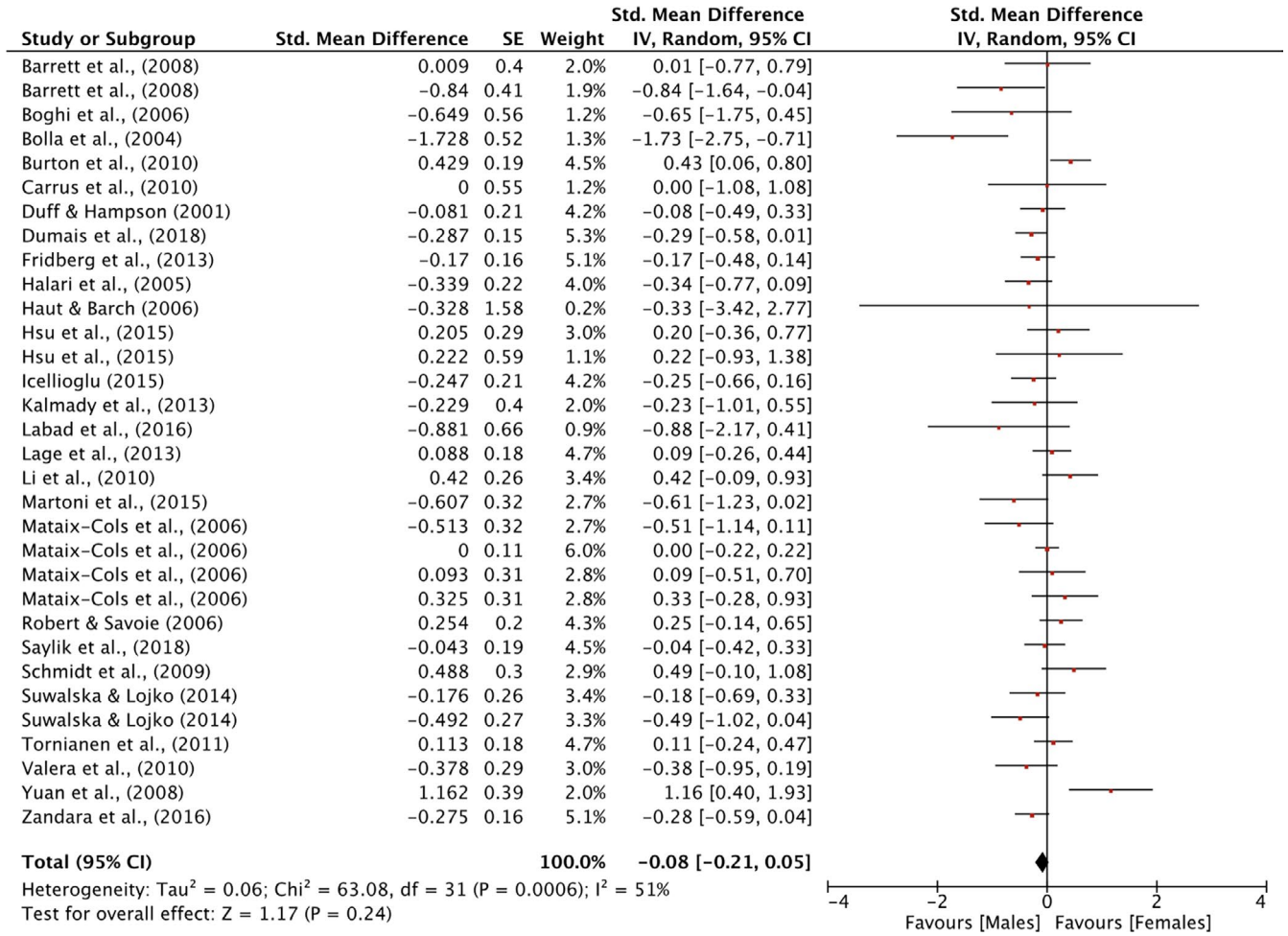

FIGURE 2 Forest plot of comparison on performance monitoring tasks between males and females

characteristic observed in several male mammalian species (Daly & Wilson, 1983). Females maximise reproductive success by seeking mates who can provide not only good genes but resources, including food and protection. This has translated to an improved delay discounting ability, as females do not obtain the same arousal from risk taking as males (Bjorklund & Kipp, 1996). Neuroimaging studies have identified critical regions involved in delay discounting, including the ventral striatum, medial PFC, and posterior cingulate cortex (PCC; Frost & McNaughton, 2017). Furthermore, these regions are related to the mesocortical dopaminergic pathways that innervate the ventral striatum and PFC. As this pathway is predominantly based in the PFC it is influenced by COMT function and oestrogen, which as aforementioned increases availability of dopamine in females, therefore improving inhibitory control. The findings of improved performance on Delay Discounting tasks in females are parallel to those of Bjorklund and Kipp (1996) meta-analysis where women and girls were better able to delay gratification ($r = 0.058$).

Together these findings show that sex differences in response inhibition are task specific, and we predict it is due to the dopaminergic pathway that is employed, whether it is subcortical or cortical. In tasks allowing contemplation regarding long-term goals, females show improved performance compared to males; however, in tasks involving immediate action, the sexes do not differ in behavioural performance.

4.3 | Quality of literature and recommendations for future research on sex differences

Our literature search returned a substantial quantity of studies; however, this meta-analysis intended to identify flaws in the current literature, therefore the final sample was substantially diminished due to inadequate reporting in the methods and results sections. Highlighted below are the most frequently over-looked and observed methodological constraints present.

One of the main issues encountered were participant sex ratios. We excluded 81 studies (19%) based on sex ratios that when calculated, were outside of a 40:60 split. This stringent

TABLE 6 Summary of effect sizes for individual performance monitoring tasks

	Effect size (g) of the sex difference								
Study	**2-back**	**CPT**	**DS B**	**IGT**	**Oddball**	**SOC**	**SWM**	**TOL**	**WCST**
Barrett *et al.* (2008)	–	–	–	–	–	0.009	−0.840	–	–
Boghi *et al.* (2006)	–	–	–	–	–	–	–	−0.649	–
Bolla *et al.* (2004)	–	–	–	−1.728	–	–	–	–	–
Burton *et al.* (2010)	–	0.429	–	–	–	–	–	–	–
Carrus *et al.* (2010)	–	–	–	–	–	–	–	–	0.000
Duff and Hampson (2001)	–	–	−0.081	–	–	–	–	–	–
Dumais *et al.* (2018)	−0.287	–	–	–	–	–	–	–	–
Fridberg *et al.* (2013)	–	–	–	−0.170	–	–	–	–	–
Halari *et al.* (2005)	−0.339	–	–	–	–	–	–	–	–
Haut and Barch (2006)	−0.328	–	–	–	–	–	–	–	–
Hsu *et al,* (2015)	0.205	–	0.222	–	–	–	–	–	–
Icellioglu (2015)	–	–	–	−0.247	–	–	–	–	–
Kalmady *et al.* (2013)	−0.229	–	–	–	–	–	–	–	–
Labad *et al.* (2016)	–	−0.881	–		–	–	–	–	–
Lage et al. (2013)	–	–	–	0.088	–	–	–	–	–
Li *et al.* (2010)	0.420	–	–	–	–	–	–	–	–
Martoni *et al.* (2015)	–	–	–	–	–	–	−0.607	–	–
Mataix-Cols et al. (2006)	–	−0.513	0.093	–	–	–	–	–	0.325
Mataix-Cols et al. (2006)	–	0.000	–	–	–	–	–	–	–
Robert and Savoie (2006)	–	–	0.254	–	–	–	–	–	–
Saylik *et al.* (2018)	–	–	–	–	–	−0.043	–	–	–
Schmidt *et al.* (2009)	0.488	–	–	–	–	–	–	–	–
Suwalska & Lojko (2014)	–	–	–	–	–	−0.176	−0.492	–	–

