



Review article

Neuropsychological functioning in adult anorexia nervosa: A meta-analysis

Kristin Stedal^{a,*}, Catherine Broomfield^b, Phillipa Hay^c, Stephen Touyz^b, Ronny Scherer^d^a Regional Department for Eating Disorders, Division of Mental Health and Addiction, Oslo University Hospital Ullevål HF, Postboks 4956 Nydalen, 0424, Oslo, Norway^b School of Psychology, Faculty of Science, The University of Sydney, NSW, 2006, Australia^c Translational Health Research Institute, Western Sydney University, Locked Bag 1797, Penrith, NSW, 2751, Australia^d Centre for Educational Measurement at the University of Oslo (CEMO), Faculty of Educational Sciences, University of Oslo, P.O. Box 1161 Blindern, 0318, Oslo, Norway

ARTICLE INFO

Keywords:

Anorexia nervosa
Neuropsychological function
Eating disorder
Meta-analysis

ABSTRACT

Several studies have conceptualized neuropsychological dysfunction as part of the core pathology and defining behaviors seen in the eating disorder anorexia nervosa (AN). The aim of the current review was to synthesize the differences in neuropsychological test performance between individuals with AN and healthy controls, quantify and explain their heterogeneity. The search and screening procedures resulted in fifty studies that comprised 186 neuropsychological test results. Utilizing random-effects meta-analyses, the results revealed evidence for significant, moderate underperformance in people with AN in overall neuropsychological functioning ($\bar{g} = -0.43$, 95 % CI [-0.50, -0.36]). Weighted mean effect sizes ranged from $\bar{g} = -0.53$ for visuospatial abilities to $\bar{g} = -0.10$ for planning. Study and participant characteristics, including body mass index (BMI) and age, had significant moderator effects, especially on executive function, memory, and visuospatial abilities. The findings from the current study provide an extensive and comprehensive overview of the possible impairments in neuropsychological functioning in adult patients diagnosed with AN.

1. Introduction

Anorexia nervosa (AN) is a severe mental illness characterized by dietary restriction leading to weight loss or a failure to gain weight, as well as body-shape and weight over-evaluation (American Psychiatric Association, 2013). It is associated with a high risk of premature death (Kask et al., 2016) and is recognized by significant concerns regarding body image and persistent efforts to lose weight despite being severely underweight. The classification, diagnosis and treatment of AN have traditionally focused on the behaviors and cognitions of patients. However, in recent years, researchers have increasingly targeted their studies towards a broader phenotypic and biological appreciation of phenomenology. One reason for emphasizing cognitive functioning as an area of interest is the association between inferior treatment outcome and poor neuropsychological functioning (Hamsher et al., 1981; Harper et al., 2017). In addition, an increased comprehension of the neuropsychological function of patients with AN has the potential to provide a better understanding of the cognitions and behaviors characterizing the illness – which could aid in diagnosis and treatment. It has also been suggested that neuropsychological deficits could be a trait marker, or endophenotype, for the disorder (Kanakam et al., 2013). Subsequently,

some of the core pathology and defining behaviors seen in AN have been conceptualized as a reflection of neuropsychological dysfunction. For example, body size estimation errors have been described as an expression of poor visuospatial abilities (Lang et al., 2016; Lang and Tchanturia, 2014), and cognitive and behavioral inflexibility have been considered a consequence of set shifting impairments (Shott et al., 2012; Steinglass et al., 2006). Consequently, more research has focused on neuropsychological functions, such as visuospatial processing and set shifting, compared to attention and inhibition (Smith et al., 2018).

Despite extensive research, however, findings have been inconsistent. Some studies have reported considerable cognitive deficits in patients with AN (e.g. Lopez et al., 2008; Tchanturia et al., 2004a; Weider et al., 2015), whereas other studies have failed to find a difference in cognitive function between patients and controls (e.g. Jones et al., 1991; Thompson, 1993; Øverås et al., 2017). These inconsistencies have been attributed to variable design, small sample sizes, heterogeneous samples and a failure to control for alternative explanations for test performance (Stedal, 2012; Tchanturia et al., 2005). This lack of coherence in the field has made it immensely challenging for clinicians and researchers to interpret findings from studies and to select tests for neuropsychological assessments. This is further highlighted in the recently published review

* Corresponding author.

E-mail address: kristin.stedal@ous-hf.no (K. Stedal).<https://doi.org/10.1016/j.neubiorev.2021.08.021>

Received 15 February 2021; Received in revised form 10 June 2021; Accepted 22 August 2021

Available online 25 August 2021

0149-7634/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

of reviews by Smith et al. (Smith et al., 2018). Out of 28 systematic reviews and meta-analyses, thirteen were based on patients with AN, and only one study (Zakzanis et al., 2010) explored a broad spectrum of cognitive functions in this patient group. The remaining studies investigated specific cognitive domains, such as set shifting (Westwood et al., 2016; Wu et al., 2014), executive functions (Hirst et al., 2017; Miles et al., 2020), decision-making (Guillaume et al., 2015; Wu et al., 2016), central coherence (Lang and Tchanturia, 2014), or attention bias (Aspen et al., 2013; Brooks et al., 2011). In addition, some previous meta-analyses have reported findings from self-reports (Miles et al., 2020), despite research showing a lack of association between performance-based neuropsychological tests and self-report measures (Herbrich et al., 2019; Stedal and Dahlgren, 2015). Consequently, there is a lack of meta-analyses providing an overall framework of cognitive function, based on traditional domain classifications (Lezak et al., 2004), using standardized neuropsychological tests.

One reason for focusing on specific domains when performing meta-analyses could be to avoid issues with dependency of effect sizes. Most primary studies investigating neuropsychological functioning in AN have assessed more than one cognitive domain and often report more than one relevant effect size for each domain. However, combining multiple effect sizes in one meta-analysis can be problematic. For example, a study investigating executive function might report scores from multiple tests which all assess executive functions or multiple scores from the same test. For traditional meta-analytic procedures this warrants concern, since a premise for the analyses is independence of effect sizes (Cheung, 2019). Until recently, the most common way of handling dependent effect sizes was to either average the effect sizes or to select only one effect size from each study (Cheung, 2019; Smith et al., 2018; Zakzanis et al., 2010) – or, in some cases, to just disregard the dependency of the data. This is concerning because “when effect sizes are not independent, conclusions based on these conventional procedures can be misleading or even wrong” (Cheung, 2019, p. 387). In addition, by selecting only one effect size from each study, there is a notable risk of selection bias in terms of which tests and/or domains are chosen, and it also limits the utilization of available data (Cheung, 2019). However, recent statistical advancements have led to the development of meta-analytical procedures which can address non-independent effect sizes. These procedures can provide more detailed information concerning both the direction and magnitude of difference between patients with AN and healthy controls on neuropsychological tests. To the authors’ knowledge, there are no studies which have applied these novel meta-analytical procedures to investigate neuropsychological functioning in patients with AN. In addition, most previous meta-analyses of neuropsychological function in AN have not taken into account possible confounding factors, including depression and anxiety, weight status, duration of illness and/or age (Smith et al., 2018).

1.1. The current meta-analysis

The literature on cognitive function in AN is inconsistent, and most previous research syntheses have focused on single rather than multiple cognitive domains and have included patients at different stages of the illness, including recovered participants. The latter challenges the interpretation of meta-analytic results, as some studies have shown improved cognitive function with weight gain during recovery (Hemmingsen et al., 2020). Further, some previous systematic reviews and meta-analyses have also combined results from different assessment methods, including self-report questionnaires, which further obfuscate the understanding of findings.

In the current meta-analysis, we focused on six major neuropsychological domains (i.e., attention, executive functions, memory, processing speed, visuospatial abilities, and working memory) and ten subdomains (see Table 1). These domains were aligned with the classification by Lezak et al. (2004) and previous studies of comparable

Table 1
Neuropsychological Domains, Subdomains, and Outcomes.

Domains	Subdomains	Outcome measure
Attention	Sustained attention	Go/No-go (omission errors), d2
	Planning	TOH, TOL, Brixton SAT
Executive functions	Response inhibition	CPT (commission errors), Go/No-go (commission errors), SST (commission errors), CWIT interference, Hayling SCT
	Set shifting/ cognitive flexibility	VFT, COWA, TMT-B, WCST, Berg CST
	Verbal memory	RAVLT, CVLT, AVLT, HVLTL, WMS logical memory
Memory	Non-verbal memory	RCFT recall
Processing Speed	Processing speed	CPT RT, SST RT, CWIT (congruent trial RT), TMT-A, WAIS digit symbol
	Visuospatial abilities	RCFT copy, WAIS block design
Visuospatial Abilities	Central coherence index	RCFT cci, RCFT organization, GEFT
	Working memory	WAIS digit span, WMS letter number sequencing

Note. RT = Reaction time; AVLT = Auditory Verbal Learning Test; ; CCI = Central coherence index; COWA = Controlled Oral Word Association test; CPT = Continuous Performance Test; CST = Card Sorting Test; CVLT = California Verbal Learning Test; CWIT = Color Word Inference Test; GEFT = Group Embedded Figures Test; HVLTL = Hopkins Verbal Learning Test; RAVLT = Rey Auditory Verbal Learning Test; RCFT = Rey Complex Figure Test; SAT = Spatial Anticipation Test; SCT = Sentence Completion Test; SST = Stop Signal Task; TMT = Trail Making Test; TOH = Tower of Hanoi; TOL = Tower of London; VFT = Verbal Fluency Test; WAIS = Wechsler Adult Intelligence Scale; WCST = Wisconsin Card Sorting Test; WMS = Wechsler Memory Scale.

patient populations (Abramovitch et al., 2013; Geller et al., 2018). The included tests were classified “according to the major functional activities they elicit” (Lezak et al., 2004, p. 335).

The primary aim of the current review was to synthesize previously published data to examine the magnitude of difference on neuropsychological tests between individuals with AN and healthy controls. In addition, since the majority of previous studies have not accounted for factors which can potentially influence test performance (Smith et al., 2018), our second aim was to assess the moderation of study and sample characteristics on neuropsychological test performance. These characteristics included, but were not limited to, AN diagnostic subtype, participants’ average age, body mass index (BMI), years of education, performance on intelligence tests, and eating disorder severity.

2. Method

2.1. Literature search

We conducted the literature search in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Moher et al., 2009). The literature search was performed in July 2019, subsequently updated in May 2020, and checked again in September 2020. We restricted the search to the databases MEDLINE, PsycINFO, ISI Web of Science, and Epistemonikos to provide an exhaustive record and documentation from these key databases which also indicate some degree of documented quality. The databases may include both published and grey literature. No further databases were searched to avoid additional duplicates and data was not drawn from other sources. A librarian at the Medical Library of the Oslo University Hospital conducted the search using the following search terms: ‘anorexia nervosa’, cross-referenced with the terms ‘neuropsych*’, ‘neurocog*’, ‘executive function’, ‘memory’, ‘processing speed’, ‘visuospatial’, ‘inhibition’, ‘planning’, ‘attention’, ‘set shifting’, ‘central coherence’, ‘flexibility’, ‘rigidity’. Terms were searched for as Medical Subject Headings (MeSH, MEDLINE) or Thesaurus of Psychological Index Terms (PIT, PsycINFO), as well as in titles and abstracts. We

limited the search to publications in English. The full search strategy for all databases, including the corresponding limits, can be found in Appendix A in Supplementary materials.

2.2. Screening procedures

The literature search resulted in 5023 titles, which were reduced to 3313 after removing duplicates. These publications were then submitted to the screening of titles and abstracts according to the inclusion and exclusion criteria specified below. A detailed overview of the screening process is presented in Fig. 1.

