Assessing inhibitory control deficits in adult ADHD:

A systematic review and meta-analysis of the stop-signal task

Daniel Senkowski[1*], Theresa Ziegler[1], Mervyn Singh[2], Andreas Heinz[1], Jason He[3], Tim Silk[2], and Robert C. Lorenz[4]

[1]Department of Psychiatry and Neurosciences, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin and Humboldt Universität zu Berlin, Charité Campus Mitte (CCM), Charitéplatz 1, 10117 Berlin, Germany

[2]Centre for Social and Early Emotional Development (SEED) and School of Psychology, Deakin University, Geelong, Victoria, Australia

[3] King's College London, WC2R 2LS, United Kingdom

[4]Lise Meitner Group for Environmental Neuroscience, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany

ORCID-IDs: DS, 0000-0002-1467-8264; AH, 0000-0001-5405-9065**;** TS, 0000-0002-7290-512X; RL, 0000-0003-3286-6792

* *Send correspondence to:* Prof. Dr. Daniel Senkowski

Dept. of Psychiatry and Neurosciences
Charité Campus Mitte (CCM)
Charité-Universitätsmedizin Berlin
Charitéplatz 1
10117 Berlin, Germany.
Phone: +49-30-450-517055
E-mail: daniel.senkowski@charite.de

Conflict of interest: No conflict of interest to disclose

Funding declaration: No funding was received to assist with the preparation of this manuscript.

Ethical approval: Not applicable

Authors contribution: DS wrote the manuscript, was involved in the study selection process and supervised the data analysis; TZ and MS were involved in the study selection process and conducted the meta-analysis and meta-regression analysis. They were also involved in writing the methods and results sections; AH, JH, TC and RCL supervised the study selection and data analysis procedures and supported writing of the manuscript. In addition, RCL was directly involved in the data analysis.

Data availability statement: Data sharing is not applicable to this article as no original datasets were generated or analyzed during the current study. A preprint of an earlier version of this work has been made available to the public:

https://www.medrxiv.org/content/10.1101/2022.07.09.22277429v1

Abstract

Background: In recent years, there has been an increasing quest in improving our understanding of neurocognitive deficits underlying adult attention-deficit/hyperactivity disorder (ADHD). Current statistical manuals of psychiatric disorders emphasize inattention and hyperactivity-impulsivity symptoms, but empirical studies have also shown consistent alterations in inhibitory control. Thus far, there is no established neuropsychological test to assess inhibitory control deficits in adult ADHD. A common paradigm for assessing response inhibition is the stop-signal task (SST).

Methods: Following PRISMA-selection criteria, our systematic review and meta-analysis integrated the findings of 26 publications with 27 studies examining the SST in adult ADHD.

Results: The meta-analysis, which included 883 patients with adult ADHD and 916 control participants, revealed reliable inhibitory control deficits, as expressed in prolonged SST response times, with a moderate effect size g = 0.51 (95% CI: 0.376-0.644, p < 0.0001). The deficits were not moderated by study quality, sample characteristics or clinical parameters, suggesting that they may be a phenotype in this disorder. The analyses of secondary outcome measures revealed greater SST omission errors and reduced go accuracy in patients, indicative of altered sustained attention. However, only few (N<10) studies were available for these measures.

Discussion: Our meta-analysis suggests that the SST could, in conjunction with other tests and questionnaires, become a valuable tool for the assessment of inhibitory control deficits in adult ADHD.

Keywords: Attention deficit hyperactivity disorder; Impulsivity; Cognition; Attention; Cognitive deficits; Neuropsychology

Introduction

Adult attention deficit hyperactivity disorder (ADHD) is a neurodevelopmental condition that emerges during childhood or young adulthood and is characterized by symptoms of inattention and/or hyperactivity-impulsivity (Adler et al., 2017; Asherson et al., 2016). Adults with ADHD show a global prevalence of 2.58 % (persistent disorder) and 6.76 % (symptomatic disorder) (Song et al., 2021). In clinical practice, adult ADHD is assessed through questionnaires, interviews of relatives and inspection of school certificates. While neurocognitive deficits are inherent in ADHD, there is still no established test or test battery that is generally used in the assessment of this disorder (Fried et al., 2021; Nikolas et al., 2019). Nevertheless, in patients with presumed cognitive deficits, neuropsychological tests should be used to objectify these deficits during the diagnostic process. To date, there is an emerging quest in establishing neurocognitive paradigms as complementary tools in the assessment of adult ADHD.

Early research on ADHD has suggested that deficits in inhibitory control are a primary phenotype in this disorder (Barkley, 1997). Support for the inhibition deficit model comes, among others, from studies that showed altered executive functions in adult ADHD (Hadas et al., 2021; Linhartová et al., 2021; Nigg et al., 2002; Silverstein et al., 2020). An important aspect of executive functions is inhibitory control, which has frequently been investigated with the stop-signal task (SST; Verbruggen and Logan, 2008), but also with other paradigms like the go/no-go task (Fisher et al., 2011). However, there are substantial differences between the two tasks. For instance, the go/no-go task does not involve a measure of individual response inhibition speed, which is explicitly obtained in the SST. Moreover, it has been shown that different neural mechanisms underly stimulus processing in the two tasks (Raud et al., 2020). Hence, the go/no-go task and the SST likely capture different facets of response inhibition and therefore, they should be examined independently. In the current review and meta-analysis, we focused on the SST in

adult ADHD (for a meta-analysis on the go/no-go task in combined children, and adult studies, see Wright et al., 2014).

In the SST, participants are instructed to perform a forced-choice response following a 'go-signal', e.g., an arrow pointing to the left or right, and to respond with a left or right button press, respectively. Crucially, in a small proportion of trials, an auditory or visual 'stop-signal' is presented after the go-signal and participants are required to withhold the behavioral response. In most studies an adaptive approach is applied to obtain the go-signal to stop-signal delay interval for which the response inhibition rates are around 50% at an individual subject level. Based on this delay interval and the response times to go-trials, the stop-signal reaction time (SSRT) is calculated, which has become an established measure of response inhibition (Logan et al., 2014). A previous meta-analysis of the SSRT, which involved studies in children and adults diagnosed with ADHD, has revealed deficits of moderate effect sizes (g = 0.62) across age groups (Lipszyc and Schachar, 2010). This meta-analysis included 68 studies with children but only 10 studies with adult ADHD. Hence, the validity of this analysis regarding adult ADHD was limited and the degree of SSRT deficits in adult ADHD remains to be investigated.

Here, we performed a review and meta-analysis that conformed to current PRISMA-guidelines focusing on response inhibition deficits, as expressed in the SSRT, in adult ADHD. Our analysis included 26 publications with 27 studies, which allowed for a reliable estimation of response inhibition deficits in patients. We conducted a quality assessment of the SST following a recent consensus paper (Verbruggen et al., 2019) and estimated the risk of bias (RoB) for each study. We thoroughly examined if the study quality as well as participant-related and clinical factors influence response inhibition deficits in adult ADHD.

Methods

The protocol of this systematic review and meta-analysis was pre-registered in the PROSPERO database (PROSPERO ID: CRD42021266709).

Study selection

The following study selection criteria were applied: 1. Patient population: Included were studies containing at least one group of adult participants (18+) with a current diagnosis of ADHD in accordance with the DSM criteria (5 or earlier) or Hyperkinetic Disorder in accordance with the ICD (10 or earlier) criteria. Studies investigating populations with only subclinical ADHD symptomatology were not considered. 2. Control group: Studies must contain at least one healthy control group. 3. Experimental task: Response inhibition performance had to be obtained by the SST or the Stop-Change Task, which is a modified version of the SST in which individuals shift to a secondary response after they have inhibited an ongoing response (Verbruggen and Logan, 2009). Studies using atypical SST paradigms such as dual tasks or the selective SST were excluded. The same applies to studies in which participants received feedback on stop-signal performance, because feedback and reward influence response inhibition (Lipszyc and Schachar, 2010; Slusarek et al., 2001). 4. Outcome measure**:** Sufficient test statistics for the stop-signal reaction time (SSRT) must be provided to calculate standardized mean differences (Hedge's g). 5. Other criteria: Empirical articles written in English or German language and published or accepted for publication in peer-reviewed journals during the time range 2000-2022.

Search strategy

To identify relevant articles, an electronic search was conducted up to April 14, 2022 in two major publication databases: Medline and PsycInfo (accessed from EBSCOhost). The following syntax, adapted from Lipszyc and Schachter (2010), was used: [(attention deficit hyperactivity disorder OR ADHD) AND Adult* AND (stop task OR stop signal OR response inhibition OR executive function)]. Limiters were set to only show articles published in peer-reviewed journals since the 1st of January 2000, in English or German. Furthermore, reference lists of the identified empirical articles, previous meta-analysis and systematic reviews were scanned to ensure that all relevant articles were captured.

Study selection

The study selection process was conducted by two authors (TZ and DS) and included two stages: 1. Initial screening of titles and abstracts using the above-described inclusion and exclusion criteria. 2. For the resulting set of studies full texts were obtained and checked in detail for eligibility. Screening of eligible articles was performed in Endnote. In case of discrepancies between authors regarding the eligibility of studies, studies were screened by other team members and disagreements during the first or the second screening process were discussed until consensus was reached.

Data extraction and outcomes

Data was extracted by two authors independently (TZ and MS). If statistical values for the meta-analysis were not sufficiently reported, the authors of the articles were contacted and asked to provide the missing information. The following measures of the SST were extracted from the articles, separately for patients and controls: SSRT as primary outcome; stop commission errors (responding on a stop trial); go discrimination errors (e.g., responding with the left arrow key, even though a rightwards pointing arrow was presented); go omission errors (not responding on a go trial) and go accuracy (percentage of correct go trials) as secondary outcomes. In addition, important variables that could influence behavioral performance in the planned analysis were extracted and tabulated for each study: Age; IQ; percentage of males; ADHD subtype; years of education; comorbidities; medication status; recruitment setting of patients.

Assessments of the SST validity and risk of bias

The validity of the SST and the risk of bias (RoB) in each study are important factors that could influence group differences between patients and controls. Therefore, they were explicitly examined in the current analysis. Both SST validity and RoB were assessed by two independent raters (TZ and MS). In case of discrepancies between authors, consensus was reached from discussions with other team members. To increase inter-rater reliability, a calibration session was conducted in which the assessments were applied to two articles not

included in this review. This way, the two raters could identify possible sources of disagreement and decide on rules on how to rate ambiguous cases (*Supplementary Text 1*).

Across studies selected for this meta-analysis, there is considerable variability in the administration of the SST, and the analytical procedures used to derive outcome measures. To determine SST validity, we used the recent consensus guide developed by Verbruggen et al. (2019), which offers 12 'best practice' recommendations on how to design, implement and analyze the SST. We selected four main criteria from this consensus guide and rephrased them into four dichotomous items, i.e., item fulfilled or not, for the critical appraisal (*Supplementary Text 2*). In case of missing information, the criterion was rated as not fulfilled.

The overall SST validity was then rated as follows: <3 criteria fulfilled = low validity; 3 criteria fulfilled = moderate validity; 4 criteria fulfilled = high validity. Cohen's unweighted kappa for nominal data were calculated for each item. In case of a bias or a prevalence problem, Byrt's bias and prevalence adjusted kappa were additionally reported (Byrt et al., 1993; Hallgren, 2012). For the RoB assessment, we applied the adapted Hombrados and Waddington criteria that have been recently used for studies with ADHD patients (Hulsbosch et al., 2021): (1) Equivalent group sizes; (2) Use of a diagnostic interview or questionnaires to determine ADHD diagnosis; (3) Sufficient sample sizes; (4) All statistical outcomes are reported; (5) Transparent report of the data analysis; (6) Report of missing/excluded data. Each item was rated "good/low RoB", "satisfactory/moderate RoB", or "poor/high RoB". In accordance with the rating system described in the Cochrane Handbook, an overall quality rating of low, moderate or high was applied to each study based on the following criteria: If at least one of the categories was rated as having a moderate RoB, the overall RoB could only be rated as moderate as well, even if all other categories were considered having a low RoB. The same principle applied when at least one category was appraised as having a high RoB. Cohen's weighted Kappa was calculated for each individual domain (Cohen, 1968; Hallgren, 2012). After all studies were rated regarding SST validity and RoB, a variable of overall study quality

combining the RoB ratings and SST validity was created. Studies with high RoB and low SST validity were rated as having low overall quality, studies with moderate or low RoB AND moderate or high SST validity were rated as having moderate to high overall quality. Studies characterized with the remaining combinations of RoB and SST validity (low RoB and low SST validity; moderate RoB and low SST validity; high RoB and moderate SST validity; high RoB and high SST validity) were designated to the category moderate to low overall quality. This categorization was used for subgroup analysis (see below).

Meta-analysis

The meta-analysis was carried out in R (version 4.0.3; R Core Team, 2020) and the metafor package (version 3.0.2; Viechtbauer, 2010). Hedges' g was calculated for each individual study and each outcome (primary outcome SSRT and secondary outcomes) displaying the effect size of the group difference. Given that various sources could account for differences in findings across studies, e.g., examination of different patient samples or use of different SST paradigms, a random-effects model was fitted to the data. Instead of the usual large-sample approximation, the sampling variance was adjusted by taking the sample-size weighted average of the Hedges' g values into the equation, as this approach has been shown to be less biased (Lin and Aloe, 2021). For computing confidence intervals, the method introduced by Knapp and Hartung (2003) was chosen. To assess for heterogeneity, (1) τ^2 was estimated using the restricted maximum-likelihood estimator (Viechtbauer, 2005), (2) the Q-test for heterogeneity and (3) the I^2 statistics (Higgins and Thompson, 2002) are reported. If heterogeneity between studies is present, i.e., $\hat{\tau}^2 > 0$, regardless of whether the Q-test reaches significance, a prediction interval for the true outcomes is provided (Riley et al., 2011). The results will be visualized using forest plots. Furthermore, the model is assessed regarding (1) potential outliers, i.e. studies with studentized residuals larger than the $100 \times (1 - 0.05/(2 \times k))th$ percentile of a standard normal distribution, considering a Bonferroni correction with α = 0.05 (two-sided) for k included studies as well as (2) potentially overinfluential studies, i.e. with a Cook's distance larger than the median plus 6 times the

interquartile range of the Cook's distances (Viechtbauer and Cheung, 2010). If outliers were detected, leave-one-out diagnostics for sensitivity analysis were conducted.

Assessment of publication bias

Evidence of publication bias was assessed using a combination of visual and statistical approaches. First, the funnel plot (Copas and Chi, 2000) of standardized mean difference (SMD) against the inverse square root of the sample size was visually inspected for asymmetries (Zwetsloot et al., 2017). If no bias exists, the funnel plot should be symmetrical and narrow down at the top, where studies with larger sample sizes are located and estimation of the effect is more precise. However, determination of publication bias using visual inspection methods (such as funnel plots) are often subjective and prone to judgmental errors (Wang and Bushman, 1998). Therefore, it is recommended to additionally compute a quantile-quantile plot (Q-Q plot) to aid in the assessment of publication bias. Next, Egger's regression test was calculated using the inverse of the square root sample size as a predictor to statistically test for asymmetry of the funnel plot (Zwetsloot et al., 2017).

Meta-regression and subgroup analysis

To assess whether the pre-specified extracted demographic and clinical variables as well as study quality influenced the meta-analytic outcome and to explore the cause of potential heterogeneity, a meta-regression analysis was conducted for continuous (age, sex, IQ) and a subgroup analysis for categorical covariates (RoB, SST validity and overall study quality, comorbidities, patient setting and medication status).

Mixed-effect models were fitted to the data for the meta-regression analysis. If enough data were available across studies, the extracted variables were included in a multivariate regression model. Otherwise, univariate models were estimated. The parameter τ^2, which displays the residual heterogeneity not explained by the included moderators (Viechtbauer, 2010) was estimated using the REML-estimator (Viechtbauer, 2005). Tests and confidence intervals were calculated by the Knapp and Hartung (2003) method. The mean values for age, IQ, and percentage of males of the study samples were computed to be included in a

regression model. To this end, the reported means of patients and the means of controls were averaged. If sample sizes of ADHD participants and healthy controls differed substantially, a sample-size weighted mean was calculated. If IQ values were reported for verbal and non-verbal IQ separately in a study, the average of those values was calculated. IQ was centered before taken into a univariate regression model. Age and gender were standardized before taken into a multivariate regression model.

The following variables were considered for subgroup analysis: (1) RoB with the levels low RoB vs. moderate RoB vs. high RoB; (2) SST validity with the levels low validity vs. moderate validity vs. high validity; (3) overall study quality with the levels low overall quality vs. low to moderate overall quality vs. moderate to high overall quality; (4) psychiatric comorbidities in patients with the levels comorbidities allowed vs. comorbidities not allowed; (5) psychiatric comorbidities in control participants with the levels comorbidities allowed vs. comorbidities not allowed; (6) patient setting with the levels subgroups recruited from a clinical-setting vs. recruited from a non-clinical setting vs. recruited from both (mixed); (7) medication status with the levels subgroups medicated vs. unmedicated. For each of these variables, separate random-effects models were fitted. Then, a fixed-effects regression, including a moderator containing the subgroups' effect estimates, was calculated to test whether it significantly moderated SSRT.

