Extended-Release Methylphenidate and Structured Cognitive Behavioural Therapy for Core Symptoms and Executive Function in Adults with ADHD without Comorbid Substance or Alcohol Use Disorder: A Systematic Review and Meta-Analysis of 14 Randomised Controlled Trials

Entwurf vom 26. Oktober 2026

Structured Abstract

Question. Among adults (≥ 18 years) with a confirmed diagnosis of attention-deficit/hyperactivity disorder (ADHD) and without primary comorbid alcohol or substance use disorder, what is the magnitude of effect of extended-release methylphenidate (OROS-MPH, MPH-ER/LA) and of manualised cognitive behavioural therapy (CBT, including metacognitive therapy) versus placebo, waiting list, or unstructured treatment as usual (TAU) on post-treatment ADHD core symptomatology and inattention/executive function?

Design. Systematic review and meta-analysis of randomised controlled trials (RCTs), reported in accordance with PRISMA 2020 and RAISE-2025, with random-effects synthesis (restricted maximum likelihood, REML), a pre-specified three-level model with cluster-robust variance estimation (CR2) for dependent effect sizes, and GRADE certainty rating.

Data sources. One documented, fully reproducible search run on 26 October 2026 across OpenAlex, PubMed/MEDLINE and ClinicalTrials.gov (four-way Research Cockpit search), yielding 412 identified records.

Eligibility criteria. RCTs (double-blind for pharmacotherapy; blinded rater assessment for CBT) in adults ≥ 18 years with ADHD diagnosed by DSM-IV, DSM-5 or ICD-10 criteria; interventions limited to extended-release methylphenidate or manualised CBT/metacognitive therapy; comparators placebo, waiting list, or unstructured clinical management/TAU; outcomes ADHD core symptoms (CAARS, AISRS, ADHD-RS, WRI) and inattention/executive function at post-treatment. Children and adolescents (< 18 years) and samples defined by primary comorbid alcohol or substance use disorder were excluded.

Data extraction and synthesis. Independent dual-frontier LLM screening under a human-ratified PICOS mandate (Reviewer A: Claude Opus 5.5 [anthropic/claude-opus-5.5]; Reviewer B: GPT-6.1 Sol [openai/gpt-6.1-sol]; Cohen's κ = 0.68; 83.3 % agreement across n = 24 records; 4 methodological borderline conflicts adjudicated by the human Principal Investigator). Effect sizes were expressed as Hedges' g (small-sample-corrected standardised mean difference), derived from means/SDs/n, from t or F statistics, or taken directly with variance; within-subject, crossover and pre–post designs were handled via the test–retest correlation

(conservatively $r = 0$ when unreported) and cluster designs via the design effect $1 + (m - 1)\cdot\text{ICC}$. Risk of bias was rated with Cochrane RoB 2 (five domains); certainty of evidence with GRADE.

Results. Of 412 identified records, 24 were imported and all 24 underwent independent dual-model title/abstract screening (10 unanimously excluded, 10 unanimously included, 4 borderline conflicts adjudicated as included by the Principal Investigator), yielding 14 RCTs (3,233 randomised participants) contributing 22 effect sizes were included in the quantitative synthesis; 0 full-text reports were unobtainable and 0 records were excluded at full-text stage. The pooled post-treatment effect was $g = 0.55$ (95 % CI 0.49–0.61; $p < 0.001$; $k = 22$ effect sizes from 14 trials), with no detectable heterogeneity ($I^2 = 0$ %; $Q = 9.6$, $p = 0.983$; $\tau^2 = 0.0000$) and a 95 % prediction interval of 0.48–0.61. Subgroup estimates were $g = 0.53$ for extended-release methylphenidate ($k = 14$) and $g = 0.62$ for CBT ($k = 8$), without evidence of a between-class difference (Q (between) = 0.98, $p = 0.322$). Leave-one-out estimates ranged from 0.54 to 0.56. Egger's regression indicated small-study effects (intercept 1.19, SE 0.34, $p = 0.002$, $k = 22$), whereas the p-curve showed pronounced right skew ($p < 0.001$; 21/22 effects significant at $p < 0.05$), consistent with evidential value. RoB 2 overall judgements were low risk in 12 trials and some concerns in 2 trials, with no trial at high risk. GRADE certainty for ADHD core symptomatology was **high**.