2.2.1. Inclusion criteria

Published studies investigating neuropsychological functioning in adult patients currently diagnosed with AN, based on criteria from the Diagnostic and Statistical Manual of Mental Disorders (DSM), 3rd edition or newer (Association, 1987), were considered for inclusion based on the following a-priori set criteria:

- (1) At least one comparison on one or more neuropsychological tests between current DSM-diagnosed (i.e. via structured interview) adult (≥ 18 years) patients with AN and a healthy (i.e. screened for absence of psychiatric or neurologic diagnosis) adult (≥ 18 years) control group was conducted.
- (2) Studies evaluated one or more of the following neuropsychological domains: Attention, executive functions, memory, processing speed, visuospatial abilities (including central coherence), and working memory.
- (3) Studies were published in English or had an available English translation.

2.2.2. Exclusion criteria

Studies were excluded if they lacked a healthy control group, if they only reported comparisons between patient groups (e.g., patients with AN compared to patients with depression), if the assessment was done within-subject (e.g., pre/post treatment), or if the study was a single-

group investigation. Treatment studies were included if it encompassed a neuropsychological pre-treatment comparison of patients with AN and healthy controls. To ensure the validity of findings, studies were excluded if they did not use validated, traditional, and standardized neuropsychological tests – as determined by experts in the field (Lezak et al., 2004). This included tests of decision making and outcome measures not considered a part of the original tests. Studies using modified versions of the original tests (e.g., Emotional Color Word Interference Test), tests administered during brain scans, or tests rarely used ($< 1\%$ of the studies) were also excluded. Organizational scores on the Rey Complex Figure Test were included, due to the large amount of research assessing organizational strategy in AN (Lang et al., 2016; Lang and Tchanturia, 2014). Books, book chapters, editorials, commentaries, reviews, theses, conference abstracts, errata, and studies presenting data where diagnostic screening was unclear were omitted. When the same dataset was reported in multiple studies, only the original article or the one with the most complete report of the relevant information was included.

2.2.3. Title screening

In this first screening process, we removed titles corresponding to studies which were obviously not eligible (e.g., “Meningioma and psychiatric symptoms: An individual patient data analysis.”). Equivocal (e.g., “Cortisol levels and vigilance in eating disorder patients”) and plausible titles (e.g., “Exploring the neurocognitive signature of poor set shifting in anorexia and bulimia nervosa”) were retained. To assess interrater agreement, two authors (KS and CB) performed an initial screening on a subset of publication titles ($n = 101$). Titles were labelled “include”, “exclude” or “inconclusive”. The resultant consistency for title extraction indicated a substantial agreement between the two raters ($\kappa = 0.80, p < .001$), according to the well-established guidelines (Landis and Koch, 1977; Viera and Garrett, 2005). The second author (CB) reviewed the remaining titles. After title screening, 2914 out of the 3313 publications were omitted, and 400 titles were retained.

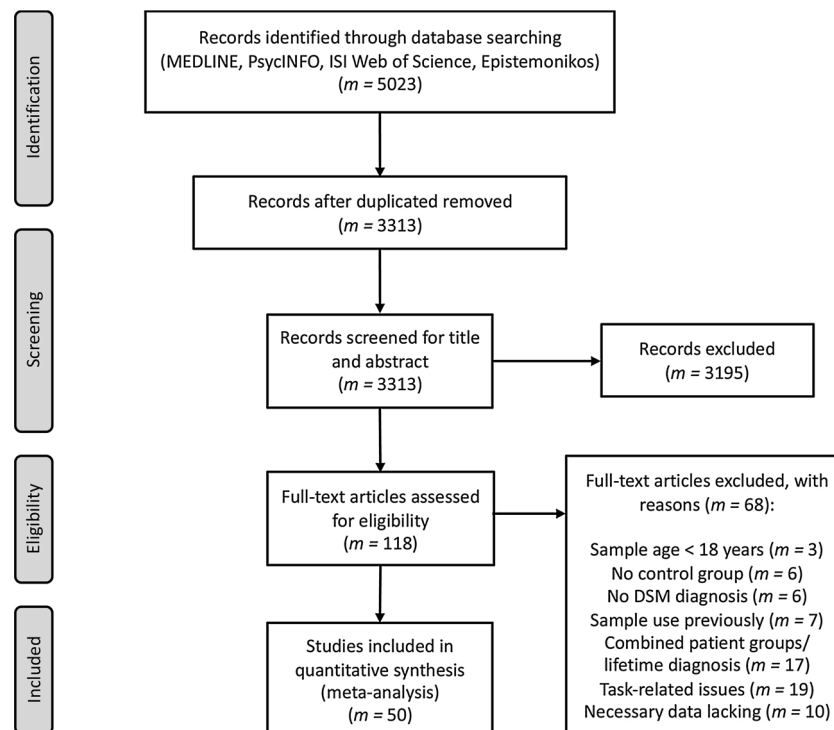


Fig. 1. PRISMA Flow Diagram of the Search, Screening, and Inclusion Processes.
Note. m = Number of studies.

2.2.4. Abstract screening

Two authors (KS and CB) screened the abstracts of the retained studies for eligibility and classified them as “include” or “exclude”. The consistency for abstract extraction on 49 % of the reviewed studies ($m = 196$) indicated almost perfect agreement between the two raters, $\kappa = 0.87$, $p < .001$. After reviewing the 400 abstracts, 118 studies were retained for full-text review and were screened by the first author. Sixty-eight of these studies were excluded for the following reasons: Not providing a DSM diagnosis ($m = 6$), investigating a population < 18 years old ($m = 3$), the results from the sample had been presented in a previous study ($m = 7$), combining patient groups (e.g., only results from AN and bulimia nervosa combined were presented), investigating patients with a lifetime diagnosis ($n = 17$), lacking necessary data to compute effect sizes ($m = 10$), or a healthy control group ($m = 6$). Finally, studies were excluded for utilizing modified/experimental tests or tests were performed in scanner ($m = 13$), or the test was not included in the current meta-analysis ($m = 6$). A total of 50 studies were included. The authors of the texts and their affiliations were disclosed in the screening and extraction processes.

2.3. Coding of primary studies and effect size measures

The status of the included and eligible studies is that of July 6, 2020. Extraction from the included studies followed recommended coding procedures (Valentine, 2009) and the coding scheme presented by Abramovitch et al. (2013). Variables were coded as “participant characteristics”, “study characteristics” and “validity and reliability assessments”. The first and second author (KS, CB) extracted the following information from the included studies: (a) Publication status, (b) publication year, and (c) the country in which the study was conducted. The subsequent participant characteristics were recorded: (a) Sample sizes for both groups (AN and healthy controls), (b) mean age (in years), (c) mean BMI (d) mean age of AN onset (in years), (e) mean duration of illness (in months), (f) years of education, (g) percentage of males in the AN group, (h) mean score on measures of AN severity (e.g., Eating Disorder Examination Questionnaire), depression severity (e.g., Beck Depression Inventory), and anxiety severity (e.g., State-Trait Anxiety Inventory), (i) percentage of AN participants with Axis I comorbid illnesses, and (j) the percentage of AN participants receiving serotonin reuptake inhibitors, neurotropic, or neuroleptic medication. Furthermore, we recorded study characteristics, including the specific neuropsychological test used and the associated domains and subdomains of functioning. Reported outcomes for neuropsychological test performance were extracted as means and standard deviations. Table 1 presents the domains, subdomains, and outcomes coded from the studies. For cases where the outcome variables were uncommon (e.g., “Time to first move” on the Tower Test), only the conventional outcome variables were recorded. Finally, validity and reliability assessments were recorded as (a) the number of tests, (b) number of testing sessions, and (c) the average length of testing sessions. When studies included more than one measure, either within the same domain or for multiple domains, we extracted all relevant outcomes instead of selecting only one. All outcome variables from the neuropsychological assessments were coded so that positive scores indicated better performance. Variations of the same or similar tests were grouped together. For instance, the Rey Auditory Verbal Learning Test, the California Verbal Learning Test, and the Auditory Verbal Learning Tests were all considered to be “verbal learning tests”. Similarly, the Tower of London and Tower of Hanoi tasks were considered to be “tower tests”.

Excluding author and publication information, a total of 20 variables were coded for each study. To assess the coding reliability, a random sample of studies was coded by two authors (KS and CB) on all variables (10 %, $m = 8$). A total of 160 variables were compared. The results revealed discrepancies between coders on only six variables, indicating high interrater agreement (96 %) – these discrepancies were resolved through discussion.

Given the means and standard deviations extracted from the primary studies for both the AN and the control group, we calculated Hedges’ g from the standardized mean difference ES , following Borenstein et al.’s (2009) procedure. Specifically, with \bar{X}_{AN} and \bar{X}_{HC} denoting the group-specific mean scores of some neuropsychological test, SD_{AN} and SD_{HC} the corresponding standard deviations, and N_{AN} and N_{HC} the sample sizes, Hedges’ g and its elements were calculated as follows:

$$ES = \frac{\bar{X}_{AN} - \bar{X}_{HC}}{SD_{Pooled}}$$

$$SD_{Pooled} = \sqrt{\frac{(N_{AN} - 1)SD_{AN}^2 + (N_{HC} - 1)SD_{HC}^2}{N_{AN} + N_{HC} - 2}}$$

$$g = \left[1 - \frac{3}{4(N_{AN} + N_{HC} - 2) - 1} \right] \cdot ES$$

The corresponding sampling variance v_g and the standard error SE_g were then calculated as:

$$v_g = \left(1 - \frac{3}{4(N_{AN} + N_{HC} - 2) - 1} \right)^2 \cdot \left(\frac{N_{AN} + N_{HC}}{N_{AN} \cdot N_{HC}} + \frac{ES^2}{2(N_{AN} + N_{HC})} \right)$$

$$SE_g = \sqrt{v_g}$$

Given this specification of Hedges’ g , negative effect sizes indicated a disadvantage of participants in the AN group in their performance on some neuropsychological test over the healthy control group. The resultant effect sizes for each primary study are displayed in Appendix B in Supplementary materials.

2.4. Quality assessment

On the basis of factors which may influence performance on neuropsychological tests (Lezak et al., 2004; Yang et al., 2018), we assessed the included studies for methodological quality and assigned quality scores to them. These scores ranged from 0 to 7, with higher scores indicating better precision of the neuropsychological test results (Yang et al., 2018). In line with the quality rating developed by Yang et al. (2018), study quality was calculated as follows: (age difference excluded [0–1]) + (gender difference excluded [0–1]) + (education difference excluded [0–1]) + (intelligence quotient [IQ] difference excluded [0–1]) + (depression excluded [0–1]) + (anxiety excluded [0–1]) + (medication excluded [0–1]). For instance, studies which accounted for (i.e. matched for) differences in age, gender, education, IQ, and medication, as well as controlling for depression and anxiety, received a score of 7.

2.5. Data analysis

2.5.1. Meta-analytic baseline models

As a first step, we synthesized the effect sizes for the overall sample of neuropsychological test and, subsequently, for each of the domains, subdomains, and tests. Given that the structure of the meta-analytic data was inherently hierarchical with multiple effect sizes per study, the independence assumption clearly did not hold (Borenstein et al., 2009). The extant literature has proposed several procedures to account for these dependencies, such as averaging multiple effect sizes per study, robust variance estimation, or multilevel random-effects modeling with or without correlated effects (Cheung, 2019; Fernández-Castilla et al., 2020; Pustejovsky and Tipton, 2021). In the current review, we performed multilevel random-effects modeling to quantify the different variance components explicitly. For instance, the three-level random-effects model quantifies the sampling variance (level 1), the variance between effect sizes within studies (level 2, variance $\tau_{(2)}^2$), and the variance of effect sizes between studies (level 3, variance $\tau_{(3)}^2$). Such a

model accounts efficiently for the dependence of effect sizes and allows researchers to test different assumptions on the variance components (Cheung, 2013). Specifically, for a given meta-analytic data set with a nested structure, the variance components can be tested against zero via model comparisons (e.g., based on information criteria and likelihood-ratio tests). However, these significant tests are performed against the boundary estimate of zero—hence; the confidence intervals of the variances should also be considered. Synthesizing the effect sizes, we tested and compared several models with different variance components (i.e., a three-level random-effects model, random-effects models with variances either between studies or effect sizes, and a fixed-effects model) to establish baseline models. These models provided the weighted mean effect sizes for neuropsychological functioning in general and the (sub-)domains specifically, next to the heterogeneity indices ($I^2_{(2)}$ and $I^2_{(3)}$) and variance components (Cheung, 2013). Moreover, we extended the baseline models to mixed-effects meta-regression models to test for moderator effects.