Finally, for both meta-regressions and subgroup analyses an omnibus test of moderators was conducted, testing all coefficients excluding the intercept against 0. If the omnibus test reaches significance, this might be an indication that some of the heterogeneity could be explained by the predictors included in the model (Viechtbauer, 2010).

Secondary outcome measures of the SST

In addition to the SSRT, the SST reveals other outcome measures that are recommended to be reported (Verbruggen et al., 2019). For the current meta-analysis, we examined stop commission errors, go discrimination errors, go omission errors and go accuracy. All studies which reported these parameters were used for the meta-analyses. If errors were reported in

numbers (i.e., means, standard deviations), then the percentage of errors was calculated. The analytical procedures were the same as for SSRT, except that no meta-regression or subgroup analyses were conducted, due to the lower number of available studies.

Results

Study selection

The electronic search resulted in 1186 articles in MEDLINE and 1353 in PsycInfo (*Figure 1*). Limiters described in the methods section excluded 215 of these articles. Search results were exported into EndNote, during which EBSCOhost automatically removed 662 duplicates, resulting in 1662 studies. After removing the remaining duplicates using EndNote's automatic deduplication tool (n = 177) and manual inspection (n = 36), there were 1449 remaining articles. Screening of titles and abstracts for eligibility led to the exclusion of 1288 studies. For the remaining 161 studies, full texts were obtained and checked thoroughly for eligibility. From the 161 studies, 116 were excluded because they did not include a healthy control group (n = 1), only subclinical symptomatology was assessed (n = 2), ADHD was only included as comorbid disorder (n = 1) or no SST paradigm was used (n = 112). A further 4 studies were excluded because they had substantially modified the SST paradigm and another 8 studies because their samples included both children and adults. In six studies not sufficient statistical values for the meta-analysis were reported and authors were contacted. We received data from four studies, which were then included in the final sample. It is important to note that Bekker et al. (2005a, 2005b) and van Dongen-Boomsa et al. (2010) reported identical SSRTs, i.e., for the same experimental session and the same participant sample. The same accounted for Nigg et al. (2005), Stavro et al. (2007) and Martel et al. (2017) as well as for Linhartová et al. (2020) and Linhartová et al. (2021). For these groups of articles, the reported SSRT value was extracted and counted as a single sample in the meta-analysis. Finally, Szekely et al. (2017) conducted two SST experiments, one implemented for fMRI and one for MEG. Even though samples for these two experiments partially overlap (63 completed the SST during MEG and fMRI, 85 during fMRI only, and 33 during MEG only), they were treated as individual

observations in the analysis. The screening of reference lists revealed no further articles. In summary, there were a total of 26 publications with 27 studies included in the meta-analysis (1799 participants; ADHD = 883; controls = 916). Sample characteristics for all included studies are found in *Table 1*.

Twenty-four out of 27 studies prohibited stimulant medication on the day of testing, two studies did not report this information and one study allowed medication (Linhartová et al., 2021). One study tested the effect of stimulant medication on task performance (Chamberlain et al., 2007) and another study allowed medication during testing (Congdon et al., 2014). To maintain similarity between studies, data of Chamberlain et al., (2007) was extracted for the placebo patient group only, and data of Congdon et al. (2014) for the unmedicated patient group only. Marx et al. (2013) used an SST paradigm that compared performance with and without reward. For this study only data for the non-reward group were extracted. In some articles, information on the presence of comorbidities or the patient setting was ambiguously reported. For example, Aron et al. (2003) report that healthy controls had “no previous contact with psychiatric services” but it is unclear whether potential comorbidities of healthy controls were screened within the study. Therefore, the coding for those two variables might be biased. Upon request, Meachon et al. (2021) provided non-published information on the age and sex distribution in the two groups. Bialystok et al. (2017) provided the means of age and male percentage for the subset that completed the SST. Demographic variables, information about the IQ and other relevant information about the study population were not available for all studies. A summary is given in *Table 2*. A detailed overview of psychiatric comorbidities in patients and in controls is provided in *Table 3*.

SST validity and risk of bias

SST validity was evaluated for all 26 articles across 4 items, resulting in 104 individual ratings. *Table 4* provides an overview of the ratings. Nineteen studies (73%) were rated low validity, 5 studies (19%) moderate and 2 studies (8%) as high validity. Marginal distributions revealed a

prevalence bias for all items to some degree. This bias was strongest for item 2 and item 4. All items showed substantial to perfect interrater agreement (Hallgren, 2012) with no systematic differences between raters (*Supplementary Table 1*). Taken together, most studies included in the meta-analysis had a low or moderate quality of the SST.

RoB was evaluated for all 26 articles across 6 domains, resulting in 156 individual ratings. *Table 5* provides an overview of the RoB ratings. Overall, most studies had a moderate or high RoB. One article (4%) received a low rating, Twelve articles (46%) a moderate rating and 13 articles (50%) a high rating. For all domains, there was substantial to perfect interrater agreement (*Supplementary Table 2*). Two major sources of interrater disagreements were in the reporting of missing data (category 6) and selective outcome reporting (category 4). Some studies did not specifically address whether the whole sample has been included in the final SST analysis. However, this could be derived from the degrees of freedom in the analysis. Hence, it was decided to consider the degrees of freedom to rate this category. Moreover, four of the included studies did not report the mean and standard deviation of SST outcomes. Upon request, the authors provided us these values and the selective data reporting for these four studies were then rated as low RoB.

Meta-analysis of stop-signal reaction time

Figure 2 presents the forest plot of the observed group differences in the SSRT for 27 observations. Across studies, Hedges' g values ranged from -0.341 to 1.230. Results of the random-effects meta-analysis revealed a statistically significant moderate average effect size estimate of 0.509 ($t(26) = 7.829$, $p < 0.0001$, 95% CI: 0.376-0.644). Adults with ADHD showed moderately higher SSRTs compared to healthy controls. The I^2 statistic demonstrated moderate evidence of heterogeneity across studies ($Q(26) = 39.546$, $p = 0.043$, $\hat{\tau}^2 = 0.030$, $I^2 = 31.224\%$). The heterogeneity reflects in a 95% prediction interval ranging between 0.129 and 0.891.

According to the Cook's distances, none of the studies was overly influential. However, the study of Szekeley et al. (2017) implementing the SST for fMRI had a studentized residual

larger than ±3.113 and, hence, is an outlier in the context of this model. Leaving out this observation would reduce $\hat{\tau}^2$ to 0.000, I^2 to 0.004% and increase g to 0.524 (95% CI 0.416 to 0.631). Linhartová et al. (2021) was the only study that allowed stable stimulant medication during testing and in Chamberlain et al. (2007) patients received a placebo treatment. In order to explore whether this might have influenced the results, we conducted a sensitivity analysis excluding these two studies. In this analysis g decreased slightly to 0.498 (95% CI 0.354, 0.641), yet heterogeneity remained comparable to the original results with $\hat{\tau}^2$ = 0.034 and I^2 = 34.44%. This provides a hint that medication has no substantial influence on SSRT deficits in adult ADHD. Taken together, the random-effects meta-analysis revealed moderate effect sizes (g = 0.509 to 0.524) with larger SSRTs in patients compared to controls.

Publication bias

Figure 3A depicts a funnel plot of the studies' SMDs plotted against the inverse of the square root of the sample sizes. Egger's regression test for funnel plot asymmetry was not significant ($t(25)$ = 1.941, p = 0.064). The funnel plot seems to converge close to the average estimate with increasing sample size. A normal quantile-quantile plot is shown in *Figure 3B*. Most dots in this plot fall inside the 95%-confidence bands. However, in the middle of the line there is a slight skewing to the left, with several dots outside the bands. This is an indication that there could be a subtle publication bias.

Meta-regression and subgroup analysis

Meta-regression analysis was conducted for continuous covariates (age; sex; IQ) and a subgroup analysis for categorical covariates (RoB, SST validity; overall study quality; comorbidities; patient setting; medication status). In this analysis, the data from the fMRI study by Szekeley et al. (2017) were an outlier and therefore, they were excluded from further analysis. In four of the studies (Bialystok et al., 2017; Cherkasova et al., 2014; Lampe et al., 2007; Meachon et al., 2021) only a subset of participants completed the SST. To assess whether this might influence robustness of meta-regression results, the analysis was conducted again, excluding these 4 studies. This did not substantially influence the study

outcome (*Supplementary Tables 3-4*). *Table 6* provides an overview of the meta-regression analyses. Bialystok et al. (2017) reported demographic and outcome variables separately for monolinguals (ML) and bilinguals (BL) in each group. Values for ML and BL were averaged to obtain only a single value per group to be included in meta-regression analysis. The analysis did not reveal any significant effects of age, sex or IQ. *Table 7* provides an overview of the subgroup analyses. Some studies reported that only some psychiatric comorbidities lead to exclusion. Those were coded as "comorbidities allowed" as well. Data for years of education was sparse and heterogeneous and therefore, not included in the meta-regression. As only one study reported that participants were medicated during testing, medication status was also dropped from analysis. There were only 5 articles that did not allow comorbidities in ADHD patients. Therefore, the estimated average SMD for these studies reported in *Table 7* may not be robust. The same accounts for the estimated SMD for the level "mixed" of the setting variable, as only 2 studies report to have recruited ADHD patients from clinical as well as from non-clinical settings. Interestingly, the differences between the setting subgroups approached significance (p = .066, *Table 7*). Therefore, we conducted a follow-up analysis to explore whether there are significant differences between studies with clinical vs. non-clinical settings only, which was not the case (p = .171). The analysis of study quality revealed that both RoB assessment and SST validity ratings did not significantly moderate SSRT. For RoB, the estimated effect was largest for studies with low RoB (g = 0.651) and smallest for studies with high RoB (g = 0.531). However, there was only one study designated as having a low RoB, therefore the result for this category should be interpreted with caution. The group of studies assigned low SST validity showed the largest average effect size (g = 0.556), whereas the group rated as having high SST validity showed the smallest average effect size (g = 0.415). There were only two studies with high SST validity, limiting the reliability of the result for this category. Similar to RoB, the study quality did not significantly moderate SSRT, with an effect size g = 0.49 for studies with moderate to high overall quality ratings. Forest plots

with subgroups are shown in *Supplementary Figures 1-3*. In summary, our analysis did not reveal variables that significantly moderated the SSRT deficits in adult ADHD.

Secondary outcome measures of the SST

Fifteen studies reported the percentage of stop commissions (*Supplementary Figure 4*); 7 studies reported the percentage of choice errors (*Supplementary Figure 5*); 9 studies reported omission errors (*Supplementary Figure 6*); and 8 studies reported go accuracy (*Supplementary Figure 7*). The analysis of secondary SST outcome measures revealed no significant differences between patients and controls regarding stop commissions (g = 0.142, p = 0.064) and choice errors (g = 0.242, p = 0.078). However, ADHD patients made significantly more omission errors (g = 0.418, p = 0.01) and had a significantly lower go accuracy (g = -0.385, p < 0.008). A more detailed description of the results is provided in *Supplementary Text 3*.

Discussion

In this systematic review and meta-analysis, we integrated the data of 27 studies examining the stop-signal task in adult ADHD. The analysis revealed inhibitory control deficits, as expressed in prolonged SSRTs, with a moderate effect size g = 0.51. These deficits were not significantly moderated by the study quality, sample characteristics, or clinical parameters. In addition, the analyses of secondary outcome measures revealed greater SST omission errors and reduced go accuracy in patients, although only few (*n* <10) studies were available for these measures.

Behavioral inhibition deficits in stop-signal response times

The main finding of our meta-analysis is that patients with adult ADHD show reliable moderate deficits in the SSRT. The magnitude of the deficits fits with the outcome of an earlier meta-analysis, which included a much smaller number of studies in adult ADHD (Lipszyc and Schachar, 2010). Our meta-analysis of 27 studies establishes the SSRT as a reliable measure for the assessment of inhibitory control deficits in adult ADHD. Extending previous work, we evaluated the SST quality, using the recommendations set out by a recent

consensus paper (Verbruggen et al., 2019) and estimated the risk of bias for each study. The large number of observations enabled us to examine whether study quality, considering RoB and the validity of the SST; demographics (age and gender); IQ or clinical parameters (comorbidities and setting) influence SSRT deficits in patients. To this end, we computed meta-regression and subgroup analyses including all studies for which the respective variables were reported. Surprisingly, none of these variables significantly influenced the magnitude of SSRT deficits in patients. This implies that the prolonged SSRT in patients can be observed in experimental settings even when the study quality and other parameters are not optimal. Taken together, the finding that there were no variables which significantly moderate the SSRT deficits suggests that inhibitory control deficits can be consistently observed and hence, may be a phenotype in adult ADHD.

Another important question is how SSRT deficits relate to clinical symptoms in adult ADHD. In a large-scale study, Kamradt et al. (2014) examined correlations between SSRTs with ratings of current inattentive, hyperactive-impulsive symptoms and executive functions in patients. The study revealed significant moderate relationships between SSRTs and all symptom domains (r = 0.23 to 0.30). Using a hierarchical linear regression model, which included other neuropsychological paradigms and demographic covariates, the authors found that only the SSRT and the continuous performance test predicted inattention and hyperactivity-impulsivity total symptom scores. In a similar vein, Stavro, Nigg and colleagues (Nigg et al., 2005; Stavro et al., 2007) also found moderate (r = 0.29) relationships between SSRT deficits and executive functions, as expressed in inattentive-disorganized and hyperactive-impulsive symptoms. Thus, inhibitory control deficits, albeit frequently reported in empirical studies, are not well reflected in the diagnostic criteria of adult ADHD. Hence, it could be that these deficits are often neglected during the diagnostic process and therefore also not treated, e.g., in the framework of neurocognitive training.

It is important to note that SSRT deficits are not only found in ADHD but also in other psychiatric disorders such as obsessive-compulsive disorder (OCD), addiction or

schizophrenia (Lipszyc and Schachar, 2010; Smith et al., 2014). For instance, Lipszyc and Schachar (2010) observed SSRT deficits with effect sizes g = 0.77 and g = 0.69 in OCD and schizophrenia, respectively. However, these analyses included only few studies (N = 4 per group) and hence, an updated evaluation of the SSRT in these groups would be desirable. Nevertheless, given the overlap in inhibitory deficits across disorders, the SST presumably provides no diagnostic value in differentiating these disorders. Hence, we suggest that the SST could be used for the quantification of inhibitory control deficits in adult ADHD, following the exclusion of other psychiatric disorders in which inhibitory control deficits have been reported. Moreover, in addition to the SST, other paradigms such as the go/no-go task, could be useful for the assessment of response inhibition in adult ADHD. For instance, a meta-analysis on the go/no-go task that combined children, adolescent and adult studies, revealed deficits with a moderate effect size g = 0.49 (Wright et al., 2014), which is comparable with the SSRT deficits in current analysis. In summary, the finding of reliable moderate deficits in the SSRT implies that the SST could become a valuable tool for the neuropsychological assessment of inhibitory control deficits in adult ADHD. To this end, it would be desirable to collect SST data from large participant samples to obtain normative SSRT distributions, considering age, gender and education. An individual's performance could be then compared against the respective SSRT distribution to determine their level of performance relative to a normative sample.

Behavioral inhibition deficits in secondary measures of the SST

In addition to the SSRT, we computed meta-analyses for stop commission errors, go discrimination errors, go omission errors and go accuracy. These analyses revealed small to moderately greater omission errors (g = 0.418) and reduced go accuracies (g = -0.385) in patients. However, only a few studies have reported omission errors (n = 9) or go accuracy (n = 8), and thus, these findings should be interpreted as preliminary evidence.

For omission errors, the study with the largest reported effect size (g = 0.73) was conducted by Roberts et al. (2011). In this study, 30 adult patients with ADHD and 28 control subjects

participated in a classical SST paradigm (Logan et al., 1984). Contrary to Roberts et al. (2011), an even negative albeit not significant effect (g = -.18) was reported by Bialystok et al. (2017). In their study monolingual and bilingual patients (*n* = 28 monolingual, *n* = 28 bilingual) and controls (*n* = 36 monolingual, *n* = 37 bilingual) participated in a slightly modified version of the SST. Hence, albeit there was some variance in effect sizes across the studies included in the analysis, there were on average significant small to moderate group differences in omission errors.

A similar variability was also observed for accuracy in go trials, where the largest group differences (g = -0.64) were reported by Epstein et al. (2001) and the smallest group differences (g = 0.24) were observed by Szekely et al. (Szekely et al., 2017; fMRI study). However, this latter study can be considered as an outlier and a meta-analysis excluding this study led to an increased *g* = -0.488. In summary, there is some evidence that, in addition to the SSRT, omission errors and accuracy in go trials during the SST also reflect neurocognitive deficits in adult ADHD. Since the deficits in omission errors and accuracies are constrained to go trials, they suggest an inability to maintain an ongoing response, which indicates attentional difficulties. This is in line with previous reports of sustained and focused attention deficits in adult ADHD (Marchetta et al., 2008). Further studies should analyze and report the secondary measures or the SST, which could be then submitted to an updated meta-analysis including a higher number of observations.