(Continues)

TABLE 6 (Continued)

	Effect size (g) of the sex difference								
Study	**2-back**	**CPT**	**DS B**	**IGT**	**Oddball**	**SOC**	**SWM**	**TOL**	**WCST**
Torniainen *et al*. (2011)	–	–	0.113	–	–	–	–	–	–
Valera *et al*. (2010)	−0.378	–	–	–	–	–	–	–	–
Yuan *et al*. (2008)	–	–	–	–	1.162	–	–	–	–
Zandara *et al*. (2016)	–	–	−0.275	–	–	–	–	–	–
Hedges' *g*	−0.18	−0.24	−0.01	−0.30	–	−0.08	−0.60**	–	0.25
95% CI (U.L, L.L)	−0.40, 0.04	−0.93, 0.45	−0.18, 0.17	−0.72, 0.12	–	−0.36, 0.20	−0.96, −0.24	–	−0.28, 0.78
Q Homogeneity Statistic	9.05	213.16***	5.33	11.16*	–	0.22	0.50	–	0.26

Note: Effect sizes are in the male direction if negative and in the female direction if positive.

Abbreviations: CI, Confidence Interval; CPT, Continuous Performance Test; DS, Digit Span; F, Female; IGT, Iowa Gambling Task; L.L, Lower Limit; M, Male; SOC, Stockings of Cambridge; SWM, Spatial Working Memory; TOL, Tower of London; U.L, Upper Limit; WCST, Wisconsin Card Sorting Test.

*$P < 0.05$;

**$P < 0.01$;

*** $P < 0.001$.

sex ratio was adhered to in the current meta-analyses, as unequal sex ratios could result in unequal variances, loss of power, and issues with confounding variables that are more likely to appear in the larger group.

Another major issue was inadequate reporting of data, alongside inadequate reporting of valid task measurements. 41 studies (10%) provided insufficient information for Cohen's d or Hedges' g to be calculated. Furthermore, a considerable amount of studies had a lack of uniformity in how outcomes were measured and did not report standard or comparable variables for the task used to be allocated to a domain, further utilising measures that have not demonstrated their reliability or validity (for example providing a delta Stroop score rather than colour-word interference score). Furthermore, two studies did not provide an age range or mean age for participants. These observations are in line with recent reviews on the quality of reporting of statistical results occurring, where despite efforts at achieving standard reporting guidelines, many studies continue to be published despite poor statistical reporting (Heroux, 2016). These reporting issues may be resolved by standardised reporting of task performance (e.g., reaction times, omission errors, and commission errors), as well as the presentation of complete participant demographics, to allow for more repeatable and comparable studies in the future.

Lastly, a portion of the studies was excluded due to the age range included. Sex hormones have been linked to cognitive performance (Colzato, Hertsig, van den Wildenberg, & Hommel, 2010), with performance differences observed in adolescents (Rubia et al., 2013). Puberty is marked by increases and changes in gonadal steroid hormone levels, occurring within adolescence and until reproductive maturation occurs. This reproductive maturation, accompanied by a sharp increase in gonadal steroid hormones is typically complete in most cases by late adolescence and emerging adulthood, between the ages of 18 and 25 (Arnett, 2000). These changes in gonadal steroid hormones, both oestrogen and testosterone, have marked effects on the brain and behaviour, not only the body. These effects include organisation and activation (Sisk & Foster, 2004), and therefore including adolescents in the same participant cohort could alter results, as some neural circuits required may be dormant, including those involved in social and reproductive behaviours (Sisk & Foster, 2004). Furthermore, few studies have assessed the reliability of the included executive control tasks in children and adolescents. Therefore the results of these studies may not be valid for comparison with an adult population.

4.4 | Limitations and future research

The results of this meta-analysis are not without limitations, and there is a clear need for further investigation of sex differences in executive control. Due to the quality of reporting in the literature included, we could not account for moderating factors, including varied task requirements, participant

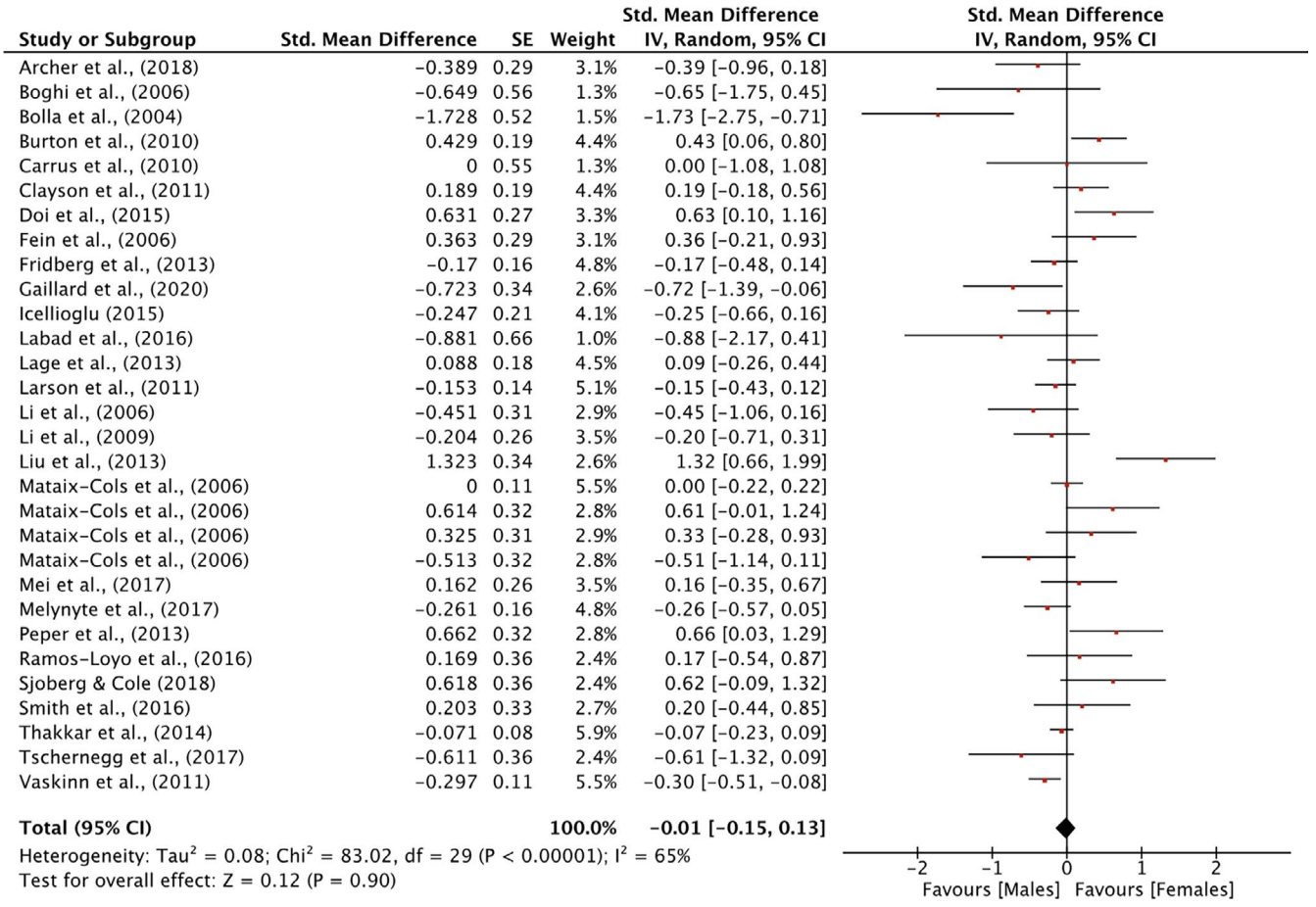

FIGURE 3 Forest plot of comparison on response inhibition tasks between males and females