2.5.2. Sensitivity analyses

To establish the robustness of our findings, we examined the sensitivity of the meta-analytic results across several conditions: (a) Type of variance estimation: Restricted maximum-likelihood (REML) estimation vs. Bayesian estimation, (b) Treatment of effect size dependencies: Robust variance estimation vs. multilevel meta-analysis with or without constant sampling correlation, (c) Handling influential effect sizes: Exclusion vs. inclusion, and (d) Treating missing data in the continuous moderators: Pairwise deletion vs. multiple multilevel imputation. Both the analytic code and the results of these analyses are documented in the Supplementary Material S1–S3.

2.5.3. Moderator analyses

As a second step, we tested the possible moderating effects of the study and sample characteristics, specifying and estimating mixed-effects models with the continuous and categorical moderators (Cheung, 2013). The moderator variables in the current meta-analysis were either related to the study (e.g., country, publication year, control for depression/anxiety, study quality score), to all the participants (e.g. BMI, IQ, age, years of education), or were specific for the patient sample (e.g., duration of illness, clinical severity). Participant moderator variables are presented in Table 2. Moderator effects for categorical variables were only considered if at least six to seven effect sizes were

Table 2
Sample Characteristics and Moderators.

Study characteristics	Mean (SD)	Range	<i>m</i>
Sample size in AN group	34.2 (27.0)	10–171	50
Sample size in HC group	43.8 (35.7)	10–199	50
AN age	25.7 (5.4)	18.8–58.8	50
HC age	25.4 (3.7)	18.6–34.9	50
AN BMI	15.8 (1.1)	13.7–19.0	46
HC BMI	21.7 (0.7)	20.2–23.2	44
AN education (years)	13.5 (1.3)	11.8–16.9	28
HC education (years)	14.9 (2.1)	12.0–21.5	28
AN IQ	106.0 (52.7)	96.1–115.8	19
HC IQ	109.4 (5.1)	100.2–119.7	19
AN illness onset age (years)	17.3 (1.7)	14.3–2.8	15
AN duration of illness (years)	6.7 (2.9)	1.7–13.1	30
AN with comorbidity (%)	26.8 (29.8)	0–77	15
AN on psychotropic medication (%)	30.1 (24.1)	0–74	25
EDI-2 scores in AN group	116.9 (77.1)	76.9–305.1	8
EDE-Q scores in AN group	3.5 (0.9)	2.1–5.0	11
BDI-I/II scores in AN group	23.5 (8.4)	11.8–39.1	18
STAI-T scores in AN group	52.8 (10.3)	35.3–62.7	9

Note. *m* = Number of studies, AN = Anorexia nervosa, HC = Healthy control, BMI = Body mass index (kg/m²), IQ = Intelligence quotient, EDI-2 = Eating Disorders Inventory-second edition, EDE-Q = Eating Disorders Examination Questionnaire, BDI-I/II = Becks Depression Inventory-version I or II, STAI-T = State-Trait Anxiety Inventory-Trait.

available per category (Rubio-Aparicio et al., 2017; Tipton et al., 2019). For moderators with many levels (e.g., countries, outcomes), we implemented the moderator variable as an additional clustering variable that indicated an explicit level of analysis. The respective models were specified as either four-level random-effects models (Fernández-Castilla et al., 2020) or cross-classified random-effects models (Fernández-Castilla et al., 2019), depending on the type of hierarchical data structure. For instance, we tested the possible differences in effects between the specific outcomes of the neuropsychological tests using a cross-classified model with variance components: Sampling variation (level 1), variation between effect sizes within studies (level 2), variation of effect sizes between studies (level 3), and variation between outcomes (level 4). While level 2 is hierarchically nested in level 3 in this example, level 4 represents a level of analysis that is independent of levels 2 and 3 (Fernández-Castilla et al., 2019). In contrast, we tested the moderator effects of countries using a four-level model with full hierarchical nesting, assuming that studies were directly nested in countries. All analyses, including the sensitivity and baseline model analyses, were performed in the R packages ‘metafor’ (Viechtbauer, 2010), ‘robmeta’ (Fisher et al., 2017), and ‘brms’ (Bürkner, 2017).

2.5.4. Publication bias, file-drawer issues, and influential effect sizes

To examine possible publication bias and file-drawer issues, we conducted several analyses: First, we conducted trim-and-fill analyses, evaluated the symmetry of the funnel plots (Duval and Tweedie, 2000), and performed Begg’s rank correlation test (Begg and Mazumdar, 1994). The trim-and-fill analyses have recently been extended to multilevel meta-analyses, resulting in the two estimates L_0^+ and R_0^+ as indicators of the number of missed effect sizes (Fernández-Castilla et al., 2021). We further tested the asymmetry of these plots via Egger’s linear regression test (Egger et al., 1997). Second, using Rosenberg’s procedure, we estimated the fail-safe *N*’s (Borenstein et al., 2009). Third, we performed the funnel plot test and evaluated the precision-effect estimate with standard error (PEESE; Fernández-Castilla et al., 2021). Fourth, we plotted the *p*-curves underlying the effects and examined their skewness (Simonsohn et al., 2014). Specifically, if a *p*-curve was right-skewed, the primary studies had evidential value, providing evidence against *p*-hacking. We used the R package ‘dmetar’ to obtain the *p*-curves (Harrer et al., 2019). Finally, we identified influential effect sizes using Viechtbauer and Cheung’s (2010) diagnostics in the R package ‘metafor’. All of these analyses and their outcomes are documented in the Supplementary Material S1.

3. Results

3.1. Description of the primary studies

A total of *m* = 50 primary studies, *k* = 186 effect sizes, and *N* = 4057 participants were included. The sample was comprised of *n* = 1778 participants diagnosed with AN and *n* = 2279 healthy controls. Notably, only two studies (Weider et al., 2014; Talbot et al., 2015) included 5 % and 4.2 % men. As noted earlier, all other studies, except for Tchanturia et al. (2002) who only reported the matching for gender, were based on female samples. The included studies were conducted in the following countries: Italy (*m* = 11), United Kingdom (*m* = 9), United States of America (*m* = 4), Spain (*m* = 7), The Netherlands (*m* = 2), Republic of South Korea (*m* = 1), Australia (*m* = 4), Germany (*m* = 3), Norway (*m* = 2), Canada (*m* = 1), Mexico (*m* = 1), Japan (*m* = 2), Argentina (*m* = 1), Belgium (*m* = 1), and France (*m* = 1). The six core domains comprised varying numbers of effect sizes: Attention (*k* = 6), executive functions (*k* = 74), memory (*k* = 38), processing speed (*k* = 30), visuospatial abilities (*k* = 29), and working memory (*k* = 7). Notably, the studies including some test of executive functions dominated the meta-analytic sample (*m* = 39 out of 50). An overview of the included studies is presented in the Appendix B in Supplementary materials (references included in

meta-analysis can be found in Appendix C in Supplementary materials), and Table 2 shows sample characteristics.

3.2. Overall effect size and moderator analyses

Combining all effect sizes across the neuropsychological functioning domains, we established a three-level random-effects model as the baseline model to report an overall effect size (see Supplementary Material S1). This model resulted in a moderate, negative, and statistically significant effect size ($\bar{g} = -0.431$, 95 % CI [-0.503, -0.359]) and indicated significant heterogeneity ($Q_E [37] = 98.4, p < .001$). The corresponding heterogeneity indices suggested moderate heterogeneity within studies and small heterogeneity between studies, and so did the variance components (0.095 and 0.014, respectively; see Table 3).

The subsequent moderator analyses revealed a marginal age effect with older participants exhibiting larger negative effect sizes ($B = -0.064, SE = 0.035, Q_M [1] = 3.3, p = .07$) and a positive BMI effect with higher average BMI in the sample resulting in the less negative effect sizes ($B = 0.075, SE = 0.036, Q_M [1] = 4.4, p = .04$). We neither found any significant moderator effects between the AN subgroups (i.e., AN, AN-restrictive, AN-binge/purge) nor with respect to other sample characteristics and study quality. More recently published studies presented more positive effect sizes (see Supplementary Material S1). However, some dependencies on the publication year were observed, with more positive effect sizes for more recently published studies ($B = 0.078, SE = 0.037, Q_M [1] = 4.4, p = .04$, for the grand-mean centered variable indicating the publication year). Finally, we examined whether significant variance existed across the different neuropsychological outcomes. The cross-level random-effects model resulted in a significant variance component across outcomes ($\tau_{(4)}^2 = 0.054$, 95 % CI [0.023, 0.116]) and outperformed the three-level random-effects model without outcomes ($\chi^2 [1] = 20.7, p < .001$). These findings testify to the significance of the differences between outcomes. Similarly, we examined possible between-country variation—however, there was no evidence

supporting such variation ($\chi^2 [1] = 0.6, p = .44$). Finally, using executive functions as the reference category, we further examined the differences between domains. The corresponding mixed-effects model revealed significantly smaller (i.e., more negative) effect sizes for memory ($B = -0.196, SE = 0.088, p = .03$) and visuospatial abilities ($B = -0.187, SE = 0.093, p = .04$).

The overall effect size provides a reference point for the more detailed analyses of effects for each domain and subdomain. Moreover, as the respective moderator analyses suggested that the effects varied between domains and outcomes, we further performed domain- and outcome-specific analyses.

3.3. Effect sizes and moderator analyses per cognitive domain and subdomain

In the following, we present the weighted mean effect sizes for each of the categories of cognitive domains, subdomains, and outcomes, along with the moderator analyses. Given that this differentiation limits the sample sizes available to meta-analyses, we evaluated the baseline models for each of these categories. Tables 3 and 4 exhibit the effect sizes, and Fig. 2 displays the corresponding forest plot.

3.3.1. Attention

Our meta-analytic sample provided six effect sizes in the domain of attention. On the basis of a fixed-effects models, we obtained a weighted mean effect size of $\bar{g} = -0.571$ (95 % CI [-0.791, -0.351]). Given the limited number of effects, we refrained from conducting moderator analyses. However, the effect sizes within this domain ranged substantially from $\bar{g} = -0.962$ (Go/No-go omission errors) to $\bar{g} = -0.439$ (Go/No-go commission errors).

3.3.2. Executive functions

Examining the effect sizes for executive functions, we did not observe significant differences between subdomains. Specifically, setting set shifting/cognitive flexibility as the reference, a mixed-effects model did

Table 3
Weighted mean effect sizes and heterogeneity tests by neuropsychological domains and subdomains.