Limitations

This review has some limitations. First, whilst we used an adapted version of the search syntax proposed by Lipszyc and Schachar (2010) to ensure compatibility with previous reviews, it is possible that the search strategy missed relevant studies due to the exclusion of other terms. To ensure that we detected all studies that fit our selection criteria, we thoroughly scanned the reference lists of the preselected empirical articles, previous meta-analyses and systematic reviews. Secondly, the literature search was restricted to peer-reviewed articles written in English or German. This excluded articles that were unpublished or published in a non-

commercial form. Therefore, a publication bias cannot be ruled out. Third, meta-regression analyses based on study-level-averages such as mean age of the overall study sample carry the risk of an ecological bias. For example, within studies, age might be correlated with the outcome (e.g., Congdon et al., 2014), but it might not be across studies, or the other way around (Higgins and Thompson, 2002). For this reason, the possibility that demographic or clinical variables might influence the results of the SST on the individual study level cannot be completely ruled out. Fourth, the SST validity assessment revealed that most studies did not apply cut-offs in order to identify invalid task behavior. It has been shown that ADHD adults frequently failed performance validity measurements, i.e., some participants in the studies might have completed the task incorrectly on purpose, in order to mimic cognitive deficits (Marshall et al., 2010, 2016). Therefore, the SSRT meta-analysis results might be biased at some degree. Lastly, the quality of most studies included in our meta-analysis was not optimal. Hence, we suggest that future studies should follow recently published best practice recommendations on how to design, implement, analyze and report the SST (Verbruggen et al., 2019) and apply the adapted Hombrados and Waddington criteria to ensure that a representative clinical sample is assessed (Hulsbosch et al. 2021).

Conclusion

This systematic review and meta-analysis revealed reliable moderate inhibitory control deficits, as reflected in the SST, in adult ADHD. Our meta-regression and subgroup analyses further demonstrated no significant contribution of demographic and study quality variables on the observed group differences in SSRTs. This indicates that inhibitory control deficits may be considered a phenotype in adult ADHD. Our review and meta-analysis suggest that the SST could, in conjunction with other neurocognitive tests and clinical questionnaires, become an important tool for the assessment of inhibitory control deficits in adult ADHD.

References

Adams, Z.W., Roberts, W.M., Milich, R., Fillmore, M.T., 2011. Does response variability predict distractibility among adults with attention-deficit/hyperactivity disorder? Psychol. Assess. 23, 427–436. doi:10.1037/a0022112

Adler, L.A., Faraone, S. V, Spencer, T.J., Berglund, P., Alperin, S., Kessler, R.C., 2017. The structure of adult ADHD. Int. J. Methods Psychiatr. Res. 26. doi:10.1002/mpr.1555

Aron, A.R., Dowson, J.H., Sahakian, B.J., Robbins, T.W., 2003. Methylphenidate improves response inhibition in adults with attention-deficit/hyperactivity disorder. Biol. Psychiatry 54, 1465–1468. doi:https://doi.org/10.1016/S0006-3223(03)00609-7

Asherson, P., Buitelaar, J., Faraone, S. V, Rohde, L.A., 2016. Adult attention-deficit hyperactivity disorder: key conceptual issues. The lancet. Psychiatry 3, 568–578. doi:10.1016/S2215-0366(16)30032-3

Barkley, R.A., 1997. Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. Psychol. Bull. 121, 65–94. doi:10.1037/0033-2909.121.1.65

Bekker, E.M., Overtoom, C.C., Kenemans, J.L., Kooij, J.J., De Noord, I., Buitelaar, J.K., Verbaten, M.N., 2005a. Stopping and changing in adults with ADHD. Psychol. Med. 35, 807–816.

Bekker, E.M., Overtoom, C.C.E., Kooij, J.J.S., Buitelaar, J.K., Verbaten, M.N., Kenemans, J.L., 2005b. Disentangling Deficits in Adults With Attention-Deficit/Hyperactivity Disorder. Arch. Gen. Psychiatry 62, 1129–1136. doi:10.1001/archpsyc.62.10.1129

Bialystok, E., Hawrylewicz, K., Wiseheart, M., Toplak, M., 2017. Interaction of bilingualism and Attention-Deficit/Hyperactivity Disorder in young adults. Biling. Lang. Cogn. 20, 588–601.

Boonstra, A.M., Kooij, J.J.S., Oosterlaan, J., Sergeant, J.A., Buitelaar, J.K., 2010. To act or not to act, that's the problem: Primarily inhibition difficulties in adult ADHD. Neuropsychology 24, 209–221. doi:10.1037/a0017670

Byrt, T., Bishop, J., Carlin, J.B., 1993. Bias, prevalence and kappa. J. Clin. Epidemiol. 46, 423–429. doi:10.1016/0895-4356(93)90018-v

Chamberlain, S.R., del Campo, N., Dowson, J., Müller, U., Clark, L., Robbins, T.W., Sahakian, B.J., 2007. Atomoxetine Improved Response Inhibition in Adults with Attention Deficit/Hyperactivity Disorder. Biol. Psychiatry 62, 977–984. doi:https://doi.org/10.1016/j.biopsych.2007.03.003

Cherkasova, M. V, Faridi, N., Casey, K.F., O'Driscoll, G.A., Hechtman, L., Joober, R., Baker, G.B., Palmer, J., Dagher, A., Leyton, M., Benkelfat, C., 2014. Amphetamine-Induced

Dopamine Release and Neurocognitive Function in Treatment-Naive Adults with ADHD. Neuropsychopharmacology 39, 1498–1507. doi:10.1038/npp.2013.349

Clark, L., Blackwell, A.D., Aron, A.R., Turner, D.C., Dowson, J., Robbins, T.W., Sahakian, B.J., 2007. Association Between Response Inhibition and Working Memory in Adult ADHD: A Link to Right Frontal Cortex Pathology? Biol. Psychiatry 61, 1395–1401. doi:https://doi.org/10.1016/j.biopsych.2006.07.020

Cohen, J., 1968. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol. Bull. 70, 213–220. doi:10.1037/h0026256

Congdon, E., Altshuler, L.L., Mumford, J.A., Karlsgodt, K.H., Sabb, F.W., Ventura, J., McGough, J.J., London, E.D., Cannon, T.D., Bilder, R.M., Poldrack, R.A., 2014. Neural activation during response inhibition in adult attention-deficit/hyperactivity disorder: Preliminary findings on the effects of medication and symptom severity. Psychiatry Res. Neuroimaging 222, 17–28. doi:https://doi.org/10.1016/j.pscychresns.2014.02.002

Copas, J., Chi, J.Q., 2000. Meta-analysis, funnel plots and sensitivity analysis. Biostatistics 1, 247–262. doi:10.1093/biostatistics/1.3.247

Crunelle, C.L., Veltman, D.J., van Emmerik-van Oortmerssen, K., Booij, J., van den Brink, W., 2013. Impulsivity in adult ADHD patients with and without cocaine dependence. Drug Alcohol Depend. 129, 18–24. doi:https://doi.org/10.1016/j.drugalcdep.2012.09.006

Cubillo, A., Halari, R., Ecker, C., Giampietro, V., Taylor, E., Rubia, K., 2010. Reduced activation and inter-regional functional connectivity of fronto-striatal networks in adults with childhood Attention-Deficit Hyperactivity Disorder (ADHD) and persisting symptoms during tasks of motor inhibition and cognitive switching. J. Psychiatr. Res. 44, 629–639. doi:https://doi.org/10.1016/j.jpsychires.2009.11.016

Epstein, J.N., Johnson, D.E., Varia, I.M., Conners, C.K., 2001. Neuropsychological Assessment of Response Inhibition in Adults With ADHD. J. Clin. Exp. Neuropsychol. 23, 362–371. doi:10.1076/jcen.23.3.362.1186

Fisher, T., Aharon-Peretz, J., Pratt, H., 2011. Dis-regulation of response inhibition in adult Attention Deficit Hyperactivity Disorder (ADHD): an ERP study. Clin. Neurophysiol. Off. J. Int. Fed. Clin. Neurophysiol. 122, 2390–2399. doi:10.1016/j.clinph.2011.05.010

Fried, R., DiSalvo, M., Kelberman, C., Biederman, J., 2021. Can the CANTAB identify adults with attention-deficit/hyperactivity disorder? A controlled study. Appl. Neuropsychol. Adult 28, 318–327. doi:10.1080/23279095.2019.1633328

Hadas, I., Hadar, A., Lazarovits, A., Daskalakis, Z.J., Zangen, A., 2021. Right prefrontal activation predicts ADHD and its severity: A TMS-EEG study in young adults. Prog. Neuro-Psychopharmacology Biol. Psychiatry 111, 110340. doi:https://doi.org/10.1016/j.pnpbp.2021.110340

Hallgren, K.A., 2012. Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. Tutor. Quant. Methods Psychol. 8, 23–34. doi:10.20982/tqmp.08.1.p023

Hamzeloo, M., Mashhadi, A., Fadardi, J.S., Ghahremanzadeh, M., 2018. Adult attention deficit/hyperactivity disorder among the prison inmates: An investigation of the executive function differences and comorbidity effects. Aust. J. Psychol. 70, 235–245. doi:10.1111/ajpy.12201

Higgins, J.P.T., Thompson, S.G., 2002. Quantifying heterogeneity in a meta-analysis. Stat. Med. 21, 1539–1558. doi:https://doi.org/10.1002/sim.1186

Hulsbosch, A.-K., De Meyer, H., Beckers, T., Danckaerts, M., Van Liefferinge, D., Tripp, G., Van der Oord, S., 2021. Systematic Review: Attention-Deficit/Hyperactivity Disorder and Instrumental Learning. J. Am. Acad. Child Adolesc. Psychiatry 60, 1367–1381. doi:10.1016/j.jaac.2021.03.009

Kamradt, J.M., Ullsperger, J.M., Nikolas, M.A., 2014. Executive function assessment and adult attention-deficit/hyperactivity disorder: tasks versus ratings on the Barkley deficits in executive functioning scale. Psychol. Assess. 26, 1095.

Knapp, G., Hartung, J., 2003. Improved tests for a random effects meta-regression with a single covariate. Stat. Med. 22, 2693–2710. doi:https://doi.org/10.1002/sim.1482

Lampe, K., Konrad, K., Kroener, S., Fast, K., Kunert, H.J., Herpertz, S.C., 2007. Neuropsychological and behavioural disinhibition in adult ADHD compared to borderline personality disorder. Psychol. Med. 37, 1717–1729. doi:10.1017/S0033291707000517

Lin, L., Aloe, A.M., 2021. Evaluation of various estimators for standardized mean difference in meta-analysis. Stat. Med. 40, 403–426. doi:https://doi.org/10.1002/sim.8781

Linhartová, P., Látalová, A., Barteček, R., Širůček, J., Theiner, P., Ejova, A., Hlavatá, P., Kóša, B., Jeřábková, B., Bareš, M., Kašpárek, T., 2020. Impulsivity in patients with borderline personality disorder: a comprehensive profile compared with healthy people and patients with ADHD. Psychol. Med. 50, 1829–1838. doi:10.1017/S0033291719001892

Linhartová, P., Širůček, J., Ejova, A., Barteček, R., Theiner, P., Kašpárek, T., 2021. Dimensions of impulsivity in healthy people, patients with borderline personality disorder, and patients with attention-deficit/hyperactivity disorder. J. Atten. Disord. 25, 584–595.

Lipszyc, J., Schachar, R., 2010. Inhibitory control and psychopathology: a meta-analysis of studies using the stop signal task. J. Int. Neuropsychol. Soc. 16, 1064–1076. doi:10.1017/S1355617710000895

Logan, G.D., Cowan, W.B., Davis, K.A., 1984. On the ability to inhibit simple and choice reaction time responses: a model and a method. J. Exp. Psychol. Hum. Percept. Perform. 10, 276–291. doi:10.1037//0096-1523.10.2.276

Logan, G.D., Van Zandt, T., Verbruggen, F., Wagenmakers, E.-J., 2014. On the ability to

inhibit thought and action: general and special theories of an act of control. Psychol. Rev. 121, 66–95. doi:10.1037/a0035230

Marchetta, N.D.J., Hurks, P.P.M., De Sonneville, L.M.J., Krabbendam, L., Jolles, J., 2008. Sustained and focused attention deficits in adult ADHD. J. Atten. Disord. 11, 664–676. doi:10.1177/1087054707305108

Martel, M.M., Nigg, J.T., Schimmack, U., 2017. Psychometrically Informed Approach to Integration of Multiple Informant Ratings in Adult ADHD in a Community-Recruited Sample. Assessment 24, 279–289. doi:10.1177/1073191116646443

Marshall, P., Schroeder, R., O'Brien, J., Fischer, R., Ries, A., Blesi, B., Barker, J., 2010. Effectiveness of symptom validity measures in identifying cognitive and behavioral symptom exaggeration in adult attention deficit hyperactivity disorder. Clin. Neuropsychol. 24, 1204–1237. doi:10.1080/13854046.2010.514290

Marshall, P.S., Hoelzle, J.B., Heyerdahl, D., Nelson, N.W., 2016. The impact of failing to identify suspect effort in patients undergoing adult attention-deficit/hyperactivity disorder (ADHD) assessment. Psychol. Assess. 28, 1290–1302. doi:10.1037/pas0000247

Marx, I., Höpcke, C., Berger, C., Wandschneider, R., Herpertz, S.C., 2013. The Impact of Financial Reward Contingencies on Cognitive Function Profiles in Adult ADHD. PLoS One 8, e67002. doi:10.1371/journal.pone.0067002

Meachon, E.J., Meyer, M., Wilmut, K., Zemp, M., Alpers, G.W., 2021. Evoked Potentials Differentiate Developmental Coordination Disorder From Attention-Deficit/Hyperactivity Disorder in a Stop-Signal Task: A Pilot Study. Front. Hum. Neurosci. 15. doi:10.3389/fnhum.2021.629479

Murphy, P., 2002. Inhibitory control in adults with attention-deficit/hyperactivity disorder. J. Atten. Disord. 6, 1–4.

Nigg, J.T., Butler, K.M., Huang-Pollock, C.L., Henderson, J.M., 2002. Inhibitory processes in adults with persistent childhood onset ADHD. J. Consult. Clin. Psychol. 70, 153–157. doi:10.1037//0022-006x.70.1.153

Nigg, J.T., Stavro, G., Ettenhofer, M., Hambrick, D.Z., Miller, T., Henderson, J.M., 2005. Executive functions and ADHD in adults: evidence for selective effects on ADHD symptom domains. J. Abnorm. Psychol. 114, 706.

Nikolas, M.A., Marshall, P., Hoelzle, J.B., 2019. The role of neurocognitive tests in the assessment of adult attention-deficit/hyperactivity disorder. Psychol. Assess. 31, 685–698. doi:10.1037/pas0000688

Ossmann, J.M., Mulligan, N.W., 2003. Inhibition and Attention Deficit Hyperactivity Disorder in Adults. Am. J. Psychol. 116, 35. doi:10.2307/1423334

Pironti, V.A., Lai, M.-C., Müller, U., Dodds, C.M., Suckling, J., Bullmore, E.T., Sahakian, B.J.,

2014. Neuroanatomical Abnormalities and Cognitive Impairments Are Shared by Adults with Attention-Deficit/Hyperactivity Disorder and Their Unaffected First-Degree Relatives. Biol. Psychiatry 76, 639–647. doi:10.1016/j.biopsych.2013.09.025

Riley, R.D., Higgins, J.P.T., Deeks, J.J., 2011. Interpretation of random effects meta-analyses. BMJ 342, d549. doi:10.1136/bmj.d549

Raud, L., Westerhausen, R., Dooley, N., Huster, R.J., 2020. Differences in unity: The go/no-go and stop signal tasks rely on different mechanisms. Neuroimage. 210:116582. doi: 10.1016/j.neuroimage.2020.116582

Roberts, W., Fillmore, M.T., Milich, R., 2011. Linking impulsivity and inhibitory control using manual and oculomotor response inhibition tasks. Acta Psychol. (Amst). 138, 419–428. doi:https://doi.org/10.1016/j.actpsy.2011.09.002

Sebastian, A., Gerdes, B., Feige, B., Klöppel, S., Lange, T., Philipsen, A., Tebartz van Elst, L., Lieb, K., Tüscher, O., 2012. Neural correlates of interference inhibition, action withholding and action cancelation in adult ADHD. Psychiatry Res. Neuroimaging 202, 132–141. doi:https://doi.org/10.1016/j.pscychresns.2012.02.010

Silverstein, M.J., Faraone, S. V, Leon, T.L., Biederman, J., Spencer, T.J., Adler, L.A., 2020. The Relationship Between Executive Function Deficits and DSM-5-Defined ADHD Symptoms. J. Atten. Disord. 24, 41–51. doi:10.1177/1087054718804347

Slusarek, M., Velling, S., Bunk, D., Eggers, C., 2001. Motivational effects on inhibitory control in children with ADHD. J. Am. Acad. Child Adolesc. Psychiatry 40, 355–363. doi:10.1097/00004583-200103000-00016

Smith, J.L., Mattick, R.P., Jamadar, S.D., Iredale, J.M., 2014. Deficits in behavioural inhibition in substance abuse and addiction: a meta-analysis. Drug Alcohol Depend. 145, 1–33. doi:10.1016/j.drugalcdep.2014.08.009

Song, P., Zha, M., Yang, Q., Zhang, Y., Li, X., Rudan, I., 2021. The prevalence of adult attention-deficit hyperactivity disorder: A global systematic review and meta-analysis. J. Glob. Health 11, 4009. doi:10.7189/jogh.11.04009

Stavro, G.M., Ettenhofer, M.L., Nigg, J.T., 2007. Executive functions and adaptive functioning in young adult attention-deficit/hyperactivity disorder. J. Int. Neuropsychol. Soc. 13, 324–334.