age, and menstrual cycle in females. Within the performance monitoring domain, there was a lack of consistency in task requirements, and thus there are issues in directly contrasting studies. Regarding age and menstrual cycle in females in most studies, these variables were not reported or controlled for (see Supplementary Materials V for more details). Sensitivity analyses were conducted to omit studies in which participants over 45 years were included (see Supplementary Materials VI for more details), as the menopausal transition has been identified as generally occurring around the ages of 45–55 (Gold, 2011). The sensitivity analyses revealed a small change in effect size, specifically on 2-back task performance. A rapid decline in oestrogens accompanies the later stages of the menopausal transition for up to two years before the final menstrual period (Su & Freeman, 2009). However, hormone changes are not only affected by ageing in women. In men, testosterone concentrations have been shown to decline by approximately 35% in serum, and up to 50% bioavailable testosterone between the ages of 25 and 75 (Gould & Petty, 2000). Future studies should test gonadal hormone levels or control for the potential differences, alongside further study of the differences in executive control task performance between younger and older populations.

Moreover some studies have adopted executive control tasks that require more than one cognitive ability to successfully perform the task; however, limit their findings to the context of one executive control domain. Executive control tasks can be multifactorial, and other abilities such as attention, processing speed, error adjustments, and reasoning can also impact on the results (Chang, Chen, Li, & Li, 2014; Deckler, Hodgins, Pinkham, Penn, & Harvey, 2018; Zhang et al., 2017). More studies that investigate the multiple mechanisms underlying tasks (especially those that overlaps domains) should employ within-task or between-task controls to ensure these effects are actually due to the executive control domain of interest; otherwise, tasks that are better able to isolate specific parameters of interest should be employed.

4.5 | Potential implications for understanding neuropsychiatric conditions

Understanding sex-specific cognitive impairments could provide important context for the study and diagnosis of neuropsychiatric illnesses, such as autism spectrum disorders (Bolte, Duketis, Poustka, & Holtmann, 2011), schizophrenia

TABLE 7 Summary of effect sizes for individual response inhibition tasks

	Effect size (*g*) of the sex difference									
Study	**ANT**	**CPT**	**DD**	**Flanker**	**Go/ No-Go**	**IGT**	**SST**	**Stroop**	**TOL**	**WCST**
Archer *et al.* (2018)	–	–	–	−0.389	–	–	–	–	–	–
Boghi *et al.* (2006)	–	–	–	–	–	–	–	–	−0.649	–
Bolla *et al.* (2004)	–	–	–	–	–	−1.728	–	–	–	–
Burton *et al.* (2010)	–	0.429	–	–	–	–	–	–	–	–
Carrus *et al.* (2010)	–	–	–	–	–	–	–	–	–	0.000
Clayson *et al.* (2011)	–	–	–	0.189	–	–	–	–	–	–
Doi *et al.* (2015)	–	–	0.631	–	–	–	–	–	–	–
Fein *et al.* (2006)	–	–	–	–	–	–	–	−0.363	–	–
Fridberg *et al.* (2013)	–	–	–	–	–	−0.170	–	–	–	–
Gaillard *et al.* (2020)	–	–	–	–	–	–	−0.72	–	–	–
Icellioglu (2015)	–	–	–	–	–	−0.247	–	–	–	–
Labad *et al.* (2016)	–	−0.881	–	–	–	–	–	–	–	–
Lage et al. (2013)	–	–	–	–	–	0.088	–	–	–	–
Larson *et al.* (2011)	–	–	–	−0.153	–	–	–	–	–	–
Li *et al.* (2006)	–	–	–	–	–	–	−0.451	–	–	–
Li et al. (2009)	–	–	–	–	–	–	−0.204	–	–	–
Liu *et al.* (2013)	1.309									
Mataix-Cols et al. (2006)	–	−0.513	–	–	–	–	–	−0.614	–	0.325
Mataix-Cols et al. (2006)	–	0.000	–	–	–	–	–	–	–	–
Mei *et al.* (2017)	–	–	–	–	–	–	0.162	–	–	–
Melynyte *et al.* (2017)	–	–	–	–	−0.261	–	–	–	–	–
Peper *et al.* (2013)	–	–	0.662	–	–	–	–	–	–	–
Ramos-Loyo *et al.* (2016)	–	–	–	–	0.169	–	–	–	–	–

(Continues)

TABLE 7 (Continued)

Study	Effect size (*g*) of the sex difference									
	ANT	CPT	DD	Flanker	Go/ No-Go	IGT	SST	Stroop	TOL	WCST
Sjoberg and Cole (2018)	–	–	–	–	0.618	–	–	–	–	–
Smith *et al*. (2016)	–	–	–	–	–	–	0.203	–	–	–
Thakkar *et al*. (2014)	–	–	–	–	–	–	−0.071	–	–	–
Tschernegg *et al*. (2017)	–	–	–	–	–	–	–	0.611	–	–
Vaskinn *et al*. (2011)	–	–	–	–	–	–	–	0.297	–	–
Hedges' *g*	–	−0.24	0.64*	−0.08	0.15	−0.30	−0.13	0.00	–	0.25
95% CI (U.L, L.L)	–	−0.93, 0.45	0.24, 1.05	−0.37, 0.21	−0.44, 0.73	−0.72, 0.12	−0.34, 0.08	−0.50, 0.51	–	−0.28, 0.78
Q Homogeneity Index	–	213.16***	0.01	3.41	8.12*	11.16*	6.79	12.21**	–	0.26

Note: Effect sizes are in the male direction if negative and in the female direction if positive.

Abbreviations: ANT, Attention Network Test; CI, Confidence Interval; CPT, Continuous Performance Test; DD, Delay Discounting; F, Female; IGT, Iowa Gambling Task; L.L, Lower Limit; M, Male; SST, Stop Signal Task; TOL, Tower of London; U.L, Upper Limit; WCST, Wisconsin Card Sorting Test.

*$P < 0.05$;

**$P < 0.01$;

*** $P < 0.001$.

(Falkenburg & Tracy, 2012), attention deficit hyperactivity disorder (Mowlem et al., 2019), and Parkinson's disease (Georgiev, Hamberg, Hariz, Forsgren, & Hariz, 2017), which have shown sex differences in performance on these tasks, but also in the age of onset, prevalence, symptom severity, and treatment success rates (Abel, Drake, & Goldstein, 2010; Li & Singh, 2014; Miller & Cronin-Golomb, 2010). As aforementioned, males and females may engage different neurophysiology and distinct regional brain activity during task-performance (sometimes in the absence of behavioural