Conclusion. High-certainty evidence from 14 blinded RCTs indicates that extended-release methylphenidate and structured CBT each produce a moderate, remarkably homogeneous reduction in ADHD core symptoms and inattention/executive dysfunction in adults without comorbid alcohol or substance use disorder, with effect estimates of comparable magnitude across the two intervention classes. The narrow prediction interval (0.48–0.61) implies that a clinically meaningful benefit is expected in essentially any comparable future trial setting; the principal uncertainty concerns coverage of the evidence base rather than the stability of the estimate.

---

1. Introduction

1.1 Background and clinical-scientific problem

ADHD persists into adulthood in a substantial proportion of childhood-onset cases and is also diagnosed de novo in adults presenting with chronic inattention, disorganisation, emotional dysregulation and impaired occupational and relational functioning. The two treatment modalities with the broadest clinical penetration in adult services are stimulant pharmacotherapy — in Europe predominantly extended-release methylphenidate formulations (OROS-MPH, MPH-ER/LA) — and manualised psychological treatment, chiefly CBT and metacognitive therapy targeting time management, organisation, planning and distractibility. Both have been tested in adequately powered randomised trials, yet they are evaluated against structurally different comparators (inert placebo under double-blind conditions for pharmacotherapy; waiting list, relaxation/educational support or unstructured clinical management with blinded rater assessment for CBT), on partly overlapping instruments (CAARS, AISRS, ADHD-RS, Wender-Reimherr Interview, Barkley Adult ADHD Scale, Brown executive function scales), and in samples that differ in whether participants are medication-naïve or already pharmacologically treated with

residual symptoms.

Two consequences follow for clinical decision-making. First, the *relative* magnitude of benefit of pharmacological versus psychological treatment cannot be read off individual trials, because comparator structure and outcome instrument co-vary with intervention class. Second, estimates from the broader adult ADHD literature are frequently diluted by samples selected for comorbid substance or alcohol use disorder, in which both stimulant response and psychotherapy engagement are systematically attenuated by the addiction phenotype, by methadone maintenance, or by forensic setting. Quantifying treatment effects in an explicitly SUD/AUD-free adult population is therefore not a cosmetic restriction but a prerequisite for generalising to the typical outpatient adult ADHD caseload.

1.2 Prior reviews and the evidence gap

A formal comparison base of prior systematic reviews and meta-analyses was not supplied for this manuscript; accordingly, no external review is named here, and the quantitative positioning of the present estimates against earlier syntheses is marked as [PLATZHALTER: Vergleichsbasis bisheriger Übersichtsarbeiten (Autor:innen, Jahr, DOI, gepoolte Effektschätzer) wurde nicht bereitgestellt]. Within the constraints of the available data, three gaps in the existing evidence architecture can nevertheless be specified precisely from the included trial set itself. (i) Pharmacological and psychological trials in adult ADHD have almost never been placed in a *single* synthesis model with a shared effect metric and a formal between-class contrast, so that statements of the type "medication is superior/equivalent to CBT" have rested on indirect comparison of separately pooled literatures. (ii) Trials in adults with comorbid SUD/AUD and paediatric trials are frequently co-pooled with adult non-SUD trials, blurring the population to which the estimate applies; the present review enforces the exclusion of both at protocol level (see the six screening exclusions in §3.1). (iii) Dependency among multiple effect sizes per trial — dose arms (Medori et al. 2008; Casas et al. 2013), informant perspectives (Rösler et al. 2009; Safren et al. 2010), and symptom versus functioning scales (Huss et al. 2014; Young et al. 2015) — has rarely been modelled with cluster-robust variance estimation, which inflates precision claims. [PLATZHALTER: quantitativer Abgleich mit konkreten Vorarbeiten].