Szekely, E., Sudre, G.P., Sharp, W., Leibenluft, E., Shaw, P., 2017. Defining the neural substrate of the adult outcome of childhood ADHD: a multimodal neuroimaging study of response inhibition. Am. J. Psychiatry 174, 867–876.

van Dongen-Boomsma, M., Lansbergen, M.M., Bekker, E.M., Sandra Kooij, J.J., van der Molen, M., Kenemans, J.L., Buitelaar, J.K., 2010. Relation between resting EEG to cognitive performance and clinical symptoms in adults with attention-deficit/hyperactivity disorder. Neurosci. Lett. 469, 102–106. doi:https://doi.org/10.1016/j.neulet.2009.11.053

Verbruggen, F., Aron, A.R., Band, G.P., Beste, C., Bissett, P.G., Brockett, A.T., Brown, J.W., Chamberlain, S.R., Chambers, C.D., Colonius, H., Colzato, L.S., Corneil, B.D., Coxon, J.P., Dupuis, A., Eagle, D.M., Garavan, H., Greenhouse, I., Heathcote, A., Huster, R.J., Jahfari, S., Kenemans, J.L., Leunissen, I., Li, C.-S.R., Logan, G.D., Matzke, D., Morein-Zamir, S., Murthy, A., Paré, M., Poldrack, R.A., Ridderinkhof, K.R., Robbins, T.W., Roesch, M., Rubia, K., Schachar, R.J., Schall, J.D., Stock, A.-K., Swann, N.C., Thakkar, K.N., Van Der Molen, M.W., Vermeylen, L., Vink, M., Wessel, J.R., Whelan, R., Zandbelt, B.B., Boehler, C.N., 2019. A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. Elife 8. doi:10.7554/elife.46323

Verbruggen, F., Logan, G.D., 2009. Models of response inhibition in the stop-signal and stop-change paradigms. Neurosci. Biobehav. Rev. 33, 647–661. doi:10.1016/j.neubiorev.2008.08.014

Verbruggen, F., Logan, G.D., 2008. Response inhibition in the stop-signal paradigm. Trends Cogn. Sci. 12, 418–424. doi:10.1016/j.tics.2008.07.005

Viechtbauer, W., 2010. Conducting Meta-Analyses in R with the metafor Package. J. Stat. Softw. 36, 1–48. doi:10.18637/jss.v036.i03

Viechtbauer, W., 2005. Bias and Efficiency of Meta-Analytic Variance Estimators in the Random-Effects Model. J. Educ. Behav. Stat. 30, 261–293. doi:10.3102/10769986030003261

Viechtbauer, W., Cheung, M.W.-L., 2010. Outlier and influence diagnostics for meta-analysis. Res. Synth. Methods 1, 112–125. doi:https://doi.org/10.1002/jrsm.11

Wang, M.C., Bushman, B.J., 1998. Using the normal quantile plot to explore meta-analytic data sets. Psychol. Methods. doi:10.1037/1082-989X.3.1.46

Wright, L., Lipszyc, J., Dupuis, A., Thayapararajah, S.W., Schachar, R., 2014. Response inhibition and psychopathology: a meta-analysis of go/no-go task performance. J. Abnorm. Psychol. 123, 429–439. doi:10.1037/a0036295

Zwetsloot, P.-P., Van Der Naald, M., Sena, E.S., Howells, D.W., Inthout, J., De Groot, J.A., Chamuleau, S.A., Macleod, M.R., Wever, K.E., 2017. Standardized mean differences cause funnel plot distortion in publication bias assessments. Elife 6. doi:10.7554/elife.24260

Tables

Table 1

Study	Patient setting	*n*	Males %	Age *M (SD)*	Subtypes *n* (%)	IQ *M (SD)*	Years of education	Medication -/+ (h)	Current CD (*n*)
Adams et al. (2011)	Not enrolled	30 ADHD	56.67	21.1 (1.7)	*n.a.*	103.2 (10.8) V 105.6 (9.8) NV	15.1 (1)	- (24h)	-
		27 HC	48.15	22.0 (1.7)		107.1 (6.9) V 111.2 (8.8) NV	15.3 (1.3)		-
Aron et al. (2003)	Patient	13 ADHD	76.92	26.2 (6.9)	3 (23.08) IA	109 (7.2) V	*n.a.*	- (≥24h)	+
		13 HC	61.54	30.5 (5.0)	8 (61.54) C 2 (15.38) PR	114 (4.3) V			-
Bekker et al. (2005a); Bekker et al. (2005b); van Dongen-Boomsma et al. (2010)	Patient	24 ADHD	50	34.3 (11.68)	24 (100) C	*n.a.*	*n.a.*	- (≥ 6 times half-life)	+
		24 HC	50	34.9 (*n.a.*)					-
Bialystok et al. (2017)	Not enrolled	56 (50) ADHD	60	23.55 (4.13)	*n.a.*	100 (12.7) ML NV 103.4 (9.6) ML V 100.9 (13.5) BL NV 100.9 (11.1) BL V	15.5 (2.0) ML 16.1 (2.2) BL	- (24h)	*n.a.*
		72 (54) HC	29.63	21.20 (2.74)		98.1 (12.1) ML NV 102.6 (10.9) ML V 100.7 (14.5) BL NV 95.8 (11.8) BL V	14.9 (1.9) ML 13.9 (1.6) BL		
Boonstra et al. (2010)	Patient	49 ADHD	53.06	38.7 (9.7)	2 (4.08) HI	100.6 (17.8)	*n.a.*	-	+
		49 HC	53.06	38.1 (9.3)	47 (95.92) C	107.71 (16.5)			-
Chamberlain et al. (2007)	Patient	20 ADHD	70	31.60 (8.33)	6 (30) IA	109.9 (9.2)	*n.a.*	- (≥12h)	+
		20 HC	70	30.90 (7.93)	13 (65) C 1 (5) PR	112.1 (6.2)			-
Cherkasova et al. (2014)	*n.a.*	15 (14) ADHD	100	29.87 (8.65)	10 (66.67) IA	107.13 (12.78)	16.20 (3.63)	-	-
		18 (12) HC	100	25.44 (6.77)	5 (33.33) C	116.83 (16.07)	17.11 (3.32)		-
Clark et al. (2007)	Patient	20 ADHD	65	28.0 (8.6)	4 (20) IA	108.3 (5.9)	13.7 (1.7)	- (24h)	+
		16 HC	87.5	25.1 (5.4)	2 (10) HI 10 (50) C	113.3 (3.5)	14.4 (3.2)		+

Study	Patient setting	*n*	Males %	Age *M (SD)*	Subtypes *n* (%)	IQ *M (SD)*	Years of education	Medication -/+ (h)	Current CD (*n*)
					2 (10) PR 2 (10) NOS				
Congdon et al. (2014)	*n.a.*	25 ADHD 62 HC	56 45	31.24 (10.37) 30.82 (8.97)	*n.a.*	*n.a.*	14.28 (1.74) 15.10 (1.75)	-	- -
Crunelle et al. (2013)	Patient	17 ADHD 17 HC	100 100	33 (7) 31 (6)	8 (47) IA 9 (53) C	105 (4) 106 (4)	*n.a.*	-	+ -
Cubillo et al. (2010)	Not enrolled	10 ADHD 14 HC	100 100	28 (1) 28 (2)	*n.a.*	90 (8) 113 (11)	*n.a.*	-	+ -
Epstein et al. (2001)	Mixed	25 ADHD 30 HC	40 50	33.6 (*n.a.*) 33.4 (*n.a.*)	14 (56) IA 1 (4) HI 10 (40) C	*n.a.*	*n.a.*	-	+ +
Hadas et al. (2021)	Not enrolled	52 ADHD 49 HC	80.36 67.31	25.7 (0.5) 26 (0.3)	*n.a.*	*n.a.*	*n.a.*	- (1 week)	- *n.a.*
Hamzeloo et al. (2018)	Not enrolled	30 ADHD 30 HC	100	29.38 (6.10)	*n.a.*	*n.a.*	7.78, (2.48)	*n.a.*	+ +
Kamradt et al. (2014)	Mixed	170 ADHD 83 HC	54.7 56.1	23.8 (4.7) 20.1 (2.9)	65 (38.2) IA 10 (5.8) HI 95 (55.8) C	*n.a.*	*n.a.*	- (24h to 48h)	+ +
Lampe et al. (2007)	Patient	22 (16) ADHD 20 (17) HC	63.63 30	29.95 (8.2) 28.7 (6.9)	14 (63.64) IA 1 (4.55) HI 7 (31.82) C	111.00 (11.6) 114.2 (8.5)	*n.a.*	- (4 weeks)	- -
Linhartová et al. (2021)	Patient	26 ADHD 26 HC	73 69	23.88 (8.14) 23.69 (7.49)	*n.a.*	*n.a.*	*n.a.*	+	+ -.
Marx et al. (2013)	Patient	18 ADHD 20 HC	61 45	27.72 (6.21) 24.75 (3.63)	4 (18.42) IA 1 (2.63) HI 13 (78.95) C	123.33 (16.82) 127.65 (12.33)	*n.a.*	- (≥72h)	+ -
Meachon et al. (2021)	*n.a.*	9 (8) ADHD 22 (19) HC	33.33 36.36	26.44 (5.83) 23.19 (5.61)	*n.a.*	*n.a.*	*n.a.*	-	+ -
Murphy et al. (2002)	Not enrolled	18 ADHD 18 HC	100 100	27-58 25-59	18 (100) C	110 (9.23) 116 (11.48)	*n.a.*	*n.a.*	+ +
Nigg et al. (2005); Stavro et al. (2007)	Not enrolled	105 ADHD 90 HC	67.6 35.6	23.70 (4.28) 24.64 (4.77)	26 (24.76) IA 5 (4.76) HI	110.80 (11.59) 113.23 (10.10)	*n.a.*	- (≥24h to 48h)	+ +

Study	Patient setting	*n*	Males %	Age *M (SD)*	Subtypes *n* (%)	IQ *M (SD)*	Years of education	Medication -/+ (h)	Current CD (*n*)
					28 (26.67) C 21 (20) IN 25 (23.08) PR				
Ossmann and Mulligan (2003)	Not enrolled	24 ADHD 24 HC	58.33 58.33	19.21 (1.18) 19.42 (1.06)	*n.a.*	116.71 (8.74) 116.33 (10.20)	*n.a.*	- (≥12h)	*n.a.* *n.a.*
Pironti et al. (2014)	Patient	20 ADHD 20 HC	85 65	32.2 (10.31) 32.55 (5.8)	4 (20) IA 16 (80) C	115.26 (6.15) 119.49 (3.27)	*n.a.*	- (≥24h)	- -
Roberts et al. (2011)	Not enrolled	30 ADHD 28 HC	56.7 46.43	21.1 (1.7) 22.1 (1.7)	*n.a.*	104.9 (10.1) 109.9 (6.9)	15.1 (1.0 15.3 (1.2)	- (≥24 h)	+ +
Sebastian et al. (2012)	Patient	20 ADHD 24 HC	55 45.83	33.3 (8.9) 30.3 (8.1)	9 (45) IA 11 (55) C	115.3 (16.7) 115.7 (16.0)	*n.a.*	- (≥2 months)	+ -
Szekely et al. (2017), fMRI	*n.a.*	24 ADHD 84 HC	45.8 57.1	23.34 (3.95) 24.46 (4.09)	*n.a.*	*n.a.*	*n.a.*	- (≥24h)	+ -
Szekely et al. (2017), MEG	*n.a.*	25 ADHD 46 HC	52.0 47.8	23.73 (4.18) 23.31 (2.96)	*n.a.*	*n.a.*	*n.a.*	- (≥24h)	+ -

Table 1: Studies investigating the stop-signal task in adult ADHD. ADHD: attention-deficit hyperactivity disorder; HC: healthy controls; Patient Setting: ADHD group recruited from in/outpatient setting or non-clinical setting; Years of education: If education status was presented in any other form than years of education, it was not included in this table. Medication: allowed or not allowed during testing (+ and -, respectively). If medication was not allowed, duration of omission prior to testing in brackets, if medication was allowed, percentage of medicated participants in brackets; CD: comorbid disorder; V: verbal; NV: nonverbal; ML: monolingual; BL: bilingual; IA: predominantly inattentive subtype; HI: predominantly hyperactive-impulsive subtype; C: combined subtype; PR: in partial remission; NOS: Not otherwise specified; IN: inconsistent, met criteria for ADHD–H, ADHD–C, or ADHD–I as children but for a different subtype as an adult.

Table 2

	ADHD patients	Healthy controls	Total
N	883	916	1799
Age (*k* = 26)	27.73 (4.92)	26.98 (4.92)	27.44 (4.72)
Males, % (*k* = 27)	67.19 (19.53)	61.30 (22.66)	64.25 (20.08)
IQ (*k* = 17)	108.41 (7.48)	113.17 (6.08)	110.85 (6.12)
Patient Setting, *k*			
Clinical	11	*n.a.*	11
Non-clinical	9		9
Mixed	2		2
Comorbidities, *k*			
Allowed	19	7	*n.a.*
Not allowed	6	17	
Subtypes, *n* (*k* = 15)			
Primarily Inattentive	167	*n.a.*	167
Primarily Hyperactive	314		314
Combined	22		22
In Partial Remission	30		30
Inconsistent	21		21
NOS	2		2

Table 2: Sample characterization. n: number of participants, k: number of studies reporting this information; Males: percentage of males in the sample; clinical: recruited from a clinical (inpatient/ outpatient) setting; non-clinical: recruited from a non-clinical setting (e.g., newspaper, university); mixed: recruited both from a clinical and a non-clinical setting; In partial remission: met at least 6 of 9 DSM-5 criteria in childhood, but only 3 to 5 of 9 criteria in adulthood. Inconsistent: met the diagnostic criteria for a different subtype in childhood than in adulthood; NOS: not otherwise specified.

Table 3

	ADHD		Healthy Controls	
	Current	*Past/Lifetime*	*Current*	*Past/Lifetime*
Adams et al., 2010	None	Depression and/or anxiety (*n* = 5) Alcohol abuse (*n* = 2) Learning disability (*n* = 1)	None	Depression and/or anxiety (*n* = 4) Alcohol abuse (*n* = 1)
Aron et al., 2003	Global Severity Index (GSI) score of the Brief Symptom Inventory[1]: *M* = 1.7 ± .9	*n.a.*	Control subjects had no previous contact with psychiatric services.	
Bekker et al., 2005	Depression (*n* = 2) Anxiety disorders (*n* = 8) Substance abuse (*n* = 3)	Depression (*n* = 13) Bipolar disorder (*n* = 1) Tic disorder (*n* = 1) Alcohol dependence (*n* = 1)	Controls reported no psychiatric disorders or developmental disorder in childhood.	
Bialystok et al., 2017	*n.a.*	*n.a.*	*n.a.*	*n.a.*
Boonstra et al., 2010	Mood disorders (*n* = 13) Anxiety disorders (*n* = 23) Substance abuse (*n* = 7) Substance dependence (*n* = 8) Others (*n* = 6)	Mood disorders (*n* = 18) Anxiety disorders (*n* = 11) Substance abuse (*n* = 14) Substance dependence (*n* = 21) Others (*n* = 6)	Substance dependence (*n* = 2, nicotine)	Mood disorders (*n* = 5) Substance abuse (*n* = 10) Substance dependence (*n* = 6)

	ADHD		Healthy Controls	
	Current	*Past/Lifetime*	*Current*	*Past/Lifetime*
Chamberlain et al., 2007	GSI: *M* = 1.28 ± .72	*n.a.*	Controls reported no current or history of axis I disorder.	
Cherkasova et al., 2014	None	None	None	None
Clark et al., 2007	Exclusion criteria were a current mood, psychotic, or substance-related diagnosis.			
Congdon et al., 2014	Participants were excluded for lifetime diagnoses of schizophrenia or other psychotic disorders, bipolar I or II disorder; or current major depressive disorder, suicidality, anxiety disorder (obsessive-compulsive disorder, panic disorder, generalized anxiety disorder, post-traumatic stress disorder), or substance abuse/dependence other than nicotine dependence.			
Crunelle et al., 2013	Participants were excluded when currently using any drugs other than alcohol, cannabis, or nicotine.		Psychiatric disorders were excluded in healthy controls.	
Cubillo et al., 2010	Anxiety disorder (*n* = 1) Mood disorder (*n* = 3) Conduct disorder (*n* = 1) Substance related (*n* = 2)	*n.a.*	For healthy controls, exclusion criteria were present or past history of any mental disorder and substance abuse.	
Epstein et al., 2001	MDD (*n* = 2) (Hypo)manic episode (*n* = 2) Alcohol abuse/dependence (*n* = 5) Other drug abuse/ dependence (*n* = 1)	*n.a.*	Alcohol abuse/ dependence (*n* = 6) Other drug abuse/ dependence (*n* = 1)	*n.a.*

	ADHD		Healthy Controls	
	Current	*Past/Lifetime*	*Current*	*Past/Lifetime*
Hadas et al., 2021	Neuropsychiatric comorbidities, and use of psychiatric drugs were ruled out.		*n.a.*	*n.a.*
Hamzeloo et al., 2018	Psychiatric comorbidity symptoms for both groups were evaluated (i.e., anxiety, depression, bipolar disorder, PTSD, Antisocial PD and BPD).			
Kamradt et al., 2014	Participants with a history of Tourette's disorder, schizophrenia or psychosis, or autism spectrum disorder were excluded.			
Lampe et al., 2007	None	BPD (*n* = 6) Substance abuse disorder (*n* = 1) Eating disorder (*n* = 4) Anxiety disorder (*n* =1) PTSD (*n* = 1)	None	None
Linhartová et al., 2021	Comorbid psychotic or affective disorder and addiction lead to exclusion.		For healthy controls the absence of psychiatric symptoms was confirmed.	
Marx et al., 2013	Depressive disorders (*n* = 2, currently remitted) Adjustment disorder (*n* = 1) Binge-eating (*n* = 2) Personality disorders (*n* = 9, except for BPD)	*n.a.*	Within the control group, no psychiatric or personality disorders were observed.	
Meachon et al., 2021	ADHD patients reported no history of brain damage or other developmental impairments.		The control group additionally reported no history of any psychiatric or other health conditions.	
Murphy et al., 2002	Subjects had to be free of psychosis, major depression, and mania, and were screened for current alcohol and drug abuse or dependence.			