Domains and subdomains	<i>k</i>	<i>m</i>	\bar{g}	<i>SE</i>	Lower 95 % CI	Upper 95 % CI	<i>Z</i>	<i>Sig.</i>	<i>Q_E</i>	<i>p(Q_E)</i>	$\tau_{(2)}^2$	$\tau_{(3)}^2$	<i>I₍₂₎</i>	<i>I₍₃₎</i>	Baseline model
Overall	184	50	-0.431	0.037	-0.502	-0.359	-11.8	< .001	516.8	< .001	0.095	0.014	56.8	8.6	REM3
Attention	6	3	-0.571	0.112	-0.791	-0.351	-5.1	< .001	4.5	.48	0.000	0.000	0.0	0.0	FEM
Executive functions	74	39	-0.343	0.052	-0.444	-0.242	-6.7	< .001	274.8	< .001	0.132	0.000	73.0	0.0	REM2b
Planning	3	3	-0.103	0.172	-0.441	0.234	-0.6	.55	0.5	.77	0.000	0.000	0.0	0.0	FEM
Response inhibition	13	13	-0.194	0.096	-0.382	-0.006	-2.0	.04	30.8	< .01	0.068	0.000	60.6	0.0	REM2a
Set shifting/cognitive flexibility	58	34	-0.385	0.061	-0.504	-0.266	-6.3	< .001	223.7	< .001	0.149	0.000	75.1	0.0	REM2b
Memory	38	17	-0.486	0.108	-0.698	-0.273	-4.5	< .001	98.4	< .001	0.152	0.000	64.7	0.0	REM2a
Verbal memory	24	8	-0.482	0.200	-0.874	-0.091	-2.4	.02	77.1	< .001	0.278	0.000	74.5	0.0	REM2a
Non-verbal memory	14	11	-0.496	0.103	-0.698	-0.294	-4.8	< .001	20.9	.08	0.059	0.000	46.2	0.0	REM2b
Processing speed	30	19	-0.390	0.071	-0.530	-0.250	-5.5	< .001	56.1	< .01	0.053	0.000	48.9	0.0	REM2a
Visuospatial abilities	29	18	-0.528	0.068	-0.661	-0.396	-7.8	< .001	58.6	< .001	0.065	0.000	51.2	0.0	REM2b
Visuospatial abilities	17	14	-0.407	0.099	-0.601	-0.213	-4.1	< .001	35.4	< .01	0.072	0.000	52.2	0.0	REM2a
Central coherence	12	12	-0.647	0.084	-0.812	-0.482	-7.7	< .001	16.5	.13	0.024	0.000	29.6	0.0	REM2a
Working memory	7	5	-0.455	0.186	-0.818	-0.091	-2.5	.01	12.3	.06	0.098	0.000	55.7	0.0	REM2a

Note. *k* = Number of effect sizes, *m* = Number of studies, \bar{g} = Weighted mean effect size (Hedges' *g*), *Q_E* = Cochran's *Q*, $\tau_{(2)}^2$ = Variance at level 2, $\tau_{(3)}^2$ = Variance at level 3, *I₍₂₎* = Heterogeneity index for level 2, *I₍₃₎* = Heterogeneity index for level 3, FEM = Fixed-effects model, REM2a = Random-effects model (level 2: studies), REM2b = Random-effects model (level 2: effect sizes), REM3 = Three-level random-effects model (level 2: effect sizes, level 3: studies). All models specify the sampling variation at level 1.

Table 4
Outcome-specific weighted mean effect sizes within the subdomains.

Subdomains/Outcomes	k	\bar{g}	SE	Lower 95 % CI	Upper 95 % CI	Z	Sig.	Q_E	$p(Q_E)$	τ^2	I^2
Attention											
d2	3	-0.548	0.156	-0.853	-0.243	-3.5	< .001	0.4	.81	0.000	0.0
Go/No-Go commission errors	1	-0.439	0.256	-0.940	0.063	-1.7	.09	1.7	.19	0.055	41.7
Go/No-Go omission errors	2	-0.962	0.290	-1.531	-0.393	-3.1	< .01				
Executive functions											
<i>Planning</i>											
Tower tests	3	-0.103	0.172	-0.441	0.234	-0.6	.55	0.5	.77	0.000	0.0
<i>Response inhibition</i>											
CPT commission errors	1	-0.482	0.244	-0.959	-0.004	-2.0	.05				
CWIT interference	6	-0.050	0.148	-0.340	0.239	-0.3	.73	13.0	.02	0.078	62.6
Go/No-Go commission errors	1	-0.056	0.267	-0.579	0.468	-0.2	.84				
Hayling	4	-0.463	0.108	-0.674	-0.251	-4.3	< .001	2.3	.52	0.000	0.0
SST commission errors	1	0.021	0.145	-0.263	0.305	0.1	.89				
<i>Set shifting/cognitive flexibility</i>											
Berg CST perseverative errors	2	-1.168	0.273	-1.703	-0.633	-4.3	< .001	0.0	.86	0.000	0.0
Brixton	1	-0.762	0.245	-1.242	-0.281	-3.1	< .01				
TMT B	13	-0.595	0.140	-0.869	-0.322	-4.3	< .001	50.3	< .001	0.180	75.1
VFT categories	6	-0.015	0.097	-0.206	0.176	-0.2	.88	2.5	.78	0.000	0.0
VFT phonemic	8	0.180	0.120	-0.056	0.415	1.5	.14	10.4	.17	0.042	37.8
WCST perseverative errors	19	-0.420	0.089	-0.594	-0.246	-4.7	< .001	54.1	< .001	0.093	67.8
WCST perseverative responses	9	-0.535	0.087	-0.706	-0.364	-6.1	< .001	14.3	.07	0.027	42.0
Memory											
<i>Verbal memory</i>											
VLT delayed recall	7	-0.207	0.157	-0.516	0.101	-1.3	.19	10.5	.11	0.069	41.1
VLT immediate recall	3	-0.152	0.181	-0.507	0.204	-0.8	.40	1.4	.51	0.000	0.0
VLT total recall	6	-0.328	0.225	-0.769	0.113	-1.5	.15	14.2	.01	0.197	67.0
WMS logical memory delayed recall	4	-1.052	0.186	-1.416	-0.688	-5.7	< .001	4.7	.20	0.044	32.0
WMS logical memory immediate recall	4	-1.101	0.152	-1.399	-0.803	-7.2	< .001	3.0	.39	0.000	0.0
<i>Non-verbal memory</i>											
RCFT delayed recall	7	-0.561	0.147	-0.849	-0.274	-3.8	< .001	11.9	.07	0.073	49.4
RCFT immediate recall	7	-0.400	0.102	-0.601	-0.200	-3.9	< .001	7.9	.25	0.009	11.9
Processing speed											
CPT reaction time	2	-0.273	0.180	-0.627	0.081	-1.5	.13	0.0	.88	0.000	0.0
CWIT congruent trial	3	-0.330	0.294	-0.907	0.247	-1.1	.26	10.1	< .01	0.194	77.3
SST reaction time	2	-0.365	0.128	-0.616	-0.114	-2.9	< .01	0.2	.65	0.000	0.0
TMT A	18	-0.391	0.090	-0.568	-0.214	-4.3	< .001	38.1	< .01	0.079	57.5
WAIS digit-symbol	5	-0.620	0.113	-0.842	-0.398	-5.5	< .001	0.9	.93	0.000	0.0
Visuospatial abilities											
<i>Visuospatial abilities</i>											
RCFT copy	10	-0.251	0.103	-0.452	-0.050	-2.5	.01	14.2	.12	0.036	36.1
WAIS block design	7	-0.660	0.126	-0.908	-0.413	-5.2	< .001	9.5	.15	0.039	35.7
<i>Central coherence</i>											
GEFT	3	-0.731	0.151	-1.027	-0.436	-4.9	< .001	0.3	.85	0.000	0.0
RCFT cci	9	-0.639	0.112	-0.858	-0.419	-5.7	< .001	15.5	.05	0.052	48.5
Working memory											
Digit span	4	-0.626	0.183	-0.985	-0.267	-3.4	< .01	4.5	.22	0.050	37.7
WMS letter number sequencing	3	-0.309	0.274	-0.845	0.227	-1.1	.26	5.8	.06	0.144	64.7

Note. k = Number of effect sizes, \bar{g} = Weighted mean effect size (Hedges' g), Q_E = Cochran's Q , τ^2 = Between-study variance, I^2 = Heterogeneity index; RT = Reaction time; AVLT = Auditory Verbal Learning Test; Brixton = Brixton Spatial Anticipation Test; CCI = Central coherence index; COWA = Controlled Oral Word Association Test; CPT = Continuous Performance Test; CST = Card Sorting Test; CVLT = California Verbal Learning Test; CWIT = Color Word Inference Test; GEFT = Group Embedded Figures Test; HVLT = Hopkins Verbal Learning Test; RAVLT = Rey Auditory Verbal Learning Test; RCFT = Rey Complex Figure Test; SST = Stop Signal Task; TMT = Trail Making Test; TOH = Tower of Hanoi; VFT = Verbal Fluency Test; WAIS = Wechsler Adult Intelligence Scale; WCST = Wisconsin Card Sorting Test; WMS = Wechsler Memory Scale.

not reveal subdomain differences ($Q_M [2] = 2.5, p = .29$). The overall effect size for executive functions was $\bar{g} = -0.343$ (95 % CI [-0.444, -0.242]).

3.3.2.1. Planning. Only three effect sizes were available to synthesize the effects for the domain of planning on the basis of a fixed-effects model (Table 3). The weighted mean effect size was $\bar{g} = -0.104$ (95 % CI [-0.441, 0.234]) and did not significantly differ from zero ($Q_E [2] = 2.5, p = .55$). No further moderator analyses were conducted, and all effect sizes were obtained from one type of outcome (i.e., "Tower tests"; see Table 4).

3.3.2.2. Response inhibition. The weighted mean effect size for response inhibition was small ($\bar{g} = -0.194$, 95 % CI [-0.382, -0.006]), with patients performing significantly worse than the control participants ($Q_E [12] = 30.8, p < .01$), and a moderate to high heterogeneity (60.6 %; Table 3). Subsequent moderator analyses revealed a marginal difference between two AN subgroups (more negative effect size for the AN-restrictive subgroup than the AN subgroup; $B = -0.329, SE = 0.182, Q_M [1] = 3.3, p = .07$). Moreover, we found a positive moderator effect of age (the higher the weighted average age, the less negative the effect size; $B = 0.198, SE = 0.099, Q_M [1] = 4.1, p = .04$), BMI (the higher the weighted average BMI, the less negative the effect size; $B = 0.253, SE = 0.065, Q_M$

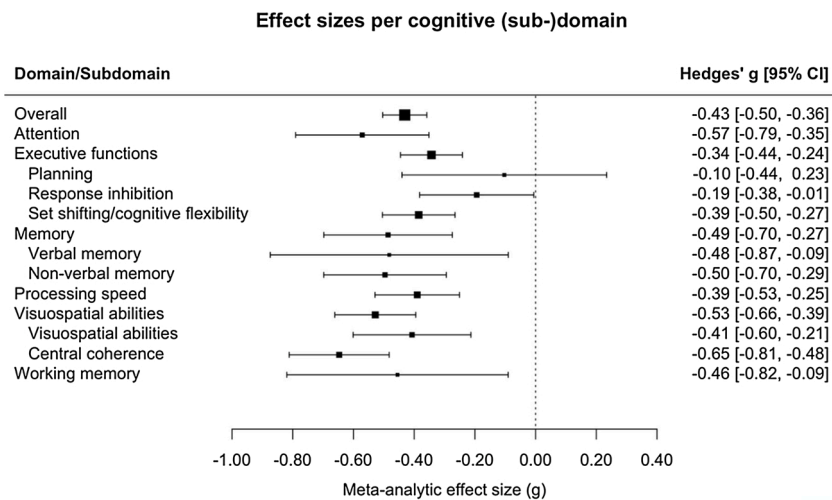


Fig. 2. Forest Plot Showing the Weighted Mean Effect Sizes per Cognitive Domain and Subdomain.

[1] = 15.0, $p < .01$), and years of education (the higher the weighted average years of education, the less negative the effect size; $B = 0.184$, $SE = 0.090$, $Q_M [1] = 4.2$, $p = .04$). Five different tasks were used to assess response inhibition (see Table 4), of which the Color Word Interference Task was most commonly administered ($k = 6$). No evidence for significant differences between outcomes existed, $\chi^2 (1) = 0.3$, $p = .61$.