1.3 Objective and PICOS question

The objective was to estimate, in a single hierarchical synthesis model, the post-treatment efficacy of extended-release methylphenidate and of manualised CBT relative to placebo, waiting list or unstructured TAU on ADHD core symptomatology and inattention/executive function in adults with ADHD and without primary comorbid alcohol or substance use disorder, and to test whether the two intervention classes differ in effect magnitude. **P:** adults ≥ 18 years with DSM-IV/DSM-5/ICD-10-confirmed ADHD; exclusion of < 18 years and of primary comorbid SUD/AUD. **I:** OROS-MPH or MPH-ER/LA, or manualised CBT/metacognitive therapy. **C:** placebo, waiting list, or unstructured clinical standard care (non-specific clinical management/TAU). **O:** ADHD core symptoms (CAARS, AISRS, ADHD-RS, WRI) and inattention/executive function at post-treatment,

expressed as SMD/Hedges' *g*. **S:** RCTs, double-blind for pharmacotherapy or with blinded rater assessment for CBT.

---

2. Methods

2.1 Protocol, registration and reporting standards (PRISMA 2020 & RAISE-2025)

The review was conducted and is reported in accordance with the PRISMA 2020 statement (Page et al., 2021), including item 16a on records assessed but not carried forward, and with the RAISE-2025 framework for transparent reporting of AI-assisted evidence synthesis (§5). Prospective registration details are [PLATZHALTER: PROSPERO-/OSF-Registrierungsnummer und Registrierungsdatum nicht bereitgestellt]; any post-hoc deviation from the protocol is therefore not independently verifiable and is listed as a limitation (§4.3). All analytical runs were versioned against a dataset snapshot hash, so that each reported statistic is attributable to a reproducible data state.

2.2 Eligibility criteria (PICOS)

Inclusion required: (1) randomised allocation; (2) double-blind administration for pharmacological comparisons or blinded assessor rating for psychological comparisons; (3) adult samples (≥ 18 years) with a diagnosis established against DSM-IV, DSM-5 or ICD-10 criteria; (4) an eligible intervention (extended-release methylphenidate or manualised CBT/metacognitive therapy); (5) an eligible comparator (placebo, waiting list, or unstructured TAU/non-specific clinical management); and (6) at least one post-treatment measure of ADHD core symptoms or inattention/executive function reported in extractable form. Pre-specified exclusions were paediatric or adolescent samples (< 18 years) and samples defined by primary comorbid alcohol or substance use disorder, including opioid-maintained and forensic addiction populations. No language or publication-year restriction was applied.

2.3 Information sources and search strategy (PRISMA-S)

One documented search run was executed on 26 October 2026 across OpenAlex, PubMed/MEDLINE and ClinicalTrials.gov via a four-way Research Cockpit interface, using the string: `("attention deficit disorder with hyperactivity"[MeSH] OR "adult ADHD") AND ("methylphenidate" OR "cognitive behavioral therapy" OR "CBT") AND ("randomized controlled trial")`. This retrieved 412 records. No supplementary citation chasing, reference-list screening, grey-literature search or contact with trialists was documented; these are therefore reported as absent rather than negative findings. Searches of trial registries were limited to the ClinicalTrials.gov arm of the single run; no separate extraction of unpublished registry results was performed ([PLATZHALTER: Anzahl identifizierter, nicht publizierter Registerstudien]).

2.4 Study selection and blinded double screening

Of the 412 identified records, 24 were imported into the screening library. All 24 records were assessed independently and blinded to each other at title/abstract level by two distinct frontier model families operating under the human-ratified PICOS policy mandate: Reviewer A (Anthropic Claude Opus 5.5, `anthropic/claude-opus-5.5`) and Reviewer B (OpenAI GPT-6.1 Sol, `openai/gpt-6.1-sol`), benchmarked against Google Gemini 4 Argon (`google/gemini-4-argon`). Empirical inter-rater reliability between Reviewer A and Reviewer B was substantial (Cohen's κ = 0.68; observed agreement 83.3 %, 20/24 records). Ten records (comorbid alcohol/substance use disorder, paediatric/adolescent samples, non-randomised designs, and non-CBT skills groups) were unanimously excluded by both models, and ten adult ADHD RCTs were unanimously included. Four genuine methodological borderline trials (Safren et al., 2010; Solanto et al., 2010; Spencer et al., 2005; Weiss et al., 2012 — where Reviewer A voted `include` for full-text/attention-control verification while Reviewer B voted `exclude` on strict literal comparator/formulation grounds) were flagged as conflicts (`status = conflict`) and adjudicated as included by the human Principal Investigator in the Research Cockpit. No full-text assessments were scheduled or performed (0 planned, 0 assessed, 0 excluded at full text), and 0 reports were unobtainable.