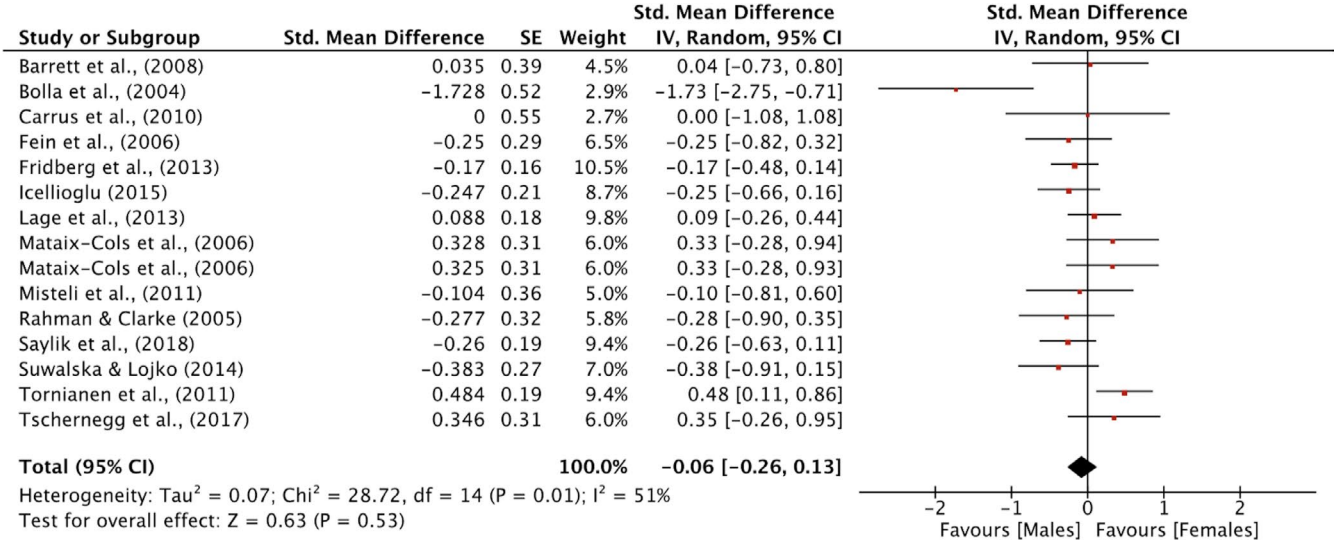

FIGURE 4 Forest plot of comparison on cognitive set shifting tasks between males and females

TABLE 8 Summary of effect sizes for individual cognitive set-shifting tasks

Study	Effect size (*g*) of the sex difference			
	IED	IGT	TMT B	WCST
Barrett *et al.* (2008)	0.035	–	–	–
Bolla *et al.* (2004)	–	−1.728	–	–
Carrus *et al.* (2010)	–	–	–	0.000
Fein *et al.* (2006)	–	–	−0.250	–
Fridberg *et al.* (2013)	–	−0.170	–	–
Icellioglu (2015)	–	−0.247	–	–
Lage et al. (2013)	–	0.088	–	–
Mataix-Cols et al. (2006)	–	–	0.328	0.325
Misteli *et al.* (2011)	–	–	−0.104	–
Rahman and Clarke (2005)	–	–	−0.277	–
Saylik *et al.* (2018)	−0.260	–	–	–
Suwalska & Lojko (2014)	–	–	−0.383	–
Torniainen *et al.* (2011)	–	–	0.487	–
Tschernegg *et al.* (2017)	–	–	0.346	–
Hedges' *g*	−0.20	−0.30	0.05	0.25
95% CI (U.L, L.L)	−0.54, 0.13	−0.72, 0.12	−0.25, 0.34	−0.28, 0.78
Q Homogeneity Index	0.46	11.16*	11.58	0.26

Note: Effect sizes are in the male direction if negative and in the female direction if positive.

Abbreviations: CI, confidence interval; F, female; IED, Intra-/Extra-Dimensional Set-shift; IGT, Iowa Gambling Task; L.L, lower limit; M, male; TMT, Trail Making Test; U.L, upper limit; WCST, Wisconsin Card Sorting Test.

* $P < 0.05$

differences). This has been observed in error-related activation (Li et al., 2009), with these differences possibly underlying sex-dependent vulnerability to executive dysfunction (Ide et al., 2018; Luo et al., 2013; Zhang, Hu, Bednarski, Erdman, & Li, 2014). Moreover demonstrating that behavioural studies alone cannot thoroughly identify these sex differences in executive control, and future multi-modal studies involving neuroimaging is required.

4.6 | Conclusion

In summary, this meta-analysis examined available research in sex differences within three domains of executive control: performance monitoring, response inhibition, and cognitive set-shifting. At a domain level, males and females did not differ on performance in these domains. However, task-specific sex differences were observed in the CANTAB Spatial Working Memory task, where males scored higher than females, and in the Delay Discounting task where females scored higher than males. Future research, with better control for age and sex hormone levels, is required to understand the observed task-specific sex differences.

ACKNOWLEDGEMENTS

SLR holds an NHMRC Senior Research Fellowship (GNT1154651). DJF holds an ARC Centre of Excellence in Integrative Brain Function Scholarship.

CONFLICTS OF INTEREST

The authors declare no conflict of interests.

AUTHOR CONTRIBUTIONS

AG and SLR prepared the research question. AG and DJF conducted searches and screened articles independently. AG conducted the statistical analyses. AG and DJF prepared the initial manuscript and all authors revised and contributed to the final manuscript.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/ejn.14946

DATA AVAILABILITY STATEMENT

Additional supporting information may be found online in the Supplementary Materials section.

ORCID

Alexandra Gaillard https://orcid.org/0000-0002-4852-0238

REFERENCES

Abel, K. M., Drake, R., & Goldstein, J. M. (2010). Sex differences in schizophrenia. *International Review of Psychiatry*, *22*, 417–428.

Alexander, G. E., & Crutcher, M. D. (1990). Functional architecture of basal ganglia circuits: Neural substrates of parallel processing. *Trends in Neurosciences*, *13*, 266–271. https://doi.org/10.1016/0166-2236(90)90107-L

Archer, J. A. , Lee, A. , Qiu, A. , & Annabel Chen, S. H. (2018). Functional connectivity of resting-state, working memory and inhibition networks in perceived stress. *Neurobiology of Stress*, *8*, 186–201. https://doi.org/10.1016/j.ynstr.2017.01.002

Arnett, J. J. (2000). Emerging adulthood. A theory of development from the late teens through the twenties. *American Psychologist*, *55*, 469–480.

Barel, E., & Tzischinsky, O. (2018). Age and sex differences in verbal and visuospatial abilities. *Advances in Cognitive Psychology*, *14*, 51–61.

Barrett, S. L., Kelly, C., Bell, R., & King, D. J. (2008). Gender influences the detection of spatial working memory deficits in bipolar disorder. *Bipolar Disorders*, *10*, 647–654. https://doi.org/10.1111/j.1399-5618.2008.00592.x

Beauchet, O. (2006). Testosterone and cognitive function: Current clinical evidence of a relationship. *European Journal of Endocrinology*, *155*, 773–781.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*, 7–15. https://doi.org/10.1016/0010-0277(94)90018-3

Becker, J. B., Perry, A. N., & Westenbroek, C. (2012). Sex differences in the neural mechanisms mediating addiction: A new synthesis and hypothesis. *Biology of Sex Differences*, *3*, 14.

Berg, E. A. (1948). A simple objective technique for measuring flexibility in thinking. *The Journal of General Psychology*, *39*, 15–22. https://doi.org/10.1080/00221309.1948.9918159

Bilder, R. M., Volavka, J., Czobor, P., Malhotra, A. K., Kennedy, J. L., Ni, X., … Lieberman, J. A. (2002). Neurocognitive correlates of the COMT Val158Met polymorphism in chronic schizophrenia. *Biological Psychiatry*, *52*, 701–707.

Bjorklund, D. F., & Kipp, K. (1996). Parental investment theory and gender differences in the evolution of inhibition mechanisms. *Psychological Bulletin*, *120*, 163–188.