3.3.2.3. Set shifting/cognitive flexibility. The meta-analysis of primary studies examining the differences in set shifting/cognitive flexibility between AN patients and control participants revealed a significant underperformance in the patient group ($\bar{g} = -0.385$, 95 % CI [-0.504, -0.266]; $Q_E [57] = 223.7$, $p < .01$) with high heterogeneity (75.1 %; see Table 3). Moderator analyses revealed a marginal negative effect of years of education (the more years of education, the more negative the effect size; $B = -0.138$, $SE = 0.082$, $Q_M [1] = 2.9$, $p = .09$), and positive moderator effects of psychotropic medication (the more participants on psychotropic medication, the less negative the effect size; $B = 0.286$, $SE = 0.076$, $Q_M [1] = 14.0$, $p < .01$) and trait anxiety (higher scores on the STAI-T were associated with less negative effect size; $B = 0.562$, $SE = 0.156$, $Q_M [1] = 13.0$, $p < .01$). Seven different tests were used to assess cognitive flexibility (see Table 4). The effect sizes varied greatly between tasks from $\bar{g} = 0.180$ for the phonemic condition of the Verbal Fluency Task to $\bar{g} = -1.168$ for the Berg Card Sorting Test. This variation was statistically significant ($\tau^2_{(3)} = 0.135$, 95 % CI [0.043, 0.173]), as the comparison between the baseline model and a three-level random-effects model with an additional outcome level indicated, $\chi^2 (1) = 6.2$, $p = .01$.

3.3.3. Memory

We also observed a moderate, significant, and negative effect size for the domain of memory ($\bar{g} = -0.485$, 95 % CI [-0.698, -0.273]; $Q_E [37] = 98.4$, $p < .01$), again favoring healthy controls with substantial heterogeneity (64.7 %; see Table 3). However, we could not find any evidence supporting the significant differences between verbal and non-verbal memory measures ($B = 0.223$, $SE = 0.163$, $Q_M [1] = 1.9$, $p = .17$). The subsequent moderator analyses revealed a marginal positive effect of BMI (the higher the average BMI in the sample, the less negative the effect size; $B = 0.164$, $SE = 0.092$, $Q_M [1] = 3.2$, $p = .07$) and eating disorder severity (higher scores on the EDE-Q were associated with less negative effect sizes; $B = 0.399$, $SE = 0.097$, $Q_M [1] = 17.0$, $p < .01$). Years of education also exhibited a marginal negative effect (more years of education was associated with a stronger negative effect size; $B = -0.223$, $SE = 0.127$, $Q_M [1] = 3.4$, $p = .07$). No further subgroup or country differences existed. Seven different tasks were used for assessing

memory—the most commonly used test were the RCFT for non-verbal memory ($k = 7$) and the delayed recall condition of list learning tests (e.g., the California Verbal Learning Test) for verbal memory ($k = 7$). The outcome-specific effect sizes ranged between $\bar{g} = -1.101$ and $\bar{g} = -0.207$ (Table 4), yet did not differ significantly, $\chi^2 (1) = 0.8$, $p = .38$.

3.3.4. Processing speed

Patients with AN had a significantly worse processing speed performance compared to the healthy adults, $\bar{g} = -0.390$ (95 % CI [-0.530, -0.250]; $Q_E [29] = 56.1$, $p < .01$). The degree of heterogeneity was high (48.9 %; see Table 3), and the moderator analyses revealed a negative moderation effect of study quality (the better the quality of the study the more negative the effect size; $B = -0.157$, $SE = 0.061$, $Q_M [1] = 6.6$, $p = .01$). No other moderator effects were detected.

3.3.5. Visuospatial abilities

The effect size for domain of visuospatial abilities was moderate, again with patients performing significantly worse than control participants ($\bar{g} = -0.528$ (95 % CI [-0.661, -0.396]; $Q_E [28] = 58.6$, $p < .01$), and exhibited moderate heterogeneity (51.2 %; see Table 3). Marginal differences in the effect sizes existed between the two subdomains, with a more negative effect size for central coherence ($B = -0.241$, $SE = 0.132$, $Q_M [1] = 3.3$, $p = .07$). In addition, there was a marginal positive effect of study quality (the better the quality of the study, the less negative the effect sizes; $B = 0.117$, $SE = 0.066$, $Q_M [1] = 3.2$, $p = .08$). Four different tests were administered to assess visuospatial abilities, two of which were used for measuring central coherence. The moderator analyses revealed marginally significant differences between tests ($\tau^2_{(3)} = 0.030$, 95 % CI [0.000, 0.251]; $\chi^2 [1] = 2.8$, $p = .09$). The range of outcome-specific effect sizes is shown in Table 4.

3.3.6. Working memory

Our meta-analytic sample contained seven effect sizes which were based on measures of working memory. A fixed-effects model resulted in a moderate, negative, and significant effect size, $\bar{g} = -0.455$ (95 % CI [-0.818, -0.091]). The underlying, outcome-specific effect sizes were $\bar{g} = -0.626$ (Digit span) and $\bar{g} = -0.309$ (WMS letter number sequencing), respectively. We did not conduct any further moderator analyses.

3.4. Sensitivity analyses and publication bias

Supplementary Material S1 and S3 show the detailed results of both the sensitivity analyses and the analyses of publication bias. Overall, the specification of the meta-analytic models via Bayesian analysis supported the choice of the baseline models—specifically, the preference of

three-level random-effects models over models with fewer variance components was backed by the respective Bayesian credibility intervals (see Supplementary Material S1). Moreover, the sizes of the effects and their variance components were almost identical to those obtained from the REML estimation. The key results obtained from the series of meta-analyses were not sensitive to the exclusion of influential effect sizes (see Supplementary Material S1). Multilevel meta-analysis did not exhibit different results as compared to the standard random-effects models with robust variance estimation in situations where one of the variance components (next to the sampling variation) was small (see Supplementary Material S1). Notably, an alternative three-level random-effects model that takes into account both hierarchical and correlated effects and assumes a constant within-study correlation between sampling errors (ρ) did not fit the data significantly better and resulted in a weighted average effects size almost identical to that of the model without this correlation (for $\rho = 0.40$: $\bar{g} = -0.43$, 95 % CI [-0.51, -0.36]). Besides, the differences in variance components were negligible (see Supplementary Material S1). Finally, one moderator effect turned statistically significant after imputing the missing data points in the continuous moderators (see Supplementary Material S3). Overall, the results presented earlier show a substantial degree of robustness with respect to the selected conditions.

The trim-and-fill analyses for the entire set of effect sizes indicated that no additional effect size was missing, $L_0^+ = 0$ and $R_0^+ = 0$ (see Supplementary Material S1). Egger's regression test was significant ($B = -1.27$, $SE = 0.51$, $p = .01$), and so was the PEESE ($B = -2.41$, $SE = 0.91$, $p < .01$), suggesting that some selection bias may be present in the data. The funnel plot test, however, did not indicate such bias ($B = 0.00$, $SE = 0.01$, $p = .77$), and neither did Begg's correlation test ($r = .08$, $p = .11$). The estimated fail-safe N was high, $N = 14767$. These findings suggested that some degree of publication bias was present. For the analysis of publication bias per domain, please refer to the Supplementary Material S1.

Besides these analyses of publication bias, we further inspected the p -curve (see Supplementary Material S1). The right-skewed binomial test of the $k = 184$ effect sizes, of which $k = 78$ were statistically significant ($p < .05$), was significant ($z = -13.2$, $p < .01$), and, conversely, the flatness test was insignificant ($z = 6.5$, $p = .99$). These findings suggested that the extracted effect sizes had evidential value.

4. Discussion

The aim of the current study was to extract all available data on neuropsychological functioning in AN in order to examine the magnitude of difference in neuropsychological test performance between individuals with AN and healthy controls. By utilizing novel statistical methods, we accounted for dependent effect sizes and examined moderating variables which could be hypothesized to influence test performance. This is the first meta-analysis since 2010 (Zakzanis et al., 2010) which examines neuropsychological performance in patients with AN across a range of cognitive domains. Several primary studies have been published since then, and subsequently been encompassed in the present meta-analysis. To facilitate interpretation and generalizability of results, we only included patients in the acute stage of illness. In addition, to keep the meta-analysis rigorous and focused, the neuropsychological domains and subdomains of interest were chosen a-priori and were based on traditional neuropsychological domain classifications (Lezak et al., 2004) as well as previous studies of comparable psychiatric samples (Abramovitch et al., 2013; Geller et al., 2018). A total of 50 studies and 1778 patients were included. An overall analysis ($k = 184$), i.e. combining all test result in one meta-analysis, can be a valuable tool to provide an indication of general cognitive functioning in the patient group. The overall analysis in the current study indicated some broad and non-specific difference in cognitive function between individuals with AN and healthy controls – with the former performing significantly

lower compared to the latter. Moderator analyses of the overall cognitive function revealed an effect of both age and BMI, demonstrating the influence of participant characteristics on task performance.

Our finding that overall cognitive function was moderated by age has some implications for the debate on whether or not cognitive inefficiencies are a trait marker of AN. Despite a large amount of research, there still appears to be no clear understanding as to why adult patients with AN perform worse on tests of cognitive function. Is it the result of malnutrition or is it pre-existing, and does it resolve over time? Longitudinal studies on adults with AN have demonstrated cognitive inefficiencies also after recovery (Tchanturia et al., 2004b, 2002), thereby providing some support for cognitive deficits as a trait marker for the illness. However, studies on children and adolescents have challenged this view, showing that cognitive functioning improves with weight recovery (Bühren et al., 2012; Lang and Tchanturia, 2014; Lozano-Serra et al., 2014). In the current study, we revealed that age was a significant moderator of overall cognitive performance, with older participants performing worse than younger. This finding lends some support to the suggestion that longer duration of illness – as commonly observed in older patients – poses greater risk for cognitive deficits (Grau et al., 2019). Thus, it could be that the inferior cognitive performance is a consequence of prolonged malnutrition, rather than a trait. However, our moderator analyses revealed no moderating effects for neither age of onset nor duration of illness. We encourage future studies to include age of onset and/or duration of illness to further investigate the potential effects of illness duration on cognitive performance. To date there are no comparable meta-analyses of children and adolescents to allow for a direct comparison with the current results.

Regarding BMI, our findings are in line with the study by Zakzanis et al. (2010) which also noted a relationship between weight and test performance. Specifically, lower weight was associated with more severely impaired test performance. However, it should be noted that the aforementioned participant characteristics (age and BMI) were not associated with all the cognitive domains investigated in the current study. For example, the impact of BMI was only associated with the effect sizes for memory, inhibition and visuospatial abilities. Whereas age had a positive effect on inhibition – i.e. older participants were better at inhibiting a pre-potent response.

Continuing with our investigation of domains and sub domains, the patients performed worse on all domains, with effect sizes ranging from -0.34 (executive function) to -0.57 (attention). For subdomains the highest effect size was found in central coherence (-0.622), whereas the smallest – and only non-significant effect size – was revealed in planning (-0.104). The moderator analyses of (sub-) domains revealed some interesting findings. Firstly, the positive moderation effect of eating disorder severity and anxiety on memory and set shifting/cognitive flexibility, respectively, was somewhat surprising. Eating disorder severity has previously been associated with poorer visuospatial memory (Zuchova et al., 2013), and anxiety is commonly reported to negatively influence cognitive performance (Clarke and Mackleod, 2013). There are some studies showing that anxiety does not impair performance when compensatory strategies are used – like enhanced effort or increased cognitive processing resources (Eysenck et al., 2007). This can be investigated by employing pure measures of inhibition and shifting (Eysenck et al., 2007) which could give an indication of recourse utilization, including performance effectiveness and processing efficiency. Nonetheless, the counter-intuitiveness of our findings accentuates that replication is essential before making recommendations for future studies.

Secondly, our moderator analyses revealed a negative effect of education for both memory and set shifting/cognitive flexibility. This is surprising, since education is usually associated with a positive influence on neuropsychological performance (Lam et al., 2013). All our moderator-analyses were carefully checked, but we are still cautious about making firm conclusions before these findings have been replicated.