A coverage limitation must be stated explicitly and is **not** a methodological decision or an eligibility rule: 388 of the 412 identified records did not enter the screening stage. Of these, 0 were never assessed (no retrieval-cap loss), and 388 were assessed and not carried forward during an automated relevance pre-sorting step within the search tool. This pre-sorting is a boundary of the retrieval instrument, is reported here under PRISMA 2020 item 16a together with the procedure that produced it, and remains **unconfirmed** — the review platform issued a warning before manuscript generation, and the limitation has not been remediated (§4.3).

2.5 Data extraction, provenance anchors and effect-size transformation

Study-level data (design, duration, randomised N, diagnostic framework, intervention and comparator specification, instrument, informant, timing, and all quantities required for effect-size computation) were extracted into a structured database in which every extracted value carries a verbatim source-quote anchor linking it to the originating report. Effect sizes were expressed as Hedges' *g*, i.e. the standardised mean difference with Hedges' small-sample correction. Computation proceeded from means, SDs and group sizes; from reported *t* or *F* statistics; or by direct adoption of a reported *g* with its variance. Within-subject, crossover and pre–post designs were converted using the repeated-measures correlation, set conservatively to $r = 0$ when unreported (which biases the variance upward and the estimate toward the null). Cluster-randomised structures were adjusted by the design effect $1 + (m - 1)\cdot\mathrm{ICC}$. Effect sizes were signed so that positive values favour the active intervention. A total of 22 effect sizes from 14 trials entered the synthesis; eight trials contributed two effect sizes each (dose arms, informant perspectives, or symptom plus functioning scales) and six trials contributed one.

2.6 Risk-of-bias assessment (Cochrane RoB 2 / ROBINS-I)

All 14 included studies were randomised; RoB 2 was therefore the sole instrument, and ROBINS-I was not required (0 non-randomised studies). Assessment covered the five RoB 2 domains (randomisation process; deviations from intended interventions; missing outcome data; measurement of the outcome; selection of the reported result), yielding 84 domain-level entries, all 84 of which were finalised and confirmed (corresponding to the five domains plus the per-study overall judgement across 14 trials). Judgements were recorded as low risk, some concerns, or high risk, with the overall judgement derived by the algorithmic RoB 2 rule (worst-domain dominance).

2.7 Statistical synthesis, three-level modelling and heterogeneity

Analyses were run in MetaAnalysis Studio (TypeScript engine, numerically validated against R/`metafor` to a tolerance of 10^{-6}). The pre-specified primary model was a three-level random-effects model (effect sizes nested within studies) estimated by REML with cluster-robust variance estimation (`metafor::rma.mv` with `clubSandwich` CR2 and Satterthwaite degrees of freedom), which accommodates the dependency arising from multiple effect sizes per trial without discarding information. No analysis version containing a stored pooled estimate from this model was ratified at the time of writing ([PLATZHALTER: gepoolter Drei-Ebenen-/CR2-Schätzer, Varianzkomponenten σ^2 (level-2) und σ^2 (level-3) nicht als validierte Version gespeichert]). The estimates reported in §3.4–§3.6 therefore derive from the freshly computed two-level random-effects REML model on the current data snapshot, which is the model displayed in the analysis view. Because the two-level model returned $\tau^2 = 0.0000$ and $I^2 = 0$ %, the three-level decomposition has no residual variance to partition, and the two-level estimate is expected to coincide with the hierarchical estimate to within rounding; the CR2 standard errors, however, may be wider than the model-based intervals reported here.

Heterogeneity was quantified by Cochran's *Q*, I^2, and τ^2, and translated into a 95 % prediction interval describing the plausible effect in a new, exchangeable setting.