Boghi, A., Rasetti, R., Avidano, F., Manzone, C., Orsi, L., D'Agata, F., … Mortara, P. (2006). The effect of gender on planning: An fMRI study using the Tower of London task. *NeuroImage*, *33*, 999–1010. https://doi.org/10.1016/j.neuroimage.2006.07.022

Bolte, S., Duketis, E., Poustka, F., & Holtmann, M. (2011). Sex differences in cognitive domains and their clinical correlates in higher-functioning autism spectrum disorders. *Autism*, *15*, 497–511. https://doi.org/10.1177/1362361310391116

Bolla, K. I., Eldreth, D. A., Matochik, J. A., & Cadet, J. L. (2004). Sex-related differences in a gambling task and its neurological correlates. *Cerebral Cortex*, *14*, 1226–1232. https://doi.org/10.1093/cercor/bhh083

Bonte, E., Flemming, T., & Fagot, J. (2011). Executive control of perceptual features and abstract relations by baboons (Papio papio). *Behavioral Brain Research*, *222*, 176–182.

Burton, L., Pfaff, D., Bolt, N., Hadjikyriacou, D., Silton, N., Kilgallen, C., … Allimant, J. (2010). Effects of gender and personality on the Conners Continuous Performance Test. *Journal of Clinical and Experimental Neuropsychology*, *32*, 66–70.

Cahill, L. (2006). Why sex matters for neuroscience. *Nature Reviews Neuroscience*, *7*, 477–484. https://doi.org/10.1038/nrn1909

Carrus, D., Christodoulou, T., Hadjulis, M., Haldane, M., Galea, A., Koukopoulos, A., … Frangou, S. (2010). Gender differences in immediate memory in bipolar disorder. *Psychological Medicine*, *40*, 1349–1355. https://doi.org/10.1017/S0033291709991644

Chang, A., Chen, C. C., Li, H. H., & Li, C. S. (2014). Event-related potentials for post-error and post-conflict slowing. *PLoS One*, *9*, e99909. https://doi.org/10.1371/journal.pone.0099909

Chartoff, E. H., & Mavrikaki, M. (2015). Sex differences in kappa opioid receptor function and their potential impact on addiction. *Frontiers in Neuroscience*, *9*, 466.

Chen, X., Sachdev, P. S., Wen, W., & Anstey, K. J. (2007). Sex differences in regional gray matter in healthy individuals aged 44–48 years: A voxel-based morphometric study. *NeuroImage*, *36*, 691–699.

Cherrier, M. M., Asthana, S., Plymate, S., Baker, L., Matsumoto, A. M., Peskind, E., … Craft, S. (2001). Testosterone supplementation improves spatial and verbal memory in healthy older men. *Neurology*, *57*, 80–88. https://doi.org/10.1212/WNL.57.1.80

Christiansen, K., & Knussmann, R. (1987). Sex hormones and cognitive functioning in men. *Neuropsychobiology*, *18*, 27–36. https://doi.org/10.1159/000118389

Clayson, P. E., Clawson, A., & Larson, M. J. (2011). Sex differences in electrophysiological indices of conflict monitoring. *Biological Psychology*, *87*, 282–289. https://doi.org/10.1016/j.biopsycho.2011.03.011

Colzato, L. S., Hertsig, G., van den Wildenberg, W. P., & Hommel, B. (2010). Estrogen modulates inhibitory control in healthy human females: Evidence from the stop-signal paradigm. *Neuroscience*, *167*, 709–715. https://doi.org/10.1016/j.neuroscience.2010.02.029

Conners, C. K., Staff, M. H. S., Connelly, V., Campbell, S., MacLean, M., & Barnes, J. (2000) Conners' continuous performance Test II (CPT II v. 5). *Multi-Health System Inc*, *29*, 175–196.

Cotto, J. H., Davis, E., Dowling, G. J., Elcano, J. C., Staton, A. B., & Weiss, S. R. (2010). Gender effects on drug use, abuse, and dependence: A special analysis of results from the National Survey on Drug Use and Health. *Gender Medicine*, *7*, 402–413. https://doi.org/10.1016/j.genm.2010.09.004

Cross, C. P., Copping, L. T., & Campbell, A. (2011). Sex differences in impulsivity: A meta-analysis. *Psychological Bulletin*, *137*, 97–130.

Cummings, J. L. (1993). Frontal-subcortical circuits and human behavior. *Archives of Neurology*, *50*, 873–880.

Daly, M., & Wilson, M. (1983). *Sex, evolution and behavior*. Boston, MA: Willard Grant Press.

Deckler, E., Hodgins, G. E., Pinkham, A. E., Penn, D. L., & Harvey, P. D. (2018). Social cognition and neurocognition in schizophrenia and healthy controls: Intercorrelations of performance and effects of manipulations aimed at increasing task difficulty. *Frontiers in Psychiatry*, *9*, 356. https://doi.org/10.3389/fpsyt.2018.00356

Doi, H., Nishitani, S., & Shinohara, K. (2015). Sex difference in the relationship between salivary testosterone and inter-temporal choice. *Hormones and Behavior*, *69*, 50–58.

Duff, S. J., & Hampson, E. (2001). A sex difference on a novel spatial working memory task in humans. *Brain and Cognition*, *47*, 470–493.

Dumais, K. M., Chernyak, S., Nickerson, L. D., & Janes, A. C. (2018). Sex differences in default mode and dorsal attention network engagement. *PLoS One*, *13*, e0199049. https://doi.org/10.1371/journal.pone.0199049

Downes, J. J., Roberts, A. C., Sahakian, B. J., Evenden, J. L., Morris, R. G., & Robbins, T. W. (1989). Impaired extra-dimensional shift performance in medicated and unmedicated Parkinson's disease: evidence for a specific attentional dysfunction. *Neuropsychologia*, *27*, 1329–1343. https://doi.org/10.1016/0028-3932(89)90128-0

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, *315*, 629–634. https://doi.org/10.1136/bmj.315.7109.629

Egner, T., & Hirsch, J. (2005). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nature Neuroscience*, *8*, 1784–1790. https://doi.org/10.1038/nn1594

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Percept. Psychophys.*, *16*, 143–149. https://doi.org/10.3758/BF03203267

Everett, J., Lavoie, K., Gagnon, J. F., & Gosselin, N. (2001). Performance of patients with schizophrenia on the Wisconsin Card Sorting Test (WCST). *Journal of Psychiatry and Neuroscience*, *26*, 123–130.

Falkenburg, J., & Tracy, D. K. (2012). Sex and schizophrenia: A review of gender differences. *Psychosis*, *6*, 61–69. https://doi.org/10.1080/17522439.2012.733405

Fan, J., McCandliss, B. D., Fossella, J., Flombaum, J. I., & Posner, M. I. (2005). The activation of attentional networks. *NeuroImage*, *26*, 471–479. https://doi.org/10.1016/j.neuroimage.2005.02.004

Fein, G., Torres, J., Price, L. J., & Di Sclafani, V. (2006). Cognitive performance in long-term abstinent alcoholic individuals. *Alcoholism: Clinical and Experimental Research*, *30*(9), 1538–1544. https://doi.org/10.1111/j.1530-0277.2006.00185.x

Frost, R., & McNaughton, N. (2017). The neural basis of delay discounting: A review and preliminary model. *Neuroscience and Biobehavioral Reviews*, *79*, 48–65.

Fridberg, D. J. , Gerst, K. R. , & Finn, P. R. (2013). Effects of working memory load, a history of conduct disorder, and sex on decision making in substance dependent individuals. *Drug and Alcohol Dependence*, *133*, 654–660. https://doi.org/10.1016/j.drugalcdep.2013.08.014

Geary, D. C., Saults, S. J., Liu, F., & Hoard, M. K. (2000). Sex differences in spatial cognition, computational fluency, and arithmetical reasoning. *Journal of Experimental Child Psychology*, *77*, 337–353.