As mentioned in the introduction, executive function and visuospatial processing have received the vast majority of attention in the field. This is also reflected in the current study, where executive function – and particularly set shifting/cognitive flexibility – was the most commonly researched domain ($k = 58$). A recently published review by Miles et al. (2020) report mixed results from studies investigating cognitive flexibility in AN. Their review found that adult patients with AN performed significantly worse on some perceptual cognitive flexibility tasks like the Brixton, but that findings were mixed for other tests of cognitive flexibility – including the Wisconsin Card Sorting Test, the Verbal Fluency Test and the Trail Making Test (Miles et al., 2020). When synthesizing findings from multiple studies, our analyses revealed a medium, significant, negative effect size for set shifting/cognitive flexibility. However, there was significant variation between tests. Thus, as noted in the review by Miles et al. (2020), measures of cognitive flexibility appear to vary in usefulness in terms of differentiating between patients with AN and healthy controls. Since “by definition, executive functions operate on other cognitive processes” (Strauss et al., 2006, p. 405), non-executive cognitive processes will be involved in solving these tasks. Thus, we encourage future studies to utilize tests which isolate specific areas of executive functioning, in order to explore which components of executive function are compromised – rather than complex task where it can be difficult to elucidate the cognitive functions involved. The Wisconsin Card Sorting Test and the shifting version of the Trail Making Test were the most commonly used measures of set shifting/cognitive flexibility ($k = 19$ and $k = 13$, respectively) in the current study. Both of these measures revealed a significantly worse performance in the patient group, with medium effect sizes for both tasks. Our findings thus confirm previous studies which demonstrate discrepancy between verbal and perceptual set shifting/cognitive flexibility tasks (Zakzanis et al., 2010). For patients with AN, the perceptual form of set shifting appeared to be associated with cognitive underperformance, whereas verbal set shifting appears comparable – or even superior – to the control group (Stedal et al., 2012; Zakzanis et al., 2010). This divergence could also account for the significant variation between tests for set shifting/cognitive flexibility domain in the current study. Performance on verbal fluency has a strong correlation with speed of information processing (Boone et al., 1998), verbal IQ, as well as working and semantic memory – in addition to executive functions like monitoring and suppression (Strauss et al., 2006). Whereas the Brixton task does not require any verbal abilities and has been shown to load on the same factor – *simple alternation* – as the Trail Making Test (shifting) in patients with eating disorders (Tchanturia et al., 2004a).

Thus, most neuropsychological tests assess combinations of different cognitive functions – and particularly tests assessing executive functions are known to operate on multiple cognitive processes (Miyake et al., 2000). For example, the Wisconsin Card Sorting Test is categorized within the set shifting/cognitive flexibility subdomain of executive functions. However, performance on this task involves several cognitive functions, including memory, visual perception, auditory perception, as well as cognitive flexibility (Keefe, 1995). Therefore we cannot be conclusive about the specificity of impairment found, and caution against overly simple interpretations that do not take account of this complexity.

Interestingly, the moderator effect of psychotropic medication was positive, indicating that patients on medication performed better on the set shifting/cognitive flexibility tasks. Previous studies have shown that some medications can indeed influence cognitive performance (Barker et al., 2004; Goldberg and Burdick, 2001; Pachet and Wisniewski, 2003). With this in mind, it is concerning that half of the included studies fail to report – or control for – medication use in their samples. It is also noteworthy that few of the included studies account for the potential impact of co-morbid disorders on neuropsychological performance. Especially since AN is commonly associated with other psychiatric disorders, such as anxiety and/or depression (Kaye et al., 2004; Kennedy et al., 1994). In fact, most studies ($m = 35$) neither assess nor report the

number of participants with a co-morbid Axis I diagnosis. This is particularly striking since the impact of depression and anxiety on neuropsychological performance in AN has been reported in previous studies (Billingsley-Marshall et al., 2013; Ely et al., 2016; Wilsdon and Wade, 2006). Both these illnesses are themselves associated with impaired neuropsychological test performance (Moran, 2016; Rock et al., 2014) and meta-analyses of affective disorders have repeatedly demonstrated that patients express deficits in areas of memory, attention and executive function when compared to controls (Burt et al., 1995; Zakzanis et al., 1998).

Since the early 2000s there has been interest in developing a neuropsychological test battery specifically aimed at assessing patients with AN (Rose et al., 2011). However, as apparent from the vast range of tests employed in the source studies included in the current meta-analysis, there is still a lack of consensus regarding which neuropsychological test(s) would be most suitable for assessing individuals suffering from this illness. The results from our meta-analysis revealed four neuropsychological tests which stood out. Firstly, because they demonstrated an effect size which would be considered medium or large (i.e. Hedges $\bar{g} \geq 0.5$). Secondly, because they encompassed more than five effect sizes, thereby indicating a fairly strong evidence base. The four tests were as follows: The switching condition of the Trail Making Test ($k = 13$, $\bar{g} = -0.595$ (Delis et al., 2001; Reitan and Wolfson, 1985), the perseverative responses condition of the Wisconsin Card Sorting Test ($k = 9$, $\bar{g} = -0.535$) (Heaton et al., 1993), the Block Design Test from The Wechsler Adult Intelligence Scale ($k = 7$, $\bar{g} = -0.660$) (Wechsler, 2008) and the Rey Complex Figure Test (Osterrieth, 1944) where both delayed recall ($k = 7$, $\bar{g} = -0.561$) and the central coherence index ($k = 9$, $\bar{g} = -0.669$) revealed a medium magnitude of difference between individuals with AN and healthy controls. Thus, based on these findings, clinicians or researchers with limited resources and/or time might wish to consider including one or more of the above-mentioned neuropsychological tests when performing cognitive assessments of individuals with AN.

It is worth noting that an inherent concern with employing neuropsychological tests to assess psychiatric populations is that the majority of these tasks have been developed to differentiate between healthy controls and patients with brain lesions or trauma – not psychiatric populations, meaning they are potentially not sensitive enough to detect subtle cognitive inefficiencies (Keefe, 1995). It has been argued that there is a need for instruments specifically tailored for assessing and detecting patients with psychiatric disorders (Kuelz et al., 2004). One first step could be to place greater emphasis on process scores which are generated by monitoring behaviors used to solve the task (Stedal et al., 2019). The vast majority of studies included in the current synthesis report global, or overall, achievement scores. The Central coherence index domain is an exception. In fact, our results reveal that central coherence was associated with the largest effect size, indicating that in comparison to healthy controls, patients with AN apply a more inefficient strategy when copying a complex figure. This is in line with a previous synthesis of studies employing the Rey Complex Figure Test to investigate central coherence in eating disorders (Lang et al., 2016). Taken together, these findings highlight the benefit of recording process scores when examining psychiatric populations with neuropsychological instruments.

Another important point of consideration when interpreting findings from neuropsychological tasks, is whether or not these effect sizes should be considered clinically significant – and if they reflect clinically significant impairments. Traditionally, a score of 2 standard deviations (the equivalent of a Cohen's d effect size of ± 2.0) below the norm is considered clinically significant (Lezak et al., 2004). On the other hand, Abramovitch and Cooperman (2015) propose that a standard deviation of 1.0 could be considered a clinically meaningful cutoff for neuropsychological test performance. In the current study, none of the effect sizes were 1 or above. It has been argued that small to moderate effect sizes – like the findings from the current study – should be labelled *cognitive underperformance*, rather than clinically significant impairments

(Abramovitch et al., 2013). Another principal consideration is how these findings relate to everyday functioning. Studies examining the ecological validity of neuropsychological test performance in AN have demonstrated only small correlations between measures considered more ecologically valid and traditional neuropsychological tests (Herbrich et al., 2019; Spitoni et al., 2018; Stedal and Dahlgren, 2015). However, the majority of these studies are based on findings from children and adolescents. Future studies should aim to include multiple formats of cognitive assessments since “self-reports of functioning, as well as observations of behavior while performing testing, are critically important pieces of information” (Harvey, 2012, p. 91).

Our quality of study index revealed that the precision of the neuropsychological test results could be questionable for some of the included studies. Despite moderator analyses revealing only one association between test results (processing speed) and the quality of study index, it is concerning that several of the included studies failed to report, or control for, variables which have been demonstrated to influence neuropsychological test results – like anxiety (Ely et al., 2016), depression (Burt et al., 1995; Wilsdon and Wade, 2006), medication (Barker et al., 2004; Goldberg and Burdick, 2001) and intelligence quotient (Diaz-Asper et al., 2004). Consequently, for future studies, we emphasize the importance of reporting possible confounding variables – both to control for in the primary study and to facilitate future meta-analyses.

4.1. Limitations and future directions

The current study has some limitations which are worth noting: First, we used an a-priori classification of neuropsychological domains and subdomains as a way to present our findings—more fine-grained and alternative classifications, however, could lead to locally different effect sizes and moderator effects. We therefore encourage researchers in the field of neuropsychological functioning to examine the convergence and diversity of such classifications in subsequent meta-analyses.

Second, for some (sub-)domains, such as working memory and central coherence, the interpretation of the effect sizes was limited by only two available outcome measures. Besides, some (sub-)domains contained only few effect sizes. As a consequence, quantifying and explaining variation between outcomes within these domains and performing meaningful moderator analyses was not possible. Moreover, the degree and consequences of publication bias is to be examined. We therefore suggest updating the sample of primary studies in future meta-analyses to increase the number of studies and effect sizes.

Third, the precision of the effect size interval estimate depends on several factors, including sample size and elements of the study design. Although the study quality index revealed that the precision of the neuropsychological test results could be questionable for some of the included studies, our moderator analyses showed that it only moderated the effect sizes for the domain of processing speed. Similar to the classification of (sub-)domains, these results may depend on the setup of the study quality index and may thus be examined further.

Fourth, with the exception of two, all studies examined all-female samples. This observation is by no means surprising – in fact, recent studies have emphasized that men have been underrepresented in the extant eating disorder literature (Limbers et al., 2018). Given that the generalizability of our finding to male populations is thus limited, it is key to future primary studies to also include male participants.

5. Conclusion

The findings from the current meta-analysis revealed consistent evidence of cognitive underperformance in individuals with AN. We have provided a framework for comparing results across studies and offer specific suggestions for neuropsychological tests which might be helpful in the assessment of this patient group. To move the field forward, there is a need for greater coherence in assessment procedures. Increased knowledge concerning a potential neuropsychological profile specific

for AN could facilitate the development of more sensitive instruments. Until then, utilizing a process oriented neuropsychological assessment approach might hold some promise.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgements

This work was funded by the Clinic for Mental Health and Addiction, Oslo University Hospital, Ullevål HF, Norway (Grant No. 44079). The funding source had no involvement in any part of the conduction of the research or preparation of the manuscript.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.neubiorev.2021.08.021>.