	ADHD		Healthy Controls	
	Current	*Past/Lifetime*	*Current*	*Past/Lifetime*
	n.a.	Alcohol abuse/ dependence (*n* = 1) Other drug abuse/ dependence (*n* = 1)	No healthy control met criteria for current alcohol or drug abuse or dependence.	
Nigg et al., 2005	Anxiety disorder (*n* = 19) Antisocial PD (*n* = 5)	Alcohol dependence (*n* = 16) Drug dependence (*n* = 11) Substance dependence (*n* = 24) MDD (*n* = 32)	Anxiety disorder (*n* = 11)	Alcohol dependence (*n* = 4) Drug dependence (*n* = 3) Substance dependence (*n* = 7) MDD (*n* = 11)
Ossmann & Mulligan, 2003	*n.a.*	*n.a.*	*n.a.*	*n.a.*
Pironti et al., 2015	Participants did not show relevant symptoms of a comorbid disorder reaching clinical significance for a formal DSM-IV Text Revision diagnosis.			
Roberts et al., 2011	Depression and/or anxiety (*n* = 5) Alcohol abuse (*n* = 2) Learning disability (*n* = 1)		Depression and/or anxiety (*n* = 4) Bipolar disorder (*n* = 1) Alcohol abuse (*n* = 1)	
Sebastian et al., 2012	Dysthymia (*n* = 4) Social phobia (*n* = 3) Specific phobia (*n* = 2) Anxiety disorder (*n* = 1) Substance abuse	MDD (*n* = 6) Obsessive compulsive disorder (*n* = 1) Substance abuse (*n* = 4) Substance dependence (*n* = 1) Eating disorder (*n* = 3)	Healthy controls had no lifetime history of axis I or axis II disorders.	

	ADHD		Healthy Controls	
	Current	*Past/Lifetime*	*Current*	*Past/Lifetime*
	(*n* = 1)			
	Dependent PD (*n* = 5)			
	Schizoid PD (*n* = 1)			
Szekely et al., 2017 fMRI	Any comorbid disorder (*n* = 12)		None	
Szekely et al., 2017 MEG	Any comorbid disorder (*n* = 12)		None	

Table 3: Detailed comorbidity information for patients and healthy controls. Current: diagnoses that patients and/or healthy controls currently meet the criteria for; Past/Lifetime: diagnoses that patients and/or healthy controls met the criteria for in the past; BPD: borderline personality disorder; PTSD: post-traumatic stress disorder; PD: personality disorder; MDD: major depressive disorder;

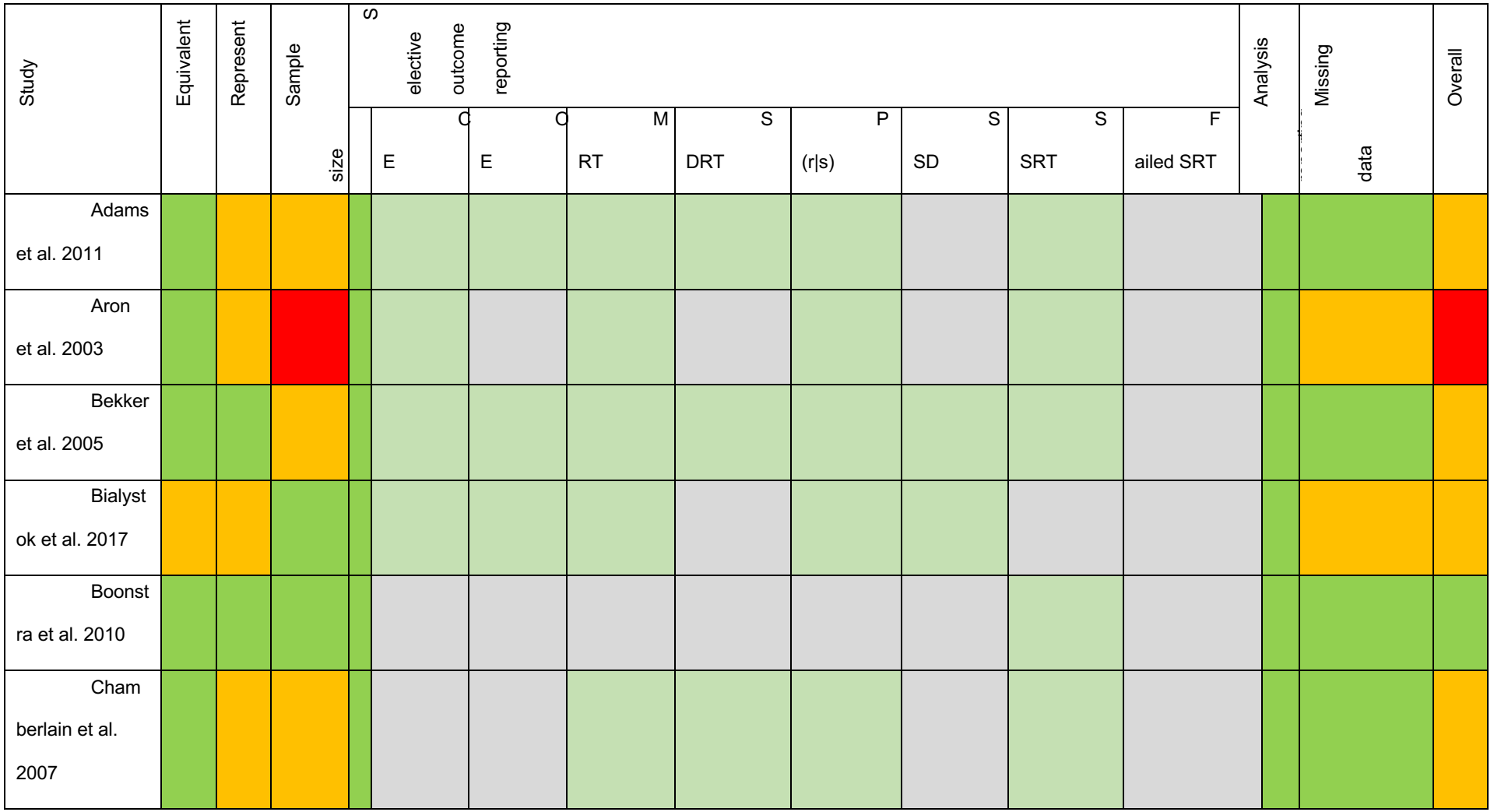
[1]Derogatis, L.R., (1993): Brief Symptom Inventory (BSI): Administrative, Scoring and Procedural Manual. 3rd ed. Minneapolis, MN: National Computer Systems.

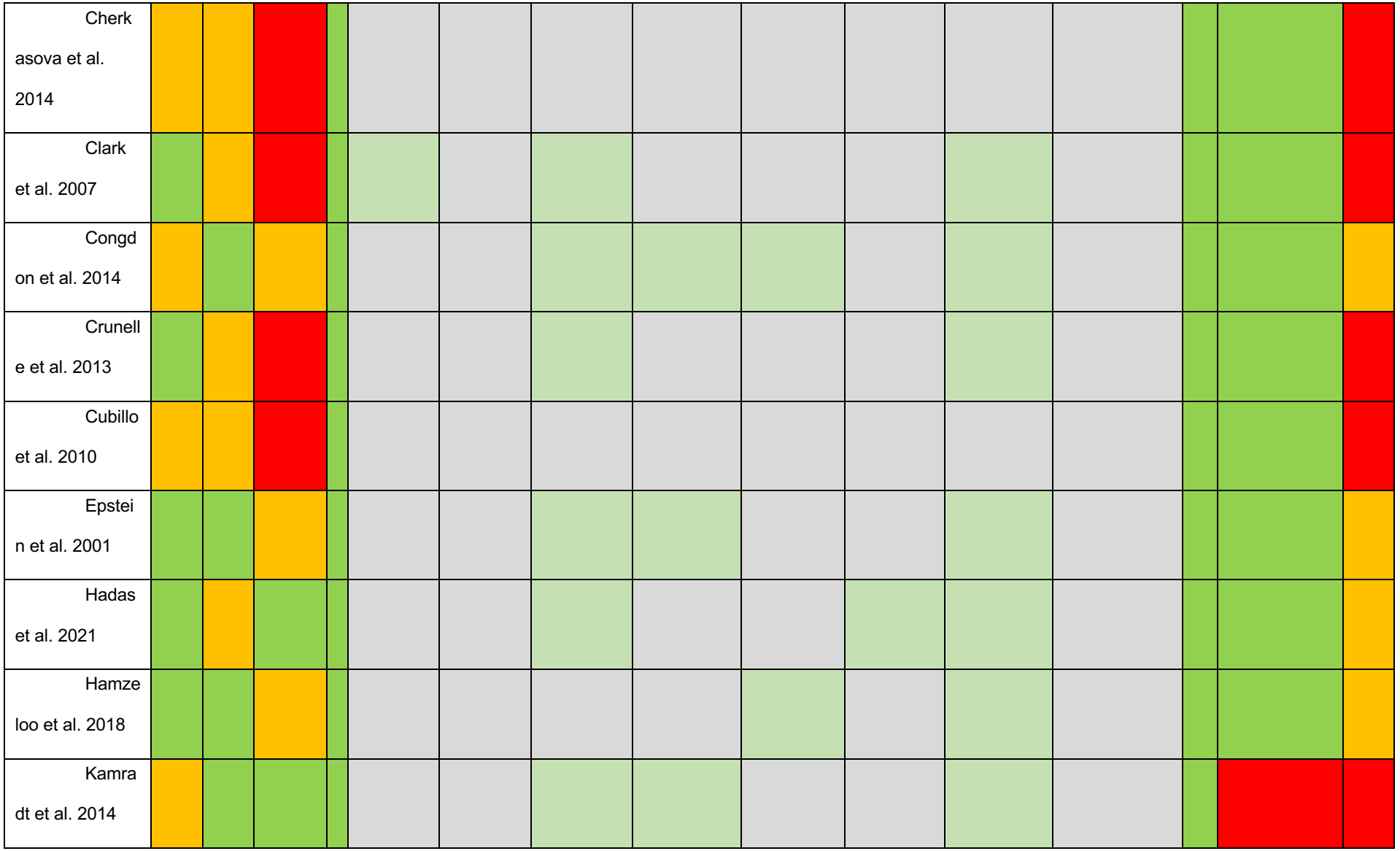
Table 4

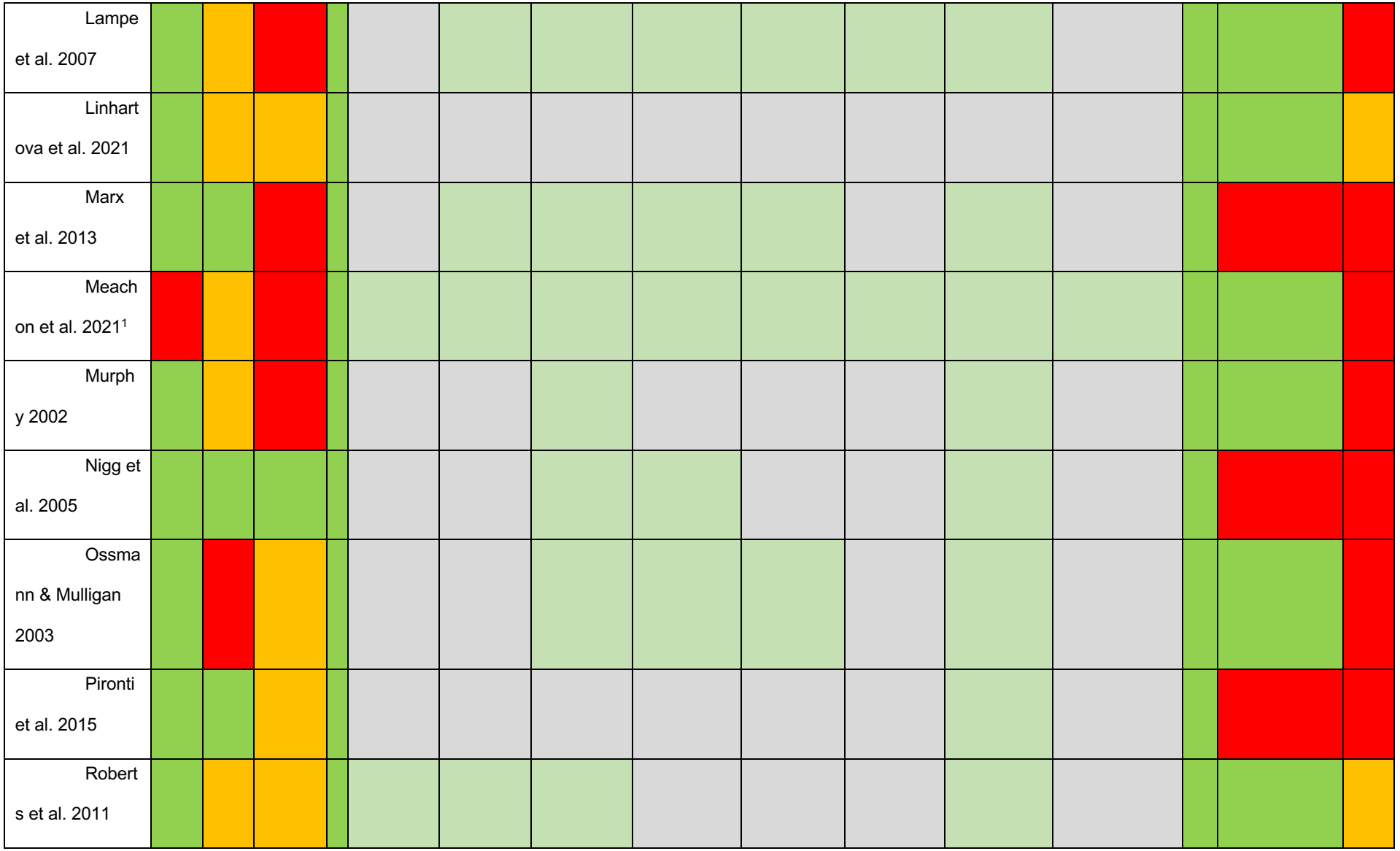
Study	Item 1	Item 2	Item 3	Item 4	Overall
Adams et al. 2011					Low
Aron et al. 2003					Low
Bekker et al. 2005					Moderate
Bialystok et al. 2017					High
Boonstra et al. 2010					Low
Chamberlain et al. 2007					Low
Cherkasova et al. 2014					Low
Clark et al. 2007					Low
Congdon et al. 2014					Moderate
Crunelle et al. 2013			-		Low
Cubillo et al. 2010					Low
Epstein et al. 2001					Low
Hadas et al. 2021					Low
Hamzeloo et al., 2018					Low
Kamradt et al. 2014					Low
Lampe et al. 2007					Moderate
Linhartová et al. 2021					Low
Marx et al. 2013					Low
Meachon et al. 2021					High
Murphy 2002					Low
Nigg et al. 2005					Moderate
Ossmann & Mulligan 2003					Low
Pironti et al. 2015					Low
Roberts et al. 2011					Low
Sebastian et al. 2012					Low
Szekely et al. 2017					Moderate

Table 4: Stop-signal task validity ratings. Item 1: ≥50 stop trials in total, stop trials constituting ≤25% of all trials; Item 2: staircase algorithm implemented; Item 3: integration method used; Item 4: Cut-Offs applied to ensure valid SSRT estimation; green: fulfilled; red: not fulfilled; Low: 1 or 2 out of 4 items fulfilled; Moderate: 3 out of 4 items fulfilled; High: 4 out of 4 items fulfilled.