Gaillard, A., Rossell, S. L., Carruthers, S. P., Sumner, P. J., Michie, P. T., Woods, W., … Hughes, M. E. (2020). Greater activation of the response inhibition network in females compared to males during stop signal task performance. *Behavioral Brain Research*, *386*, 112586. https://doi.org/10.1016/j.bbr.2020.112586

Georgiev, D., Hamberg, K., Hariz, M., Forsgren, L., & Hariz, G. M. (2017). Gender differences in Parkinson‘s disease: A clinical perspective. *Acta Neurologica Scandinavica*, *136*, 570–584.

Gold, E. B. (2011). The timing of the age at which natural menopause occurs. *Obstetrics and Gynecology Clinics of North America*, *38*, 425–440.

Gould, D. C., & Petty, R. (2000). The male menopause – Does it exist? *BMJ*, *320*, 858–861.

Grissom, N. M., & Reyes, T. M. (2019). Let‘s call the whole thing off: Evaluating gender and sex differences in executive function. *Neuropsychopharmacology*, *44*, 86–96. https://doi.org/10.1038/s41386-018-0179-5

Gur, R. C., Turetsky, B. I., Matsui, M., Yan, M., Bilker, W., Hughett, P., & Gur, R. E. (1999). Sex differences in brain gray and white matter in healthy young adults: Correlations with cognitive performance. *Journal of Neuroscience*, *19*, 4065–4072. https://doi.org/10.1523/JNEUROSCI.19-10-04065.1999

Halari, R., Hines, M., Kumari, V., Mehrotra, R., Wheeler, M., Ng, V., & Sharma, T. (2005). Sex differences and individual differences in cognitive performance and their relationship to endogenous gonadal hormones and gonadotropins. *Behavioral Neuroscience*, *119*, 104–117. https://doi.org/10.1037/0735-7044.119.1.104

Hausmann, M., Slabbekoorn, D., Van Goozen, S. H. M., Cohen-Kettenis, P. T., & Gunturkun, O. (2000). Sex hormones affect spatial abilities during the menstrual cycle. *Behavioral Neuroscience*, *114*, 1245–1250.

Haut, K. M., & Barch, D. M. (2006). Sex influences on material-sensitive functional lateralization in working and episodic memory: men and women are not all that different. *NeuroImage*, *32*, 411–422.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. London: Academic Press Inc..

Heroux, M. (2016). Inadequate reporting of statistical results. *Journal of Neurophysiology*, *116*, 1536–1537.

Hidalgo-Lopez, E., Mueller, K., Harris, T., Aichhorn, M., Sacher, J., & Pletzer, B. (2020). Human menstrual cycle variation in subcortical functional brain connectivity: A multimodal analysis approach. *Brain Structure and Function*, *225*, 591–605. https://doi.org/10.1007/s00429-019-02019-z

Hjelmervik, H., Hausmann, M., Osnes, B., Westerhausen, R., & Specht, K. (2014). Resting states are resting traits – An FMRI study of sex differences and menstrual cycle effects in resting state cognitive control networks. *PLoS One*, *9*, e103492. https://doi.org/10.1371/journal.pone.0103492

Hsu, H. L., Chen, D. Y., Tseng, Y. C., Kuo, Y. S., Huang, Y. L., Chiu, W. T., … Chen, C. J. (2015). Sex Differences in Working Memory after Mild Traumatic Brain Injury: A Functional MR Imaging Study. *Radiology*, *276*, 828–835. https://doi.org/10.1148/radiol.2015142549

Hyde, J. S. (1981). How large are cognitive gender differences? A meta-analysis using !w2 and d. *American Psychologist*, *36*, 892–901.

Ide, J. S., Zhornitsky, S., Chao, H. H., Zhang, S., Hu, S., Wang, W., … Li, C. R. (2018). Thalamic cortical error-related responses in adult social drinkers: Sex differences and problem alcohol use. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *3*, 868–877. https://doi.org/10.1016/j.bpsc.2018.04.008

Icellioglu, S. (2015). Iowa gambling test: Normative data and correlation with executive functions. *Düşünen Adam*, *28*, 222–230.

Janowsky, J. S., Oviatt, S. K., & Orwoll, E. S. (1994). Testosterone influences spatial cognition in older men. *Behavioral Neuroscience*, *108*, 325–332.

Kalmady, S. V., Agarwal, S. M., Shivakumar, V., Jose, D., Venkatasubramanian, G., & Reddy, Y. C. (2013). Revisiting Geschwind‘s hypothesis on brain lateralisation: a functional MRI study of digit ratio (2D:4D) and sex interaction effects on spatial working memory. *Laterality*, *18*, 625–640.

Keenan, P. A., Ezzat, W. H., Ginsburg, K., & Moore, G. J. (2001). Prefrontal cortex as the site of estrogen‘s effect on cognition. *Psychoneuroendocrinology*, *26*, 577–590. https://doi.org/10.1016/S0306-4530(01)00013-0

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *J. Exp. Psychol.*, *55*, 352–358. https://doi.org/10.1037/h0043688

Labad, J., Gutierrez-Zotes, A., Creus, M., Montalvo, I., Cabezas, A., Sole, M., … Vilella, E. (2016). Hypothalamic-pituitary-adrenal axis measures and cognitive abilities in early psychosis: Are there sex differences? *Psychoneuroendocrinology*, *72*, 54–62.

Labouvie, E. W., & McGee, C. R. (1986). Relation of personality to alcohol and drug use in adolescence. *Journal of Consulting and Clinical Psychology*, *54*, 289–293.

Lage, G. M., Albuquerque, M. R., Fuentes, D., Correa, H., & Malloy-Diniz, L. F. (2013). Sex differences in dimensions of impulsivity in a non-clinical sample. *Perceptual and Motor Skills*, *117*, 601–607.

Li, C. S. , Huang, C. , Constable, R. T. , & Sinha, R. (2006). Gender differences in the neural correlates of response inhibition during a stop signal task. *NeuroImage*, *32*, 1918–1929. https://doi.org/10.1016/j.neuroimage.2006.05.017

Li, T., Luo, Q., & Gong, H. (2010). Gender-specific hemodynamics in prefrontal cortex during a verbal working memory task by near-infrared spectroscopy. *Behavioral Brain Research*, *209*, 148–153.

Lappin, J. S., & Eriksen, C. W. (1966). Use of a delayed signal to stop a visual reaction-time response. *J. Exp. Psychol.*, *72*, 805–811. https://doi.org/10.1037/h0021266

Larson, M. J., South, M., & Clayson, P. E. (2011). Sex differences in error-related performance monitoring. *NeuroReport*, *22*, 44–48. https://doi.org/10.1097/WNR.0b013e3283427403

Li, R., & Singh, M. (2014). Sex differences in cognitive impairment and Alzheimer's disease. *Frontiers in Neuroendocrinology*, *35*, 385–403.

Liu, G., Hu, P. P., Fan, J., & Wang, K. (2013). Gender differences associated with orienting attentional networks in healthy subjects. *Chinese Medical Journal*, *126*, 2308–2312.

Logan, G. D., Cowan, W. B., & Davis, K. A. (1984). On the ability to inhibit simple and choice reaction time responses: a model and a method. *Journal of Experimental Psychology: Human Perception and Performance*, *10*, 276–291. https://doi.org/10.1037/0096-1523.10.2.276

Li, C. S., Zhang, S., Duann, J. R., Yan, P., Sinha, R., & Mazure, C. M. (2009). Gender differences in cognitive control: An extended investigation of the stop signal task. *Brain Imaging and Behavior*, *3*, 262–276.