References

- Abramovitch, A., Cooperman, A., 2015. The cognitive neuropsychology of obsessive-compulsive disorder: a critical review. *J. Obsessive-Compulsive Relat. Disord.* 5, 24–36. <https://doi.org/10.1016/j.jocrd.2015.01.002>.
- Abramovitch, A., Abramowitz, J.S., Mittelman, A., 2013. The neuropsychology of adult obsessive-compulsive disorder: a meta-analysis. *Clin. Psychol. Rev.* 33 (8), 1163–1171. <https://doi.org/10.1016/j.cpr.2013.09.004>.
- American Psychiatric Association, 2013. *Diagnostic and Statistical Manual of Mental Disorders, fifth edition*. American Psychiatric Publishing, Arlington, VA.
- Aspen, V., Darcy, A.M., Lock, J., 2013. A review of attention biases in women with eating disorders. *Cogn. Emot.* 27 (5), 820–838. <https://doi.org/10.1080/02699931.2012.749777>.
- Association, A. P., 1987. *Diagnostic and Statistical Manual, 3rd revised ed.* APA Press, Washington, DC.
- Barker, M.J., Greenwood, K.M., Jackson, M., Crowe, S.F., 2004. Persistence of cognitive effects after withdrawal from long-term benzodiazepine use: a meta-analysis. *Arch. Clin. Neuropsychol.* 19 (3), 437–454. [https://doi.org/10.1016/S0887-6177\(03\)00096-9](https://doi.org/10.1016/S0887-6177(03)00096-9). %J Archives of Clinical Neuropsychology.
- Begg, C.B., Mazumdar, M., 1994. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 50 (4), 1088–1101. <https://doi.org/10.2307/2533446>.
- Billingsley-Marshall, R.L., Basso, M.R., Lund, B.C., Hernandez, E.R., Johnson, C.L., Drevets, W.C., et al., 2013. Executive function in eating disorders: the role of state anxiety. *Int. J. Eat. Disord.* 46 (4), 316–321. <https://doi.org/10.1002/eat.22086>.
- Boone, K.B., Pontón, M.O., Gorsuch, R.L., González, J.J., Miller, B.L., 1998. Factor analysis of four measures of prefrontal lobe functioning. *Arch. Clin. Neuropsychol.* 13 (7), 585–595.
- Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R., 2009. *Introduction to Meta-Analysis*. John Wiley & Sons, Ltd, Chichester, UK.
- Brooks, S., Prince, A., Stahl, D., Campbell, I.C., Treasure, J., 2011. A systematic review and meta-analysis of cognitive bias to food stimuli in people with disordered eating behaviour. *Clin. Psychol. Rev.* 31 (1), 37–51. <https://doi.org/10.1016/j.cpr.2010.09.006>.
- Bühren, K., Mainz, V., Herpertz-Dahlmann, B., Schäfer, K., Kahraman-Lanzerath, B., Lente, C., Konrad, K., 2012. Cognitive flexibility in juvenile anorexia nervosa patients before and after weight recovery. *J. Neural Transm.* 119 (9), 1047–1057. <https://doi.org/10.1007/s00702-012-0821-z>.
- Bürkner, P.-C., 2017. brms: an R package for Bayesian multilevel models using stan. *J. Stat. Softw.* 80 (1), 28. <https://doi.org/10.18637/jss.v080.i01>.
- Burt, D.B., Zembard, M.J., Niederehe, G., 1995. Depression and memory impairment: a meta-analysis of the association, its pattern, and specificity. *Psychol. Bull.* 117 (2), 285–305. <https://doi.org/10.1037/0033-2909.117.2.285>.
- Cheung, M.W.L., 2013. *Meta-Analysis: A Structural Equation Modeling Approach*. John Wiley & Sons Ltd., Chichester, UK.
- Cheung, M.W.L., 2019. A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychol. Rev.* 29 (4), 387–396. <https://doi.org/10.1007/s11065-019-09415-6>.
- Clarke, P., Mackleod, C., 2013. The impact of anxiety on cognitive task performance. In: Arnett, P. (Ed.), *Secondary Influences on Neuropsychological Test Performance*. Oxford university press., New York, NY.
- Delis, D.C., Kaplan, E., Kramer, J., 2001. *Delis Kaplan Executiv Function System*. The Psychological Corporation, San Antonio.
- Diaz-Asper, C.M., Schretlen, D.J., Pearson, G.D., 2004. How well does IQ predict neuropsychological test performance in normal adults? *J. Int. Neuropsychol. Soc.* 10 (1), 82–90. <https://doi.org/10.1017/s1355617704101100>.
- Duval, S., Tweedie, R., 2000. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics* 56 (2), 455–463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>.

- Egger, M., Smith, G.D., Schneider, M., Minder, C., 1997. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 315 (7109), 629–634. <https://doi.org/10.1136/bmj.315.7109.629>. %J BMJ.
- Ely, A.V., Wierenga, C.E., Kaye, W.H., 2016. Anxiety impacts cognitive inhibition in remitted anorexia nervosa. *Eur. Eat. Disord. Rev.* 24 (4), 347–351. <https://doi.org/10.1002/erv.2451>.
- Eysenck, M.W., Derakshan, N., Santos, R., Calvo, M.G., 2007. Anxiety and cognitive performance: attentional control theory. *Emotion* 7 (2), 336–353. <https://doi.org/10.1037/1528-3542.7.2.336>.
- Fernández-Castilla, B., Maes, M., Declercq, L., Jamshidi, L., Beretvas, S.N., Onghena, P., Van den Noortgate, W., 2019. A demonstration and evaluation of the use of cross-classified random-effects models for meta-analysis. *Behav. Res. Methods* 51 (3), 1286–1304. <https://doi.org/10.3758/s13428-018-1063-2>.
- Fernández-Castilla, B., Jamshidi, L., Declercq, L., Beretvas, S.N., Onghena, P., Van den Noortgate, W., 2020. The application of meta-analytic (multi-level) models with multiple random effects: a systematic review. *Behav. Res. Methods*. <https://doi.org/10.3758/s13428-020-01373-9>.
- Fernández-Castilla, B., Declercq, L., Jamshidi, L., Beretvas, S.N., Onghena, P., Van den Noortgate, W., 2021. Detecting selection bias in meta-analyses with multiple outcomes: a simulation study. *J. Exp. Educ.* 89 (1), 125–144. <https://doi.org/10.1080/00220973.2019.1582470>.
- Fisher, Z., Tipton, E., Zhipeng, H., 2017. Robumeta: Robust Variance Meta-regression. R Package Version 2.0.
- Geller, D.A., Abramovitch, A., Mittelman, A., Stark, A., Ramsey, K., Cooperman, A., et al., 2018. Neurocognitive function in paediatric obsessive-compulsive disorder. *World J. Biol. Psychiatry* 19 (2), 142–151. <https://doi.org/10.1080/15622975.2017.1282173>.
- Goldberg, J.F., Burdick, K.E., 2001. Cognitive side effects of anticonvulsants. *J. Clin. Psychiatry* 62 (Suppl. 14), 27–33.
- Grau, A., Magallón-Neri, E., Faus, G., Feixas, G., 2019. Cognitive impairment in eating disorder patients of short and long-term duration: a case-control study. *Neuropsychiatr. Dis. Treat.* 15, 1329–1341. <https://doi.org/10.2147/NDT.S199927>.
- Guillaume, S., Gorwood, P., Jollant, F., Van den Eynde, F., Courtet, P., Richard-Davyant, S., 2015. Impaired decision-making in symptomatic anorexia and bulimia nervosa patients: a meta-analysis. *Psychol. Med.* 45 (16), 3377–3391. <https://doi.org/10.1017/s003329171500152x>.
- Hamsher, Kd.S., Halmi, K.A., Benton, A.L., 1981. Prediction of outcome in anorexia nervosa from neuropsychological status. *Psychiatry Res.* 4 (1), 79–88. [https://doi.org/10.1016/0165-1781\(81\)90011-1](https://doi.org/10.1016/0165-1781(81)90011-1).
- Harper, J.A., Brodrick, B., Van Enkevort, E., McAdams, C.J., 2017. Neuropsychological and cognitive correlates of recovery in anorexia nervosa. *Eur. Eat. Disord. Rev.* 25 (6), 491–500. <https://doi.org/10.1002/erv.2539>.
- Harrell, M., Cuijpers, P., Furukawa, T., Ebert, D.D., 2019. dmetar: Companion R Package for the Guide 'Doing Meta-analysis in R'. R Package Version 0.0.9000.
- Harvey, P.D., 2012. Clinical applications of neuropsychological assessment. *Dialogues Clin. Neurosci.* 14 (1), 91–99.
- Heaton, R.K., Chelune, G.J., Talley, J.L., Kay, G.G., Curtiss, G., 1993. *Wisconsin Card Sorting Test Manual: Revised and Expanded*. Psychological Assessment Resources, Odessa, FL.
- Hemmingsen, S.D., Wesselhoeft, R., Lichtenstein, M.B., Sjögren, J.M., Støving, R.K., 2020. Cognitive improvement following weight gain in patients with anorexia nervosa: a systematic review. *European Eating Disorders Review*. <https://doi.org/10.1002/erv.2796> n/a(n/a).
- Herbrich, L.R., Kappel, V., Winter, S.M., van Noort, B.M., 2019. Executive functioning in adolescent anorexia nervosa: neuropsychology versus self- and parental-report. *Child Neuropsychol.* 25 (6), 816–835. <https://doi.org/10.1080/09297049.2018.1536200>.
- Hirst, R.B., Beard, C.L., Colby, K.A., Quittner, Z., Mills, B.M., Lavender, J.M., 2017. Anorexia nervosa and bulimia nervosa: a meta-analysis of executive functioning. *Neurosci. Biobehav. Rev.* 83, 678–690. <https://doi.org/10.1016/j.neubiorev.2017.08.011>.
- Jones, B.P., Duncan, C.C., Brouwers, P., Mirsky, A.F., 1991. Cognition in eating disorders. *J. Clin. Exp. Neuropsychol.* (5), 711–728.
- Kanakam, N., Raoult, C., Collier, D., Treasure, J., 2013. Set shifting and central coherence as neurocognitive endophenotypes in eating disorders: a preliminary investigation in twins. *World J. Biol. Psychiatry* 14 (6), 464–475. <https://doi.org/10.3109/15622975.2012.665478>.
- Kask, J., Ekselius, L., Brandt, L., Kollia, N., Ekbo, A., Papadopoulos, F.C., 2016. Mortality in women with anorexia nervosa: the role of comorbid psychiatric disorders. *Psychosom. Med.* 78 (8), 910–919. <https://doi.org/10.1097/psy.0000000000000342>.
- Kaye, W., Bulik, C., Thornton, L., Barbarich, N., Masters, K., Fichter, M., et al., 2004. Comorbidity of anxiety disorders with anorexia and bulimia nervosa. *Am. J. Psychiatry* 161. <https://doi.org/10.1176/appi.ajp.161.12.2215>.
- Keefe, R.S., 1995. The contribution of neuropsychology to psychiatry. *Am. J. Psychiatry* 152 (1), 6–15. <https://doi.org/10.1176/ajp.152.1.6>.
- Kennedy, S.H., Kaplan, A.S., Garfinkel, P.E., Rockert, W., Toner, B., Abbey, S.E., 1994. Depression in anorexia nervosa and bulimia nervosa: discriminating depressive symptoms and episodes. *J. Psychosom. Res.* 38 (7), 773–782. [https://doi.org/10.1016/0022-3999\(94\)90030-2](https://doi.org/10.1016/0022-3999(94)90030-2).
- Kuelz, A.K., Hohagen, F., Voderholzer, U., 2004. Neuropsychological performance in obsessive-compulsive disorder: a critical review. *Biol. Psychol.* 65 (3), 185–236. <https://doi.org/10.1016/j.biopsycho.2003.07.007>.
- Lam, M., Eng, G.K., Rapisarda, A., Subramanian, M., Kraus, M., Keefe, R.S.E., Collinson, S.L., 2013. Formulation of the age-education index: measuring age and education effects in neuropsychological performance. *Psychol. Assess.* 25 (1), 61–70. <https://doi.org/10.1037/a0030548>.
- Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. *Biometrics* 33 (1), 159–174.
- Lang, K., Tchanturia, K., 2014. A systematic review of central coherence in young people with anorexia nervosa. *J. Child Adolesc. Behav.* 2 (140), 2.
- Lang, K., Roberts, M., Harrison, A., Lopez, C., Goddard, E., Khondoker, M., et al., 2016. Central coherence in eating disorders: a synthesis of studies using the Rey Osterrieth complex figure test. *PLoS One* 11 (11), e0165467. <https://doi.org/10.1371/journal.pone.0165467>.
- Lezak, M.D., Howieson, D.B., Loring, D.W., Hannay, H.J., Fischer, J.S., 2004. *Neuropsychological Assessment*. Oxford University Press, Oxford, UK.
- Limbers, C.A., Cohen, L.A., Gray, B.A., 2018. Eating disorders in adolescent and young adult males: prevalence, diagnosis, and treatment strategies. *Adolesc. Health Med. Ther.* 9, 111–116. <https://doi.org/10.2147/AHMT.S147480>.
- Lopez, C., Tchanturia, K., Stahl, D., Booth, R., Holliday, J., Treasure, J., 2008. An examination of the concept of central coherence in women with anorexia nervosa. *Int. J. Eat. Disord.* 41 (2), 143–152.
- Lozano-Serra, E., Andrés-Perpiña, S., Lázaro-García, L., Castro-Fornieles, J., 2014. Adolescent anorexia nervosa: cognitive performance after weight recovery. *J. Psychosom. Res.* 76 (1), 6–11. <https://doi.org/10.1016/j.jpsychores.2013.10.009>.
- Miles, S., Gnaatt, I., Phillipou, A., Nedeljkovic, M., 2020. Cognitive flexibility in acute anorexia nervosa and after recovery: a systematic review. *Clin. Psychol. Rev.* 81, 101905. <https://doi.org/10.1016/j.cpr.2020.101905>.
- 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. *Cogn. Psychol.* 41 (1), 49–100. <https://doi.org/10.1006/cogp.1999.0734>.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., The, P.G., 2009. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 6 (7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>.
- Moran, T.P., 2016. Anxiety and working memory capacity: a meta-analysis and narrative review. *Psychol. Bull.* 142 (8), 831–864. <https://doi.org/10.1037/bul0000051>.
- Øverås, M., Kapstad, H., Brunborg, C., Landrø, N.I., Rø, Ø., 2017. Is overestimation of body size associated with neuropsychological weaknesses in anorexia nervosa? *Eur. Eat. Disord. Rev.* 25 (2), 129–134. <https://doi.org/10.1002/erv.2500>.
- Osterrieth, P.A., 1944. Le test de copie d'une figure complexe. *Arch. Psychol.* 30, 206–356.
- Pachet, A.K., Wisniewski, A.M., 2003. The effects of lithium on cognition: an updated review. *Psychopharmacology (Berl.)* 170 (3), 225–234. <https://doi.org/10.1007/s00213-003-1592-x>.
- Pustejovsky, J.E., Tipton, E., 2021. Meta-analysis with robust variance estimation: expanding the range of working models. *Prev. Sci.* <https://doi.org/10.1007/s11212-021-01246-3>.
- Reitan, R.M., Wolfson, D., 1985. *The Halstead-Reitan Neuropsychological Test Battery: Theory and Clinical Interpretation*. Neuropsychology Press, Tucson, Ariz.
- Rock, P.L., Roiser, J.P., Riedel, W.J., Blackwell, A.D., 2014. Cognitive impairment in depression: a systematic review and meta-analysis. *Psychol. Med.* 44 (10), 2029–2040. <https://doi.org/10.1017/s0033291713002535>.
- Rose, M., Davis, J., Frampton, I., Lask, B., 2011. The Ravello Profile: development of a global standard neuropsychological assessment for young people with anorexia nervosa. *Clin. Child Psychol. Psychiatry* 16 (2), 195–202. <https://doi.org/10.1177/1359104511401191>.
- Rubio-Aparicio, M., Sánchez-Meca, J., López-López, J.A., Botella, J., Marín-Martínez, F., 2017. Analysis of categorical moderators in mixed-effects meta-analysis: consequences of using pooled versus separate estimates of the residual between-studies variances. *Br. J. Math. Stat. Psychol.* 70 (3), 439–456. <https://doi.org/10.1111/bmsp.12092>.
- Shott, M.E., Filoteo, J.V., Bhatnagar, K.A.C., Peak, N.J., Hagman, J.O., Rockwell, R., et al., 2012. Cognitive set-shifting in anorexia nervosa. *Eur. Eat. Disord. Rev.* 20 (5), 343–349. <https://doi.org/10.1002/erv.2172>.
- Simonson, U., Nelson, L.D., Simmons, J.P., 2014. P-curve: a key to the file-drawer. *J. Exp. Psychol.: Gen.* 143 (2), 534–547. <https://doi.org/10.1037/a0033242>.
- Smith, K.E., Mason, T.B., Johnson, J.S., Lavender, J.M., Wonderlich, S.A., 2018. A systematic review of reviews of neurocognitive functioning in eating disorders: the state-of-the-literature and future directions. *Int. J. Eat. Disord.* 51 (8), 798–821. <https://doi.org/10.1002/eat.22929>.
- Spitoni, G.F., Aragona, M., Bevacqua, S., Cotugno, A., Antonucci, G., 2018. An ecological approach to the behavioral assessment of executive functions in anorexia nervosa. *Psychiatry Res.* 259, 283–288. <https://doi.org/10.1016/j.psychres.2017.10.029>.
- Stedal, K., 2012. *The Neuropsychological Performance of Patients with Anorexia Nervosa*. AIT Oslo AS, Oslo.
- Stedal, K., Dahlgren, C.L., 2015. Neuropsychological assessment in adolescents with anorexia nervosa – exploring the relationship between self-report and performance-based testing. *J. Eat. Disord.* 3, 27. <https://doi.org/10.1186/s40337-015-0062-2>.
- Stedal, K., Frampton, I., Landro, N.I., Lask, B., 2012. An examination of the ravello profile—a neuropsychological test battery for anorexia nervosa. *Eur. Eat. Disord. Rev.* 20 (3), 175–181. <https://doi.org/10.1002/erv.1160>.
- Stedal, K., Ely, A., Kurniadi, N., Lopez, E., Kaye, W.H., Wierenga, C.E., 2019. A process approach to verbal memory assessment: exploratory evidence of inefficient learning in women remitted from anorexia nervosa. *J. Clin. Exp. Neuropsychol.* 41 (6), 653–663. <https://doi.org/10.1080/13803395.2019.1610160>.
- Steinglass, J.E., Walsh, B.T., Stern, Y., 2006. Set shifting deficit in anorexia nervosa. *J. Int. Neuropsychol. Soc.* 12 (3), 431–435.
- Strauss, E., Sherman, E., Spreen, O., 2006. *A Compendium of Neuropsychological Tests. Administration, Norms and Commentary*, 3rd ed. Oxford University Press, Oxford.