Table 5


Study
Equivalent
Represent
Sample size
Selective outcome reporting
CE
OE
MRT
SDRT
P(r|s)
SSD
SSRT
Failed SRT
Analysis
Missing data
Overall
Adams et al. 2011
Aron et al. 2003
Bekker et al. 2005
Bialystok et al. 2017
Boonstra et al. 2010
Chamberlain et al. 2007


Cherkasova et al. 2014
Clark et al. 2007
Congdon et al. 2014
Crunelle et al. 2013
Cubillo et al. 2010
Epstein et al. 2001
Hadas et al. 2021
Hamzeloo et al. 2018
Kamradt et al. 2014


Lampe et al. 2007
Linhartova et al. 2021
Marx et al. 2013
Meachon et al. 2021[1]
Murphy 2002
Nigg et al. 2005
Ossmann & Mulligan 2003
Pironti et al. 2015
Roberts et al. 2011

Sebastian et al. 2012															
Szekely et al. 2017															

Table 5: Risk of bias ratings. Ratings were based on the adapted Hombrados and Waddington criteria (Hulsbosch et al., 2021)*;* CE: choice errors; OE: omission errors; MRT: mean reaction time; SDRT: intrasubject variability; P(r|s): probability to respond on a stop trial; SSD: stop signal delay; SSRT: stop signal reaction time; Failed SRT: failed stop reaction time; green: good/low RoB; orange: satisfactory/moderate RoB; red: poor/high RoB. [1]This study was a pilot study, which might be the reason for small sample sizes.

Table 6

Moderator	*B (SE)*	*t*	*p*	*ci*	*F*-Test	p_F
Age, Sex (*k* = 25, *n* = 1655)					*F*(3,21) = 0.558	.649
Intercept	0.513 (0.056)	9.182	<.001	0.397, 0.629		
Age	0.085 (0.078)	1.097	.285	-0.077, 0.248		
Sex	-0.003 (0.072)	-0.037	.971	-0.153, 0.148		
Age:Sex	0.054 (0.116)	0.468	.644	-0.187, 0.296		
IQ (*k* = 17, *n* = 937)					*F*(1,15) = 0.345	.566
Intercept	0.601 (0.079)	7.574	<.001	0.432, 0.770		
IQ	0.008 (0.013)	0.587	.566	-0.020, 0.036		

Table 6: Meta-regression analyses for SSRT. k: number of studies for which data was available; n: number of participants used for analysis. B: regression coefficient. For categorical variables, B is the average estimated effect size for each individual factor level; SE: standard error of regression coefficient; t: t-test for the regression coefficient; p: p-value for regression coefficient t-test; ci: confidence interval; F-Test: omnibustest of moderator; p_F: p-value for test of moderator; Sex: percentage of males in the individual study samples; IQ: for ADHD and control group combined; Setting: patient setting of ADHD group.

Table 7

Moderator	*B (SE)*	*z*	*p*	*ci*	Q_M-*Test*	p_Q
Risk of bias					$Q_M(2) = 0.260$	.878
High (*k* = 13, *n* = 816)	0.531 (0.116)	4.583	<.001	0.304, 0.759		
Moderate (*k* = 12, *n* = 777)	0.023 (0.126)	0.186	0.852	-0.224, 0.271		
Low (*k* = 1, *n* = 98)	0.120 (0.236)	0.508	0.611	-0.342, 0.582		
SST validity					$Q_M(2) = 0.593$	.743
Low (*k* = 19, *n* = 1126)	0.556 (0.062)	8.919	<.001	0.434, 0.678		
Moderate (*k* = 5, *n* = 434)	-0.036 (0.158)	-0.227	0.820	-0.345, 0.273		
High (*k* = 2, *n* = 131)	-0.141 (0.186)	-0.760	0.447	-0.505 ,0.223		
Overall quality					$Q_M(2) = 0.173$	.917
Low (*k* =10, *n* = 561)	0.558 (0.140)	3.974	<.001	0.283, 0.833		
Moderate/Low (*k* = 12, *n* = 820)	-0.010 (0.151)	-0.066	0.948	-0.306, 0.286		
Moderate/High (*k* = 4, *n* = 310)	-0.065 (0.189)	-0.343	0.732	-0.435, 0.306		
Psychiatric comorbidities						
In Patients					$Q_M(1) = 0.132$	.716
Allowed (*k* = 18, *n* = 1195)	0.520 (0.065)	7.980	<.001	0.392, 0.648		
Not allowed (*k* = 6, *n* = 344)	0.056 (0.155)	0.363	0.716	-0.247, 0.360		
In Controls					$Q_M(1) = 1.592$	.207
Allowed (*k* = 7, *n* = 693)	0.446 (0.097)	4.584	<.001	0.256, 0.637		
Not allowed (*k* = 16, *n* = 745)	0.157 (0.124)	1.262	0.207	-0.087, 0.400		
Setting					$Q_M(2) = 5.432$	.066
Mixed (*k* = 2, *n* = 308)	0.399 (0.085)	4.698	<.001	0.233, 0.565		
Non-clinical (*k* = 9, *n* = 683)	0.099 (0.123)	0.805	0.421	-0.142, 0.341		
Clinical (*k* = 11, *n* = 489)	0.259 (0.113)	2.288	0.022	0.037, 0.480		

Table 7: Subgroup analysis for SSRT. k: number of studies for which data was available; n: number of participants used for analysis; B: regression coefficients (first group is the intercept, for the other groups the coefficients are contrasts); SE: standard error of regression coefficient; z: Wald-type z-test for the regression coefficient; p: p-value for regression coefficient z-test; ci: confidence interval; Q_M-Test: test for subgroup differences; p_Q: p-value for test for subgroup differences; risk of bias: as assessed by the Hulsbosch Ratings; SST validity: stop-signal task validity; overall quality: risk of bias and SST validity ratings combined; Setting: setting of recruitment for ADHD group.

Figures

Figure 1

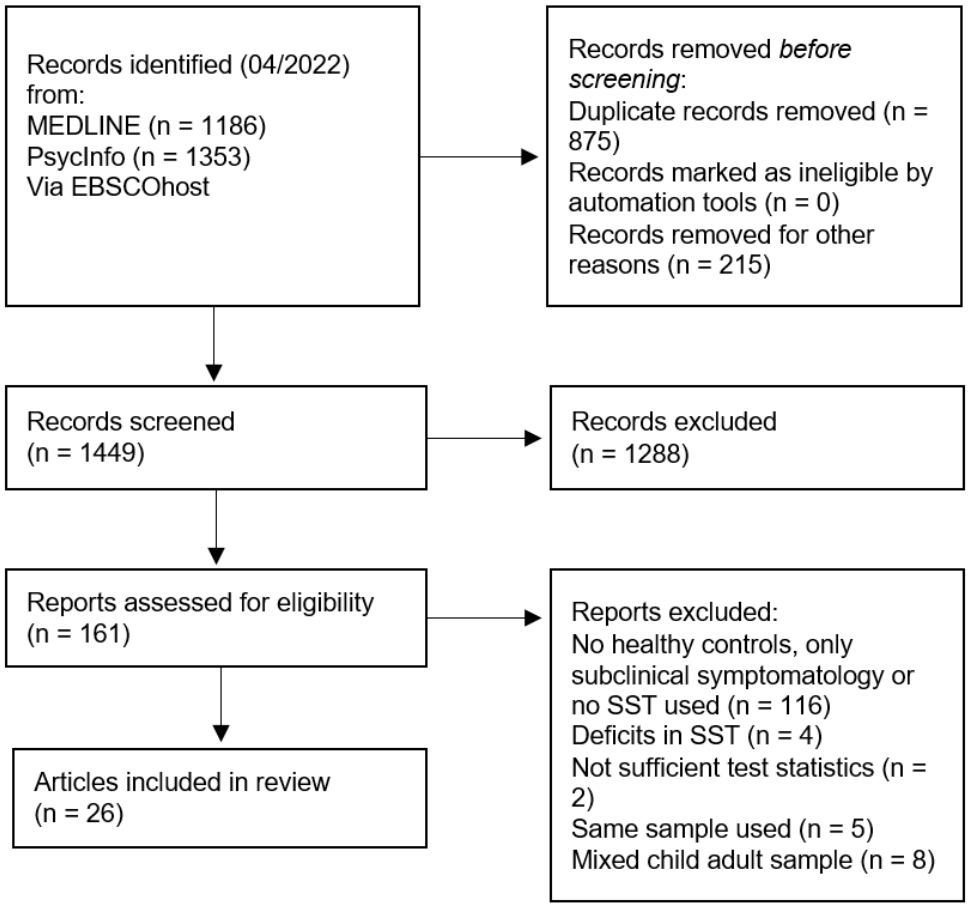


Figure 1: PRISMA chart of study selection (in accordance with Page et al., 2021)

Figure 2

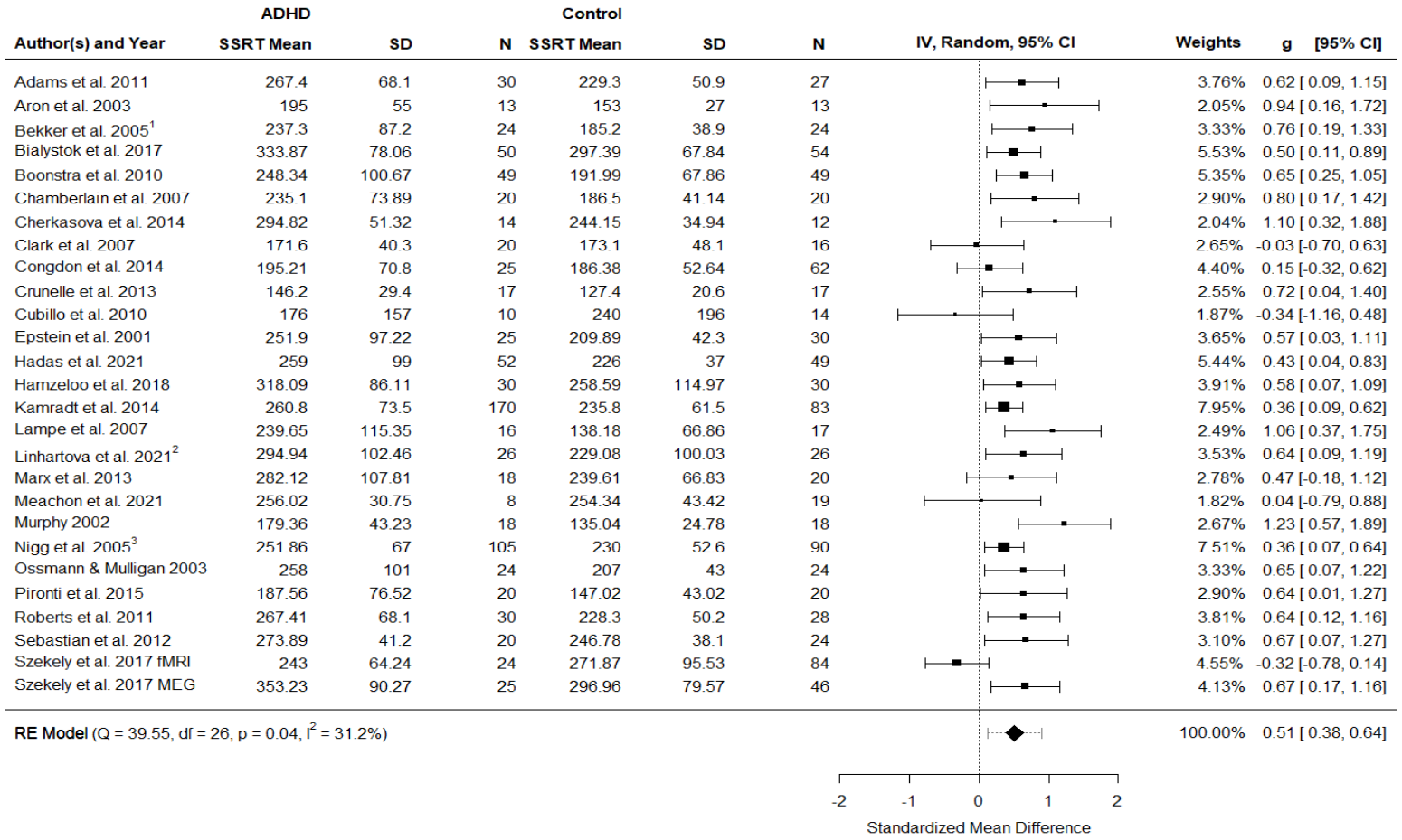


Figure 2: Forest plot showing the observed standardized mean differences (Hedges' g) for SSRT, the estimate of the random-effects model on the right, and the results for the test of heterogeneity on the left. The dashed line at the overall effect estimate (diamond) represents the prediction interval that is shown due to the present heterogeneity. [1]Data was extracted from Bekker et al. (2005a), Bekker et al. (2005b) and van Dongen-Boomsma et al. (2010); [2] Data was extracted from Linhartová et al. (2020) and Linhartová et al. (2021); [3]Data was extracted from Nigg et al. (2005), Stavro et al. (2007) and Martel et al. (2017).

Figure 3

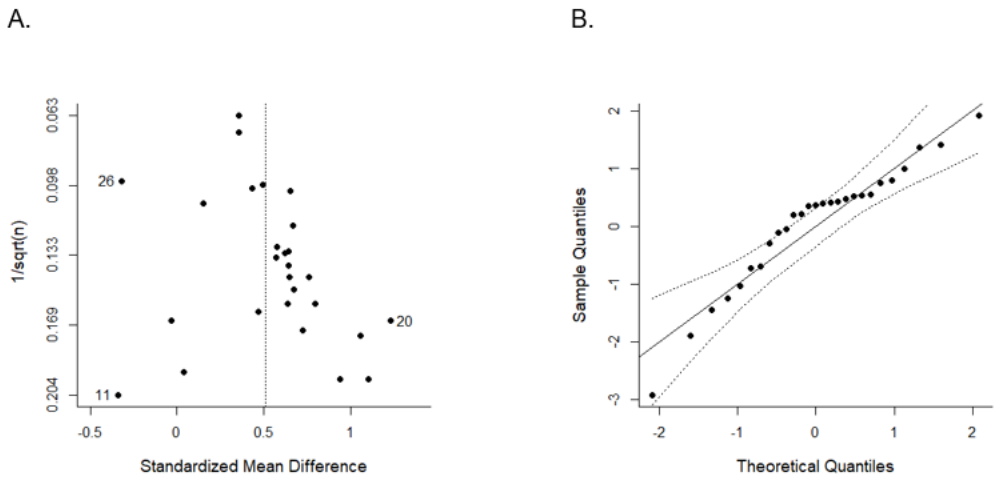


Figure 3: Plots for assessment of publication bias. (A) Funnel plot for SSRT plotting SMDs against the inverse of the square root of the sample size. (B) Normal quantile-quantile plot, plotting the quantiles of a standard normal distribution against the quantiles of the observed distribution. The points should fall on a straight line and inside the 95%-confidence bands.

Supplementary Material

Supplementary text 1

Results of calibration session

In order to increase inter-rater reliability, a calibration session was conducted. This way, the two raters could identify possible sources of disagreement and decide on rules on how to rate ambiguous cases. Moreover, the calibration session revealed a different understanding regarding some of the criteria. Therefore, the following additional rules were applied:

Representative sample. We didn't consider a prior diagnosis or medical records if the diagnosis was not confirmed within the study. A judgment by a consultant psychiatrist which is not further described was not considered an interview.

Sample size. If the sample size of ADHD patients has a different RoB than the sample size of controls, the overall rating for the sample sizes is oriented towards the category with the higher RoB. For example, if the ADHD sample size is n = 20 (therefore rated as having a moderate RoB) and the sample size of controls however is n = 31 (therefore rated as having a low RoB), the overall rating for sample sizes will be moderate RoB.

Analysis reporting. We will not consider whether the method of SSRT estimation was reported, as this aspect is covered by the checklist for SST validity.

Outcome reporting. Mean and SD for SSRT must be reported. Studies who did not report SSRT, but kindly provided us with the data after we contacted them, were also rated as having a low RoB in this category. It had to be apparent that non-significant results were equally reported.

Missing data. If studies do not explicitly report on missing data, however, it is obvious that data from all participants has been used (e.g., if it can be inferred from the degrees of freedom in the analysis), we rated studies as having a low RoB in this domain, even though the original tool requires a clear statement on missing data.

Supplementary Text 2

Assessment of SST validity

This short checklist is based on the consensus guide by Verbruggen et al. (2019) and focuses only on the validity of the assessment and analysis of the SST. Therefore, the four item checklist refers to the recommendations 3 to 9 in the consensus guide and represent these recommendations.

1. Stop trials

Item**:** Did the task include a sufficient number of stop trials and, furthermore, are stop signals only presented on a minority of trials?

- The task includes 50 or more stop trials AND
- Percentage of stop trials is 25% or less (please see also note).

Note: It is also possible to have a higher percentage of stop signals, additional measures to minimize slowing are required (explicitly instruct participants not to wait and include block-based feedback).

Rating:

- Yes (both conditions fulfilled)
- No (at least one condition is not fulfilled)

2. Tracking procedure

Item: Was the stop-signal delay adapted using a tracking procedure (also called staircase)?

- A tracking procedure was implemented with a sufficient step size (usually 50ms steps are used, 16ms steps are too small).

Rating

- Yes (condition fulfilled)
- No (not fulfilled)

3. Estimation Method for SSRT

Item: Was the integration method used to estimate the SSRT?