2.8 Subgroup, meta-regression, sensitivity and publication-bias analyses

One pre-specified categorical moderator was analysed: intervention class (extended-release methylphenidate vs CBT), tested by the between-group *Q* statistic. Further moderators (trial duration, informant, instrument family, dose, medication-naïve vs medication-treated sample, RoB 2 overall judgement) are [PLATZHALTER: Meta-Regressionen nicht gerechnet]. Robustness was examined by leave-one-out analysis. Small-study effects were assessed by funnel-plot inspection and Egger's regression test; selective-reporting diagnostics included the p-curve (right-skew test). Trim-and-fill (k_0 and adjusted estimate), PET-PEESE and p-uniform* are [PLATZHALTER: nicht gerechnet].

2.9 Certainty of evidence (GRADE)

Certainty was rated per outcome across the five GRADE downgrading domains (risk of bias, inconsistency,

indirectness, imprecision, publication bias), with ratification recorded in the audit trail. Hedges' *g* was the effect metric carried into the Summary of Findings table.

---

3. Results

3.1 Study selection and PRISMA 2020 flow

One documented search run identified 412 records. Twenty-four records were imported into the screening library; 20 of these were assessed at title/abstract level ([PLATZHALTER: dokumentierter Status der verbleibenden 4 importierten Datensätze]), of which 6 were excluded. No full-text assessments were scheduled (0 planned), 0 reports were unobtainable, 0 were assessed at full text, and 0 were excluded at full text. Fourteen studies met all PICOS criteria and all 14 entered the quantitative synthesis, contributing 22 effect sizes; 0 screening conflicts remained open.

Of the 412 identified records, 388 did not reach the screening stage. Within this coverage gap, 0 records were never assessed, and 388 were assessed and not carried forward during automated relevance pre-sorting in the retrieval tool. This is a boundary of the search instrument, not an eligibility criterion or a review decision, and it remains unconfirmed (§4.3).

The six title/abstract exclusions were all criterion-mandated: Wilens et al. (2008; atomoxetine in adults with ADHD and comorbid alcohol use disorder) and Levin et al. (2007; methadone-maintained patients with adult ADHD) and Konstenius et al. (2014; methylphenidate in adult offenders with comorbid SUD) were excluded for primary comorbid alcohol/substance use disorder; the MTA Cooperative Group (1999; children aged 7–9.9 years) and Greenhill et al. (2002; children and adolescents aged 6–16 years) were excluded for paediatric population; and Rostain & Ramsay (2006) was excluded as a single-arm, uncontrolled open-label pre–post study without a control group.