- Tchanturia, K., Morris, R.G., Surguladze, S., Treasure, J., 2002. An examination of perceptual and cognitive set shifting tasks in acute anorexia nervosa and following recovery. *Eat. Weight Disord.* 7 (4), 312–315.
- Talbot, A., Hay, P., Buckett, G., Touyz, S., 2015. Cognitive deficits as an endophenotype for anorexia nervosa: an accepted fact or a need for re-examination? *Int. J. Eat. Disord.* 48 (1), 15–25. <https://doi.org/10.1002/eat.22332>.
- Tchanturia, K., Anderluh, M.B., Morris, R.G., Rabe-Hesketh, S., Collier, D.A., Sanchez, P., Treasure, J.L., 2004a. Cognitive flexibility in anorexia nervosa and bulimia nervosa. *J. Int. Neuropsychol. Soc.* 10 (4), 513–520.
- Tchanturia, K., Morris, R.G., Anderluh, M.B., Collier, D.A., Nikolaou, V., Treasure, J., 2004b. Set shifting in anorexia nervosa: an examination before and after weight gain, in full recovery and relationship to childhood and adult OCPD traits. *J. Psychiatr. Res.* 38 (5), 545–552. <https://doi.org/10.1016/j.jpsychires.2004.03.001>.
- Tchanturia, K., Campbell, I.C., Morris, R., Treasure, J., 2005. Neuropsychological studies in anorexia nervosa. *Int. J. Eat. Disord.* 37 (Suppl), S72–S76. <https://doi.org/10.1002/eat.20119> discussion S87–79.
- Thompson, S.B., 1993. Implications of neuropsychological test results of women in a new phase of anorexia nervosa. *Eur. Eat. Disord. Rev.* 1 (3), 152–165.
- Tipton, E., Pustejovsky, J.E., Ahmadi, H., 2019. Current practices in meta-regression in psychology, education, and medicine. *Res. Synth. Methods* 10 (2), 180–194. <https://doi.org/10.1002/jrsm.1339>.
- Valentine, J.C., 2009. Judging the quality of primary research. In: Cooper, H., Hedges, L. V., Valentine, J.C. (Eds.), *The Handbook of Research Synthesis and Meta-Analysis*, 2nd edition. Russel Sage Foundation, New York, pp. 129–145.
- Viechtbauer, W., 2010. Conducting meta-analyses in R with the metafor package. *J. Stat. Softw.* 1 (3) <https://doi.org/10.18637/jss.v036.i03> (2010).
- Viechtbauer, W., Cheung, M.W.-L., 2010. Outlier and influence diagnostics for meta-analysis. *Res. Synth. Methods* 1 (2), 112–125. <https://doi.org/10.1002/jrsm.11>.
- Viera, A.J., Garrett, J.M., 2005. Understanding interobserver agreement: the kappa statistic. *Fam. Med.* 37 (5), 360–363.
- Wechsler, D.A., 2008. *Wechsler Adult Intelligence Scale, 4 ed.* Psychological Corporation, San Antonio, TX.
- Weider, S., Indreavik, M.S., Lydersen, S., Hestad, K., 2014. Intellectual function in patients with anorexia nervosa and bulimia nervosa. *Eur. Eat. Disord. Rev.* 22 (1), 15–24.
- Weider, S., Indreavik, M.S., Lydersen, S., Hestad, K., 2015. Neuropsychological function in patients with anorexia nervosa or bulimia nervosa. *Int. J. Eat. Disord.* 48 (4), 397–405. <https://doi.org/10.1002/eat.22283>.
- Westwood, H., Stahl, D., Mandy, W., Tchanturia, K., 2016. The set-shifting profiles of anorexia nervosa and autism spectrum disorder using the Wisconsin Card Sorting Test: a systematic review and meta-analysis. *Psychol. Med.* 46 (9), 1809–1827. <https://doi.org/10.1017/S0033291716000581>.
- Wilsdon, A., Wade, T.D., 2006. Executive functioning in anorexia nervosa: exploration of the role of obsessiveness, depression and starvation. *J. Psychiatr. Res.* 40 (8), 746–754. <https://doi.org/10.1016/j.jpsychires.2005.10.006>.
- Wu, M., Brockmeyer, T., Hartmann, M., Skunde, M., Herzog, W., Friederich, H.-C., 2014. Set-shifting ability across the spectrum of eating disorders and in overweight and obesity: a systematic review and meta-analysis. *Psychol. Med.* 1–21.
- Wu, M., Brockmeyer, T., Hartmann, M., Skunde, M., Herzog, W., Friederich, H.-C., 2016. Reward-related decision making in eating and weight disorders: a systematic review and meta-analysis of the evidence from neuropsychological studies. *Neurosci. Biobehav. Rev.* 61, 177–196. <https://doi.org/10.1016/j.neubiorev.2015.11.017>.
- Yang, Y., Shields, G.S., Guo, C., Liu, Y., 2018. Executive function performance in obesity and overweight individuals: a meta-analysis and review. *Neurosci. Biobehav. Rev.* 84, 225–244. <https://doi.org/10.1016/j.neubiorev.2017.11.020>.
- Zakzanis, K.K., Leach, L., Kaplan, E., 1998. On the nature and pattern of neurocognitive function in major depressive disorder. *Neuropsychiatry Neuropsychol. Behav. Neurol.* 11 (3), 111–119.
- Zakzanis, K.K., Campbell, Z., Polsinelli, A., 2010. Quantitative evidence for distinct cognitive impairment in anorexia nervosa and bulimia nervosa. *J. Neuropsychol.* 4 (1), 89–106. <https://doi.org/10.1348/174866409X459674>.
- Zuchova, S., Kubena, A.A., Erler, T., Papezova, H., 2013. Neuropsychological variables and clinical status in anorexia nervosa: relationship between visuospatial memory and central coherence and eating disorder symptom severity. *Eat. Weight. Disord. Stud. Anorex. Bulim. Obes.* 18 (4), 421–428. <https://doi.org/10.1007/s40519-013-0062-7>.