- The integration method was used to estimate the SSRT.

Note: There are cases that the integration method is only applied when p(respond|signal) is not about 0.50, otherwise the mean or median method is applied. This case also meets the criterion.

Rating:

- Yes (condition fulfilled)
- No (not fulfilled)

4. Check for invalid estimation of SSRT

Item: SSRT should only be estimated, when the assumptions of the horse race model are not violated. There are different methods to check for non-compliant behaviour, e.g., values for p(respond|signal) were lower than 0.25 or higher than 0.75. Did the researchers inspect behaviour for invalid SSRT estimation for each participant for a potential exclusion?

- A rule for inspection for invalid SSRT estimation was applied for each participant.

Note: There are various methods to identify non-compliant participants like reaction time on unsuccessful stop trials should not be numerically longer than reaction time on go trials. This item is not intended for the exclusion of specific trials within the participants (e.g., exclusion of trials of with reactions shorter than 100ms).

Rating:

- Yes (condition fulfilled)
- No (no condition fulfilled)

Overall Rating of SST validity

- High validity: all four items fulfilled
- Moderate validity: three items fulfilled
- Low validity: two or less items fulfilled

Supplementary Text 3

Secondary outcome measures of the SST

Fifteen studies have reported the percentage of stop commissions. Hedges' g of those studies ranged from -0.234 to 0.660, with 73% of estimates indicating that ADHD patients had a higher stop commission percentage *(Supplementary Figure 4*). The two studies that reported stop commissions but did not use a tracking algorithm displayed larger deviations from 50% than the other studies (i.e., Adams et al. 2011: ADHD *M* = 58.2, HC *M* = 62; Marx et al. 2013: ADHD *M* = 64.1, HC *M* = 57.86). The estimated average Hedges' g based on the random-effects model was *g* = 0.142 (95% CI: -0.009 to 0.293), which did not significantly differ from zero ($t(14) = 2.014$, $p = 0.064$). Moreover, there was no significant heterogeneity ($Q(14) = 13.519$, $p = 0.486$, $\hat{\tau}^2 = 0.002$, $I^2 = 2.757\%$) with a 95% prediction interval given by -0.039 to 0.324. In addition, there was no indication of outliers as indicated by the studentized residuals (no values larger than ±2.935) and none of the studies could be considered overly influential according to the Cook's distances. Egger's regression test did not indicate funnel plot asymmetry $t(13) = 2.037$, $p = 0.063$ *(Supplementary Figure 8).*

Only 7 studies reported the percentage of choice errors, a forest plot is shown in *Supplementary Figure 5.* Hedges' g of those studies ranged from -0.467 to 0.541, with 86% of estimates indicating that ADHD patients made more choice errors. The estimated average based on the random-effects model was $g = 0.242$ (95% CI: -0.037 to 0.521), which did not significantly differ from zero ($t(6) = 2.119$, $p = 0.078$). There was no significant heterogeneity ($Q(6) = 7.213$, $p = 0.302$, $\hat{\tau}^2 = 0.002$, $I^2 = 2.853\%$). Clark et al. (2007) had a studentized residual larger than ±2.6901 and may be a potential outlier. Cook's distances, however, revealed no overly influential studies. Leaving the study out leads to an average estimate of *g* = 0.315 (95% CI: 0.125 to 0.505, $t(5) = 4.251$, $p = 0.008$), $\hat{\tau}^2$ and I^2 decreases to 0. There were not enough studies to test for funnel plot asymmetry, as at least ten studies are recommended for reliable results (Sterne et al., 2011).

A forest plot with 9 studies that reported omission errors is shown in *Supplementary Figure*

6. The range of Hedge's g was -0.176 to 0.731, with 78% of estimates hinting that most ADHD patients made more omission errors. The estimated average effect based on the random-effects model was g = 0.418 (95% CI: 0.132 to 0.703), which differed significantly from zero ($t(8)$ = 3.373, p = 0.01). The test for heterogeneity reached significance ($Q(8)$ = 15.780, p = 0.046, $\hat{\tau}^2$ = 0.078, I^2 = 48.506%) and the I^2 statistics indicated moderate heterogeneity in the results. For the true outcomes, the 95% prediction interval was -0.285 to 1.120. Bialystok et al. (2017) had a studentized residual larger than ±2.773 and may be a potential outlier as well as potentially over influential according to Cook's distances. Leaving the study out leads to an average estimate of g = 0.524 (95% CI: 0.286 to 0.762, $t(7)$ = 5.206, p = 0.001), $\hat{\tau}^2$ decreases to 0.002 and I^2 decreases to 1.561. However, there were not enough studies to evaluate funnel plot asymmetry.

Finally, eight of the selected studies provided go accuracy. A forest plot of these studies is shown in *Supplementary Figure 7.* Observed Hedges' g ranged from -0.644 to 0.238, with 88% of estimates indicating that go accuracy was lower for ADHD patients. The estimated average was g = -0.385 (95% CI: -0.635 to -0.136), which significantly differed from zero ($t(7)$ = -3.650, p = 0.008). Even though the test for heterogeneity failed to reach significance ($Q(7)$ = 9.786, p = 0.201, $\hat{\tau}^2$ = 0.031, I^2 = 32.09%), I^2 indicated moderate heterogeneity, reflected by a 95% prediction interval between -0.871 and 0.100. The fMRI observation of Szekely et al. (2017) had a studentized residual larger than ±2.734 and may be a potential outlier as well as potentially overinfluential according to Cook's distances. Leaving out this observation increases g to -0.488 (95% CI: -0.608 to -0.368, $t(6)$ = -9.963, p < 0.0001) and decreases both $\hat{\tau}^2$ and I^2 to 0. Again, funnel plot asymmetry could not be evaluated due to the small number of studies.

Supplementary Tables

Supplementary Table 1

	Unadjusted kappa	Adjusted kappa
Item 1	0.920	0.923
Item 2	1	1
Item 3	0.752	0.769
Item 4	0.785	0.846

Supplementary Table 1: Inter-rater reliability for stop signal task validity ratings. Item 1: ≥50 stop trials in total, stop trials constituting ≤25% of all trials; Item 2: staircase algorithm implemented; Item 3: integration method used; Item 4: cut-offs applied to ensure valid SSRT estimation.

Supplementary Table 2

	Weighted kappa
Equivalent groups	0.743
Representative sample	0.792
Sample sizes	0.861
Selective outcome reporting	1
Analysis reporting	1
Missing data	0.667

Supplementary Table 2: Inter-rater reliability for RoB ratings. Domains in accordance with the adapted Hombrados and Waddington criteria (Hulsbosch et al., 2021).

Supplementary Table 3

Moderator	*B (SE)*	*t*	*p*	*ci*	*F*-Test	p_F
Age, Sex (*k* = 21, *n* = 1465)					F(3,17) = 0.885	.469
Intercept	0.477 (0.054)	8.875	<.001	0.364, 0.591		
Age	0.122 (0.076)	1.604	.127	-0.039, 0.284		
Sex	-0.077 (0.083)	-0.922	.370	-0.253, 0.099		
Age:Sex	0.146 (0.126)	1.161	.262	-0.120, 0.412		
IQ (*k* = 14, *n* = 774)					F(1,12) = 0.098	.759
Intercept	0.564 (0.088)	6.438	<.001	0.373, 0.755		
IQ	0.005 (0.016)	0.313	.759	-0.030, 0.040		

Supplementary Table 3: Meta-regression analyses for SSRT. k: number of studies for which data was available; n: number of participants used for analysis; B: unstandardized regression coefficient. For categorical variables, B is the average estimated effect size for each individual factor level; SE: standard error of regression coefficient; t: t-test for the regression coefficient; p: p-value for regression coefficient t-test; ci: confidence interval; F-Test: test of moderator; p_F: p-value for test of moderator; Sex: percentage of males in the individual study samples; IQ: for ADHD and control group combined; Setting: patient setting of ADHD group.

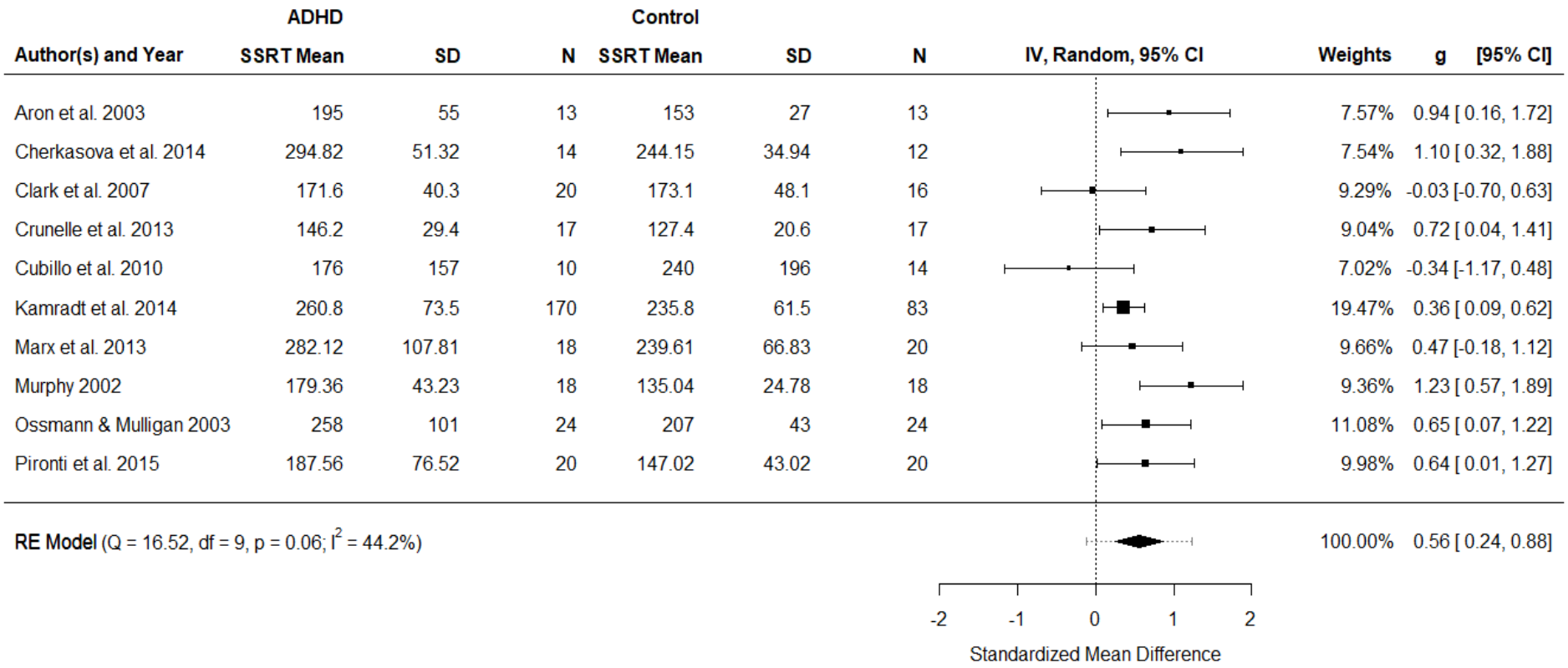
Supplementary Table 4

Moderator	*B (SE)*	*z*	*p*	*ci*	Q_M-*Test*	p_Q
Comorbidities						
In Patients					Q_M (1) = 0.679	.410
Allowed (*k* = 17, *n* = 1168)	0.532 (0.066)	8.071	<.001	0.403, 0.662		
Not allowed (*k* = 4, *n* = *285*)	-0.104 (0.127)	-0.824	.410	-0.353, 0.144		
In Controls					Q_M (1) = 1.167	.280
Allowed (*k* = 7, *n* = 693)	0.446 (0.097)	4.584	<.001	0.256, 0.637		
Not allowed (*k* = 13, *n* = 659)	0.133 (0.123)	1.080	.280	-0.108, 0.373		
Setting					Q_M (2) = 4.287	.117
Mixed (*k* = 2, *n* = 308)	0.399 (0.085)	4.698	<.001	0.233, 0.565		
Non-clinical (*k* = 8, *n* = 579)	0.105 (0.136)	0.771	.441	-0.162, 0.371		
Clinical (*k* = 10, *n* = 456)	0.230 (0.112)	2.057	.040	0.011, 0.448		

Supplementary Table 4: Subgroup analysis for SSRT. k: number of studies for which data was available; n: number of participants used for analysis; B: regression coefficients (first group is the intercept, for the other groups the coefficients are contrasts); SE: standard error of regression coefficient; Wald-type z-test for the regression coefficient; p: p-value for regression coefficient z-test; ci: confidence interval; Q_M-Test: test for subgroup differences; p_Q: p-value for test for subgroup differences; risk of bias: as assessed by the Hulsbosch Ratings; SST validity: stop-signal task validity; overall quality: risk of bias and SST validity ratings combined; Setting: setting of recruitment for ADHD group.

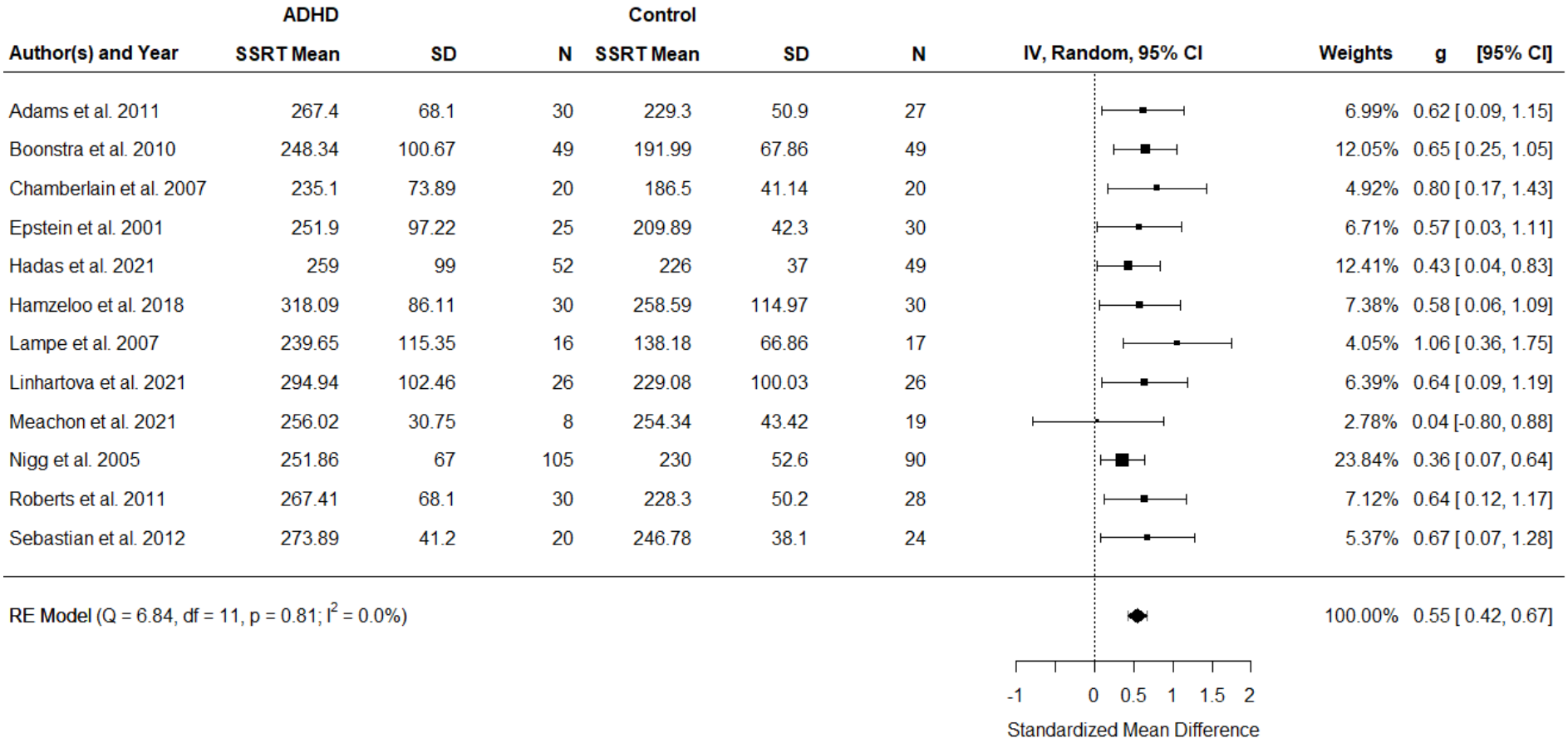
Supplementary Figures

Supplementary Figure 1



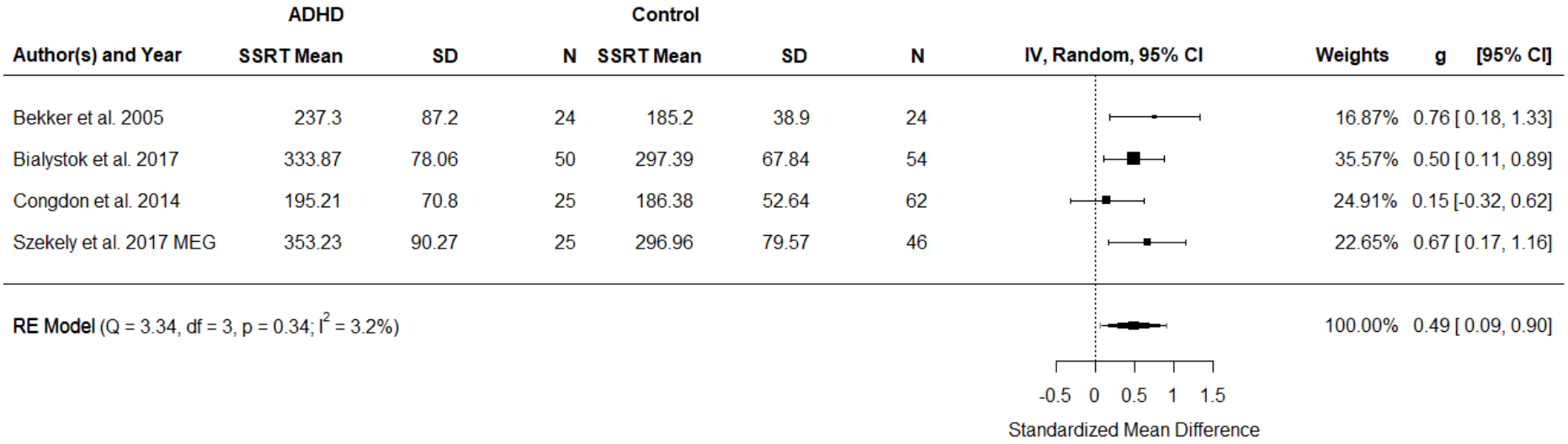
Supplementary Figure 1: SSRT for studies with low quality**.** Forest plot showing the observed standardized mean differences (Hedges' g) for SSRT, the estimates of the random-effects model, and the results for the test of heterogeneity for studies designated as having low quality.