Luo, X., Zhang, S., Hu, S., Bednarski, S. R., Erdman, E., Farr, O. M., … Li, C. S. (2013). Error processing and gender-shared and -specific neural predictors of relapse in cocaine dependence. *Brain*, *136*, 1231–1244. https://doi.org/10.1093/brain/awt040

Maeda, Y., & Yoon, S. Y. (2013). A meta-analysis on gender differences in mental rotation ability measured by the Purdue spatial visualization tests: Visualization of rotations (PSVT:R). *Educational Psychologist*, *25*, 69–94. https://doi.org/10.1007/s10648-012-9215-x

Mansouri, F. A., Buckley, M. J., Fehring, D. J., & Tanaka, K. (2020). The role of primate prefrontal cortex in bias and shift between visual dimensions. *Cerebral Cortex*, *30*, 85–99. https://doi.org/10.1093/cercor/bhz072

Mansouri, F. A., Buckley, M. J., & Tanaka, K. (2014). The essential role of primate orbitofrontal cortex in conflict-induced executive control adjustment. *Journal of Neuroscience*, *34*, 11016–11031. https://doi.org/10.1523/JNEUROSCI.1637-14.2014

Mansouri, F. A., Tanaka, K., & Buckley, M. J. (2009). Conflict-induced behavioural adjustment: A clue to the executive functions of the prefrontal cortex. *Nature Reviews Neuroscience*, *10*, 141–152.

Margolis, E. B., Hjelmstad, G. O., Bonci, A., & Fields, H. L. (2003). Kappa-opioid agonists directly inhibit midbrain dopaminergic neurons. *Journal of Neuroscience*, *23*, 9981–9986.

Martin, M., Clare, L., Altgassen, A. M., Cameron, M. H., & Zehnder, F. (2011) Cognition-based interventions for healthy older people and people with mild cognitive impairment. *Cochrane Database of Systematic Reviews*, CD006220. https://doi.org/10.1002/14651858.CD006220.pub2

Mataix-Cols, D., Rahman, Q., Spiller, M., Alonso, M. P., Pifarre, J., Menchon, J. M., & Vallejo, J. (2006). Are there sex differences in neuropsychological functions among patients with obsessive-compulsive disorder? *Applied Neuropsychology*, *13*, 42–50.

Martoni, R. M., Salgari, G., Galimberti, E., Cavallini, M. C., & O'Neill, J. (2015). Effects of gender and executive function on visuospatial working memory in adult obsessive-compulsive disorder. *European Archives of Psychiatry and Clinical Neuroscience*, *265*, 707–718. https://doi.org/10.1007/s00406-015-0604-2

Mei, X., Tian, L., Xue, Z., & Li, X. (2017). A working memory task reveals different patterns of impulsivity in male and female college students. *Behavioural Processes*, *138*, 127–133. https://doi.org/10.1016/j.beproc.2017.02.023

Melynyte, S., Ruksenas, O., & Griskova-Bulanova, I. (2017). Sex differences in equiprobable auditory Go/NoGo task: effects on N2 and P3. *Experimental Brain Research*, *235*, 1565–1574. https://doi.org/10.1007/s00221-017-4911-x

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167–202.

Miller, I. N., & Cronin-Golomb, A. (2010). Gender differences in Parkinson's disease: Clinical characteristics and cognition. *Movement Disorders*, *25*, 2695–2703.

Misteli, M., Duschek, S., Richter, A., Grimm, S., Rezk, M., Kraehenmann, R., … Schuepbach, D. (2011). Gender characteristics of cerebral hemodynamics during complex cognitive functioning. *Brain and Cognition*, *76*, 123–130. https://doi.org/10.1016/j.bandc.2011.02.009

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*, 49–100. https://doi.org/10.1006/cogp.1999.0734

Moffat, S. D., & Hampson, E. (1996). A curvilinear relationship between testosterone and spatial cognition in humans: Possible influence of hand preference. *Psychoneuroendocrinology*, *21*, 323–337. https://doi.org/10.1016/0306-4530(95)00051-8

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., … … PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, *4*, 1.

Moola, S., Munn, Z., Tufanaru, C., Aromataris, E., Sears, K., Sfetcu, R., … Mu, P. (2017). Chapter 7: Systematic reviews of etiology and risk. In E. Aromataris, & Z. Munn (Eds.), *Joanna Briggs Institute reviewer's manual*, (217–269). Adelaide, Australia: The Joanna Briggs Institute.

Mowlem, F. D., Rosenqvist, M. A., Martin, J., Lichtenstein, P., Asherson, P., & Larsson, H. (2019). Sex differences in predicting ADHD clinical diagnosis and pharmacological treatment. *European Child and Adolescent Psychiatry*, *28*, 481–489.

Mulvihill, L. E., Skilling, T. A., & Vogel-Sprott, M. (1997). Alcohol and the ability to inhibit behavior in men and women. *Journal of Studies on Alcohol and Drugs*, *58*, 600–605.

Niemeier, J. P., Marwitz, J. H., Lesher, K., Walker, W. C., & Bushnik, T. (2007). Gender differences in executive functions following traumatic brain injury. *Neuropsychological Rehabilitation*, *17*, 293–313.

Peper, J. S., Mandl, R. C., Braams, B. R., de Water, E., Heijboer, A. C., Koolschijn, P. C., & Crone, E. A. (2013). Delay discounting and frontostriatal fiber tracts: a combined DTI and MTR study on impulsive choices in healthy young adults. *Cerebral Cortex*, *23*, 1695–1702. https://doi.org/10.1093/cercor/bhs163

Rahman, Q., & Clarke, C. D. (2005). Sex differences in neurocognitive functioning among abstinent recreational cocaine users. *Psychopharmacology (Berl)*, *181*, 374–380. https://doi.org/10.1007/s00213-005-2257-8

Ramos-Loyo, J., Angulo-Chavira, A., Llamas-Alonso, L. A., & Gonzalez-Garrido, A. A. (2016). Sex differences in emotional contexts modulation on response inhibition. *Neuropsychologia*, *91*, 290–298. https://doi.org/10.1016/j.neuropsychologia.2016.08.023

Reitan, R. M. (1955). The relation of the trail making test to organic brain damage. *Journal of Consulting Psychology*, *19*, 393–394. https://doi.org/10.1037/h0044509

Richards, J. B. , Zhang, L. , Mitchell, S. H. , & de Wit, H. (1999). Delay or probability discounting in a model of impulsive behavior: Effect of alcohol. *Journal of the Experimental Analysis of Behavior*, *71*, 121–143. https://doi.org/10.1901/jeab.1999.71-121

Richardson, J. T. (2007). Measures of short-term memory: A historical review. *Cortex*, *43*, 635–650. https://doi.org/10.1016/S0010-9452(08)70493-3

Robert, M., & Savoie, N. (2006). Are there gender differences in verbal and visuospatial working-memory resources? *European Journal of Cognitive Psychology*, *18*, 378–397.

Rubia, K., Lim, L., Ecker, C., Halari, R., Giampietro, V., Simmons, A., … Smith, A. (2013). Effects of age and gender on neural networks of motor response inhibition: From adolescence to mid-adulthood. *NeuroImage*, *83*, 690–703. https://doi.org/10.1016/j.neuroimage.2013.06.078

Ruigrok, A. N., Salimi-Khorshidi, G., Lai, M. C., Baron-Cohen, S., Lombardo, M. V., Tait, R. J., & Suckling, J. (2014). A meta-analysis of sex differences in human brain structure. *Neuroscience and Biobehavioral Reviews*, *39*, 34–50. https://doi.org/10.1016/j.neubiorev.2013.12.004

Sacher, J., Neumann, J., Okon-Singer, H., Gotowiec, S., & Villringer, A. (2013). Sexual dimorphism in the human brain: Evidence from neuroimaging. *Magnetic Resonance Imaging*, *31*, 366–375. https://doi.org/10.1016/j.mri.2012.06.007

Saylik, R., Raman, E., & Szameitat, A. J. (2018). Sex differences in emotion recognition and working memory tasks. *Frontiers in Psychology*, *9*, 1072. https://doi.org/10.3389/fpsyg.2018.01072

Schmidt, H. , Jogia, J. , Fast, K. , Christodoulou, T. , Haldane, M. , Kumari, V. , & Frangou, S. (2009). No gender differences in brain activation during the N-back task: An fMRI study in healthy individuals. *Human Brain Mapping*, *30*, 3609–3615.