Supplementary Figure 2



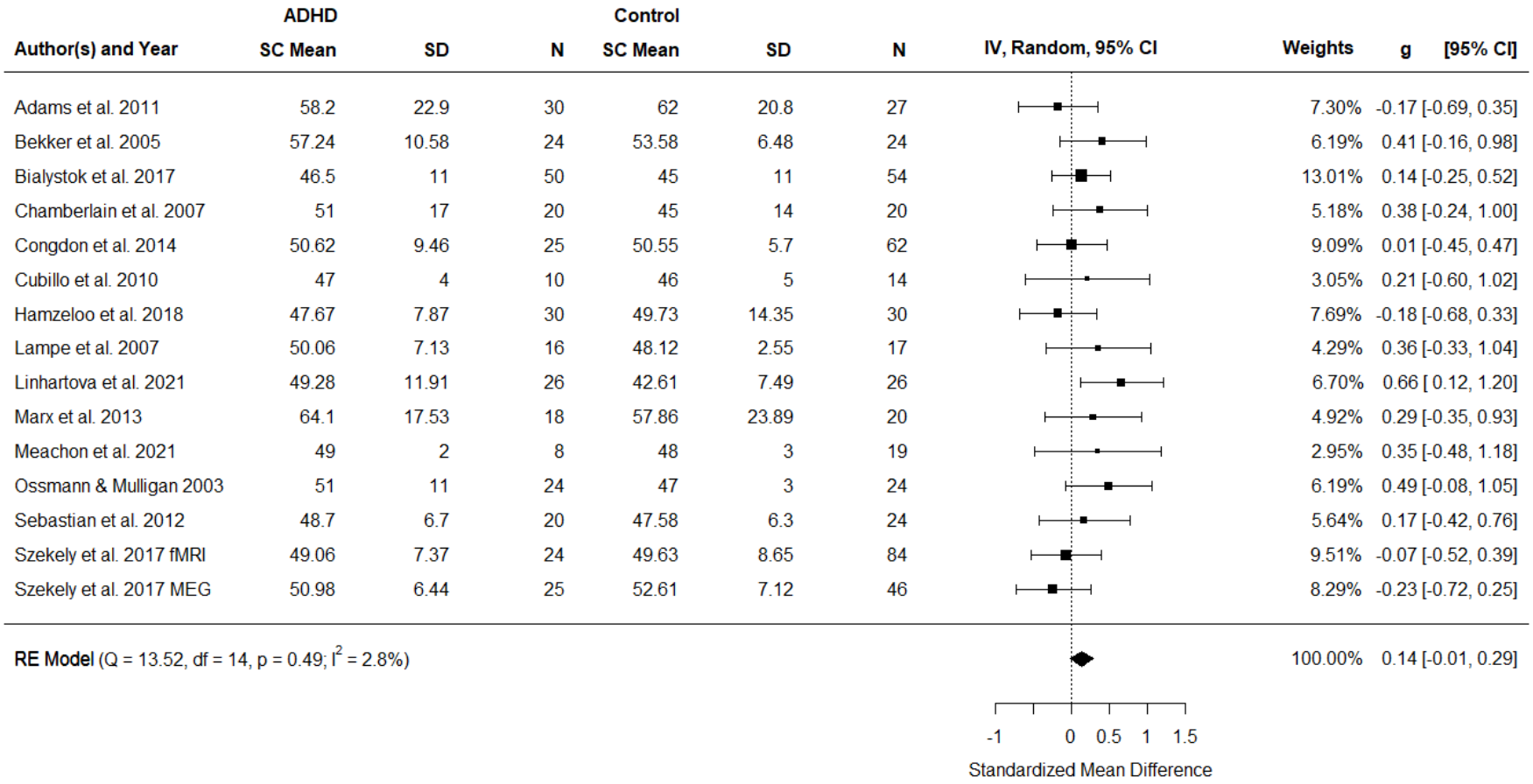
Supplementary Figure 2: SSRT for studies with moderate quality. Forest plot showing the observed standardized mean differences (Hedges' g) for SSRT, the estimates of the random-effects model, and the results for the test of heterogeneity for studies designated as having low to moderate quality.

Supplementary Figure 3



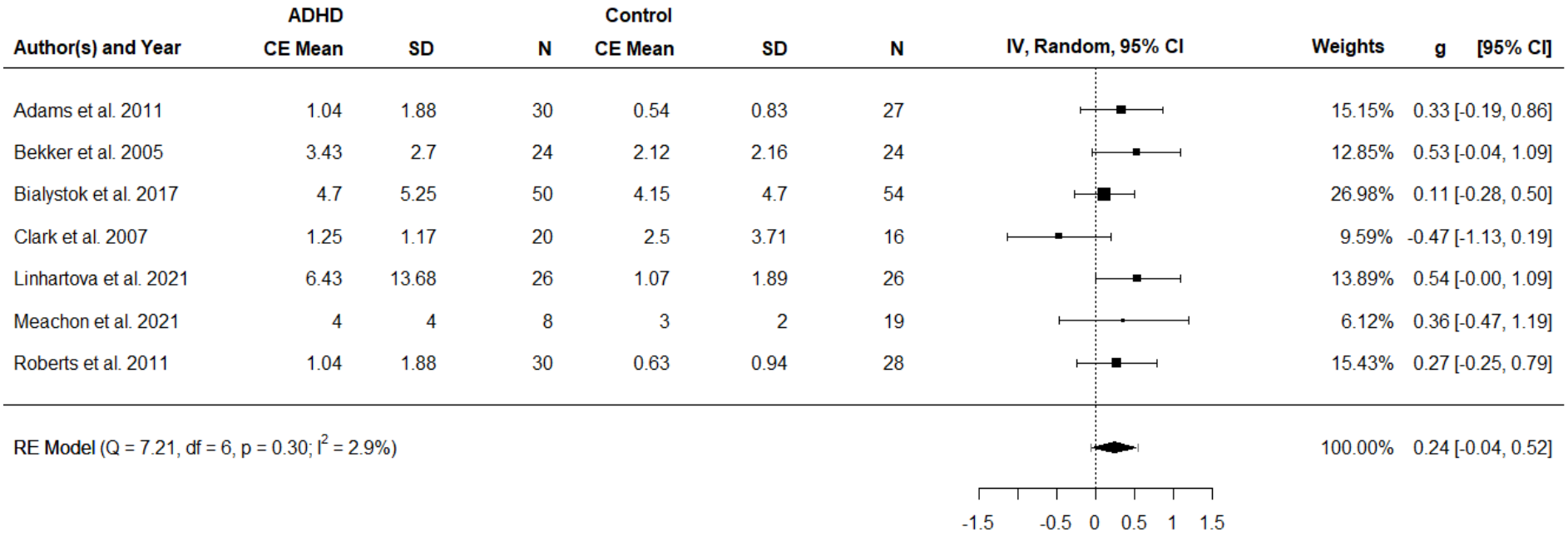
Supplementary Figure 3: SSRT for studies with high quality**.** Forest plot showing the observed standardized mean differences (Hedges' g) for SSRT, the estimates of the random-effects model and the results for the test of heterogeneity for studies designated as having moderate to high quality.

Supplementary Figure 4



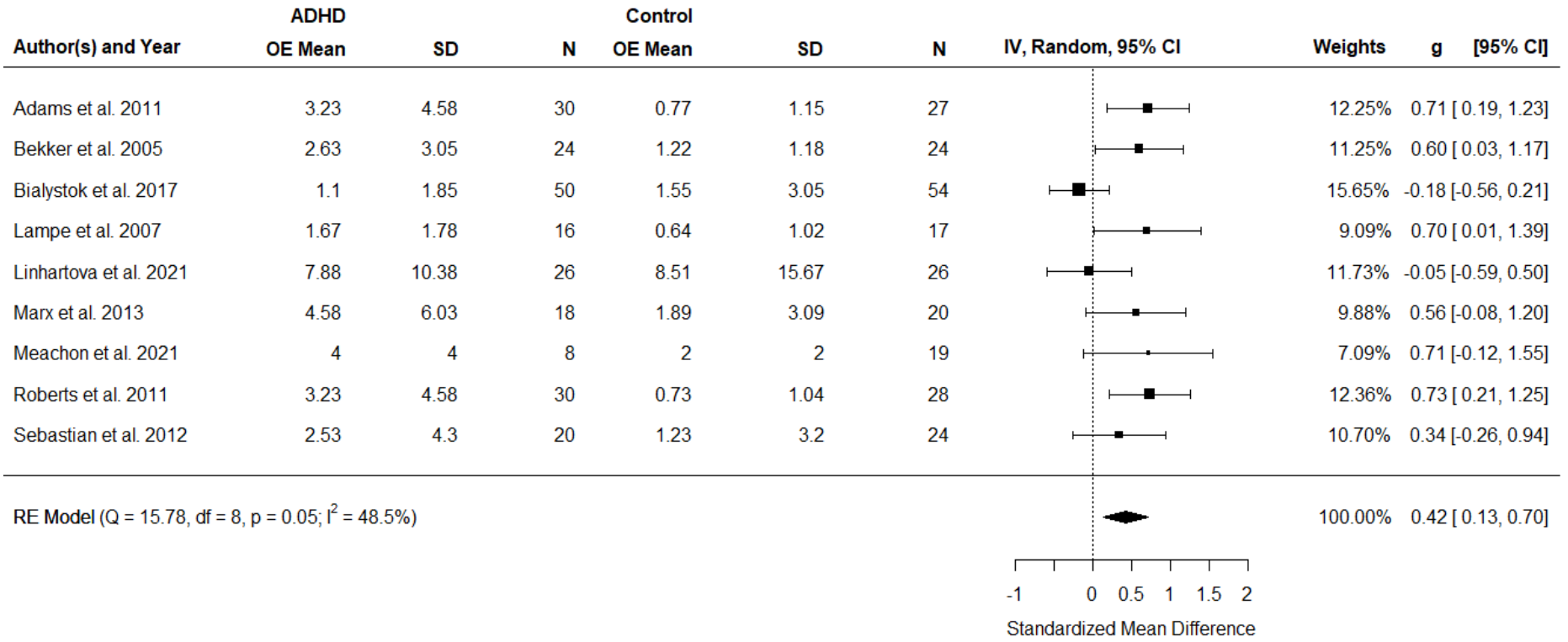
Supplementary Figure 4: Forest plot showing the observed standardized mean differences (Hedges' g) for stop commissions (SC, %), the estimate of the random-effects model and the results for the test of heterogeneity.

Supplementary Figure 5



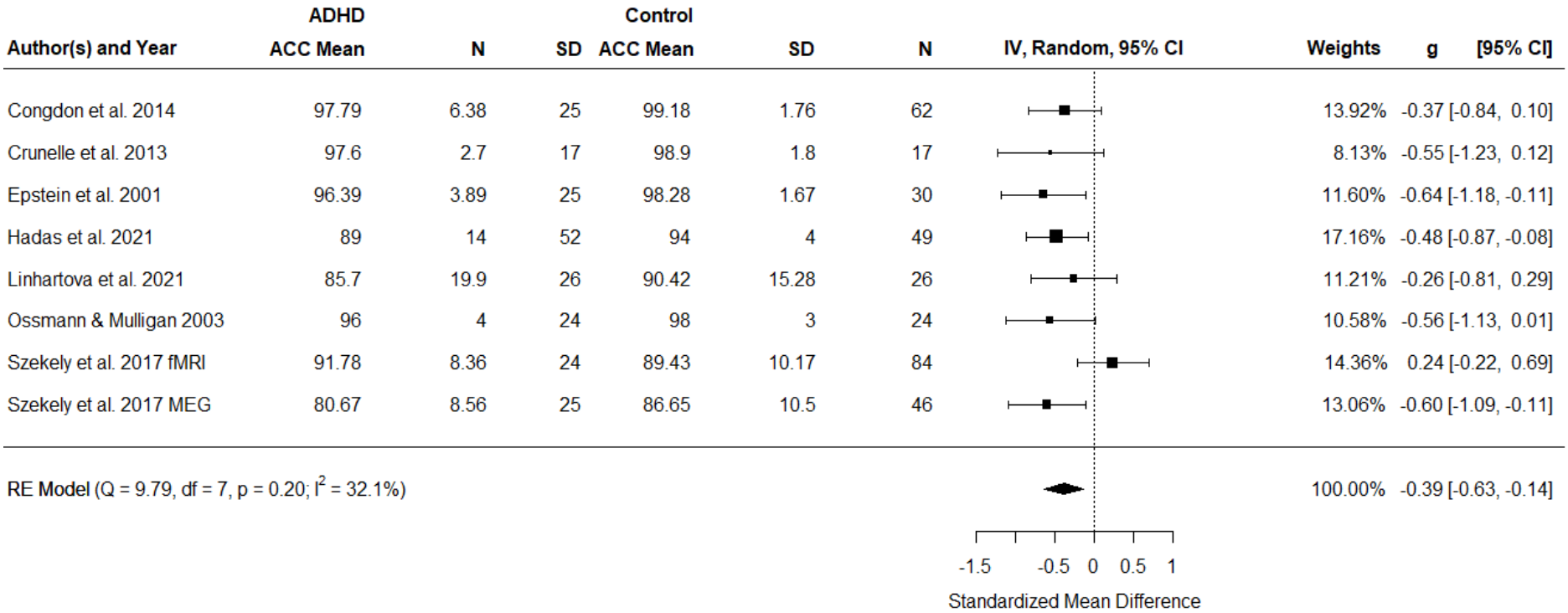
Supplementary Figure 5: Forest plot showing the observed standardized mean differences (Hedges' g) for choice errors (CE) in go trials (%), the estimate of the random-effects model and the results for the test of heterogeneity.

Supplementary Figure 6



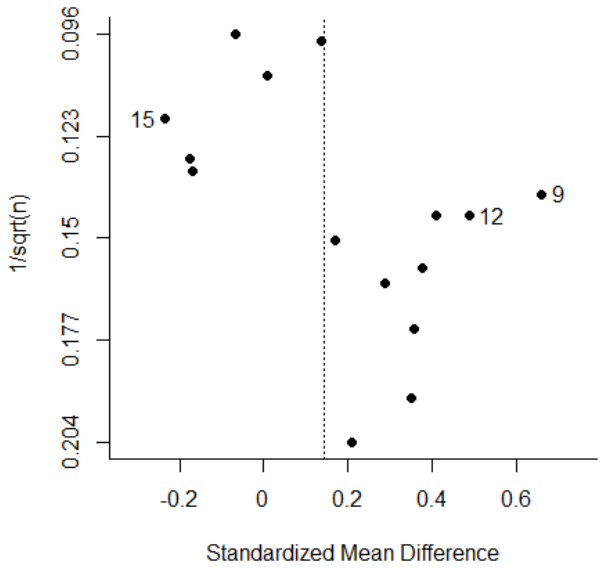
Supplementary Figure 6: Forest plot showing the observed standardized mean differences (Hedges' g) for omission errors (OE) in go trials (%), the estimate of the random-effects model and the results for the test of heterogeneity.

Supplementary Figure 7



Supplementary Figure 7: Forest plot showing the observed standardized mean differences (Hedges' g) for accuracy (ACC) in go trials (%), the estimate of the random-effects model, and the results for the test of heterogeneity.

Supplementary Figure 8



Supplementary Figure 8: Funnel plot for stop commissions plotting SMDs against the inverse of the square root of the sample size. [9]Linhartova et al. (2021); [12]Ossman & Mulligan (2003); [15]Szekely et al. (2017) MEG.