Schoning, S., Engelien, A., Kugel, H., Schafer, S., Schiffbauer, H., Zwitserlood, P., … Konrad, C. (2007). Functional anatomy of visuo-spatial working memory during mental rotation is influenced by sex, menstrual cycle, and sex steroid hormones. *Neuropsychologia*, *45*, 3203–3214. https://doi.org/10.1016/j.neuropsychologia.2007.06.011

Shallice, T. (1982). Specific impairments of planning. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *298*, 199–209.

Shanmugan, S., & Epperson, C. N. (2014). Estrogen and the prefrontal cortex: Towards a new understanding of estrogen's effects on executive functions in the menopause transition. *Human Brain Mapping*, *35*, 847–865.

Silverman, I., Kastuk, D., Choi, J., & Phillips, K. (1999). Testosterone levels and spatial ability in men. *Psychoneuroendocrinology*, *24*, 813–822. https://doi.org/10.1016/S0306-4530(99)00031-1

Sisk, C. L., & Foster, D. L. (2004). The neural basis of puberty and adolescence. *Nature Neuroscience*, *7*, 1040–1047.

Sjoberg, E. A., & Cole, G. G. (2018). Sex Differences on the Go/No-Go Test of Inhibition. *Archives of Sexual Behavior*, *47*, 537–542. https://doi.org/10.1007/s10508-017-1010-9

Smith, J. L., Iredale, J. M., & Mattick, R. P. (2016). Sex differences in the relationship between heavy alcohol use, inhibition and performance monitoring: Disconnect between behavioural and brain functional measures. *Psychiatry Res. Neuroimaging*, *254*, 103–111. https://doi.org/10.1016/j.pscychresns.2016.06.012

Squires, N. K., Squires, K. C., & Hillyard, S. A. (1975). Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. *Electroencephalography and Clinical Neurophysiology*, *38*, 387–401. https://doi.org/10.1016/0013-4694(75)90263-1

Sterne, J. A., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., … Higgins, J. P. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*, *343*, d4002. https://doi.org/10.1136/bmj.d4002

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

Su, H. I., & Freeman, E. W. (2009). Hormone changes associated with the menopausal transition. *Minerva Ginecologica*, *61*, 483–489.

Sun, J., Walker, A. J., Dean, B., van den Buuse, M., & Gogos, A. (2016). Progesterone: The neglected hormone in schizophrenia? A focus on progesterone-dopamine interactions. *Psychoneuroendocrinology*, *74*, 126–140. https://doi.org/10.1016/j.psyneuen.2016.08.019

Suwalska, A., & Lojko, D. (2014). Sex dependence of cognitive functions in bipolar disorder. *Scientific World Journal*, *2014*, 418432. https://doi.org/10.1155/2014/418432

Takeuchi, H., Taki, Y., Nouchi, R., Yokoyama, R., Kotozaki, Y., Nakagawa, S., … Kawashima, R. (2017). Global associations between regional gray matter volume and diverse complex cognitive functions: Evidence from a large sample study. *Scientific Reports*, *7*, 10014. https://doi.org/10.1038/s41598-017-10104-8

Thakkar, K. N., Congdon, E., Poldrack, R. A., Sabb, F. W., London, E. D., Cannon, T. D., & Bilder, R. M. (2014). Women are more sensitive than men to prior trial events on the Stop-signal task. *British Journal of Psychology*, *105*, 254–272. https://doi.org/10.1111/bjop.12034

Torniainen, M., Suvisaari, J., Partonen, T., Castaneda, A. E., Kuha, A., Perala, J., … Tuulio-Henriksson, A. (2011). Sex differences in cognition among persons with schizophrenia and healthy first-degree relatives. *Psychiatry Research*, *188*, 7–12. https://doi.org/10.1016/j.psychres.2010.11.009

Tschernegg, M., Neuper, C., Schmidt, R., Wood, G., Kronbichler, M., Fazekas, F., … Koini, M. (2017). FMRI to probe sex-related differences in brain function with multitasking. *PLoS One*, *12*, e0181554. https://doi.org/10.1371/journal.pone.0181554

Valera, E. M., Brown, A., Biederman, J., Faraone, S. V., Makris, N., Monuteaux, M. C., … Seidman, L. J. (2010). Sex differences in the functional neuroanatomy of working memory in adults with ADHD. *American Journal of Psychiatry*, *167*, 86–94.

Vaskinn, A., Sundet, K., Simonsen, C., Hellvin, T., Melle, I., & Andreassen, O. A. (2011). Sex differences in neuropsychological performance and social functioning in schizophrenia and bipolar disorder. *Neuropsychology*, *25*, 499–510. https://doi.org/10.1037/a0022677

Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences*, *12*, 418–424.

Verdejo-Garcia, A., Bechara, A., Recknor, E. C., & Perez-Garcia, M. (2006). Decision-making and the Iowa Gambling Task: Ecological validity in individuals with substance dependence. *Psychol. Belg.*, *46*, 55–78. https://doi.org/10.5334/pb-46-1-2-55

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*, 250–270. https://doi.org/10.1037/0033-2909.117.2.250

Weafer, J., & de Wit, H. (2014). Sex differences in impulsive action and impulsive choice. *Addictive Behaviors*, *39*, 1573–1579.

White, T. P., Loth, E., Rubia, K., Krabbendam, L., Whelan, R., Banaschewski, T., … … IMAGEN Consortium. (2014). Sex differences in COMT polymorphism effects on prefrontal inhibitory control in adolescence. *Neuropsychopharmacology*, *39*, 2560–2569. https://doi.org/10.1038/npp.2014.107

Yoest, K. E., Cummings, J. A., & Becker, J. B. (2014). Estradiol, dopamine and motivation. *Central Nervous System Agents in Medicinal Chemistry*, *14*, 83–89.

Yuan, J., He, Y., Qinglin, Z., Chen, A., & Li, H. (2008). Gender differences in behavioral inhibitory control: ERP evidence from a two-choice oddball task. *Psychophysiology*, *45*, 986–993. https://doi.org/10.1111/j.1469-8986.2008.00693.x

Zandara, M., Garcia-Lluch, M., Pulopulos, M. M., Hidalgo, V., Villada, C., & Salvador, A. (2016). Acute stress and working memory: The role of sex and cognitive stress appraisal. *Physiology & Behavior*, *164*, 336–344. https://doi.org/10.1016/j.physbeh.2016.06.022

Zhang, S., Hu, S., Bednarski, S. R., Erdman, E., & Li, C. S. (2014). Error-related functional connectivity of the thalamus in cocaine dependence. *NeuroImage Clinical*, *4*, 585–592. https://doi.org/10.1016/j.nicl.2014.01.015

Zhang, Y., Ide, J. S., Zhang, S., Hu, S., Valchev, N. S., Tang, X., & Li, C. R. (2017). Distinct neural processes support post-success and post-error slowing in the stop signal task. *Neuroscience*, *357*, 273–284. https://doi.org/10.1016/j.neuroscience.2017.06.011

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Gaillard A, Fehring DJ, Rossell SL. A systematic review and meta-analysis of behavioural sex differences in executive control. *Eur J Neurosci*. 2020;00:1–24. https://doi.org/10.1111/ejn.14946