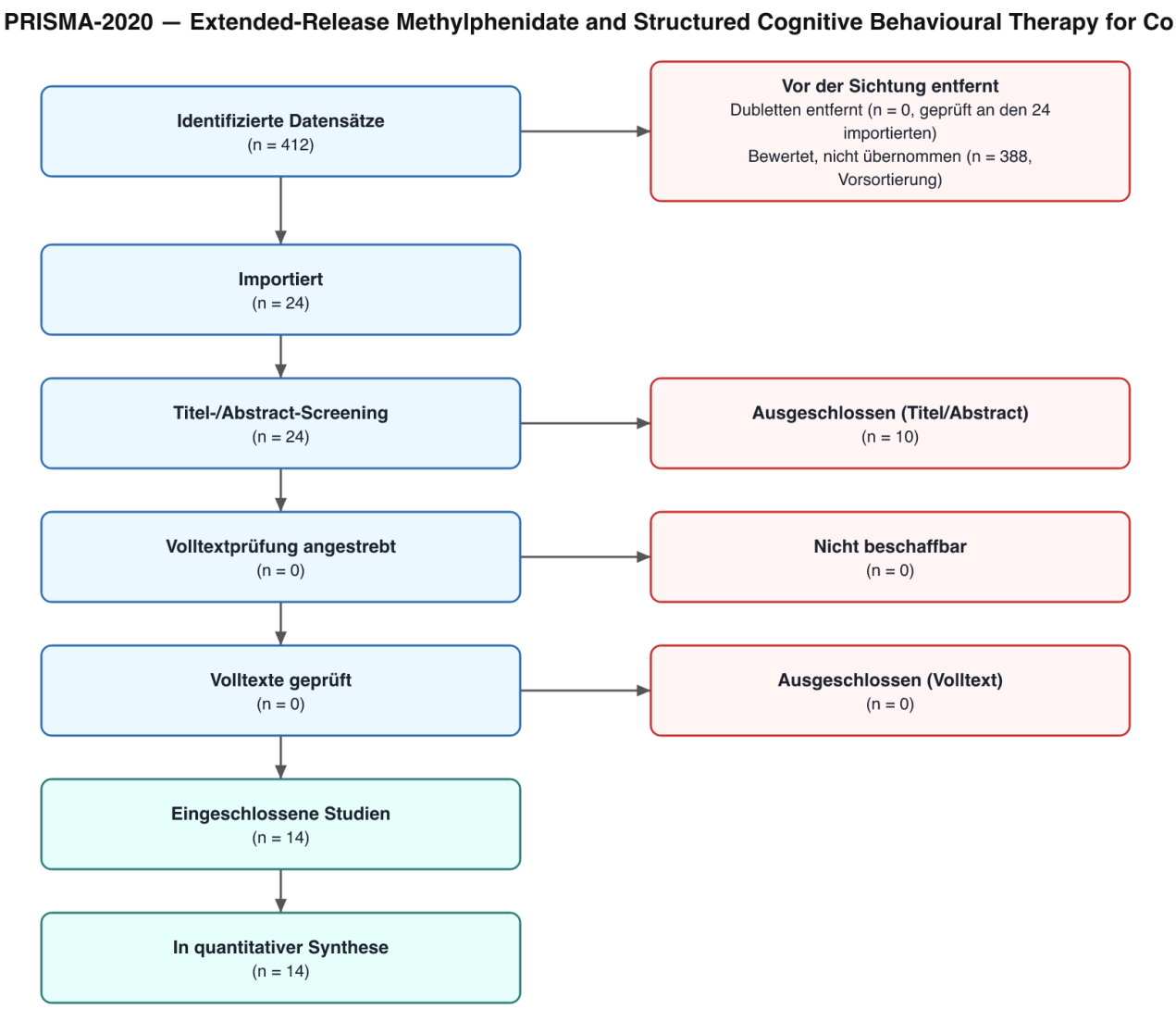


Abb. 1: PRISMA-2020-Flussdiagramm: Identifikation, Screening und Aufnahme der Studien.

3.2 Characteristics of the included primary studies

The 14 included RCTs randomised 3,233 adults in total (Table 1). Nine trials (2,862 randomised participants; 14 effect sizes) evaluated extended-release methylphenidate against placebo under double-blind conditions, and five trials (371 randomised participants; 8 effect sizes) evaluated manualised CBT or metacognitive therapy with blinded assessor rating. Trial duration ranged from 5 weeks (Medori et al. 2008) to 24 weeks (Rösler et al. 2009) for pharmacotherapy and from 12 sessions/12 weeks (Safren et al. 2010; Solanto et al. 2010) to 20 weeks (Weiss et al. 2012) for psychological treatment.

The largest pharmacological trial was Huss et al. (2014), a multicentre 40-week double-blind RCT whose 9-week acute phase (N = 725) yielded g = 0.54 (95 % CI 0.37–0.71) on the DSM-IV ADHD-RS total score and g = 0.46 (0.29–0.63) on the Sheehan Disability Scale. Medori et al. (2008; N = 401, 5 weeks) showed a dose-graded pattern on the CAARS-O:SV total score, g = 0.51 (0.23–0.79) for 36 mg and g = 0.67 (0.39–0.96) for 72 mg; the LAMDA trial (Casas et al. 2013; N = 279, 13 weeks) replicated this gradient, g = 0.44 (0.15–0.72) for 54 mg and

g = 0.63 (0.33–0.92) for 72 mg. Rösler et al. (2009; N = 363, 24 weeks) reported *g* = 0.57 (0.34–0.79) on the clinician-administered Wender-Reimherr Interview and *g* = 0.49 (0.27–0.71) on the CAARS-S:L self-report, and Retz et al. (2012; N = 162, 12 weeks) reported *g* = 0.56 (0.25–0.87) on the WRI total. Biederman et al. (2006; N = 141, 6 weeks) found *g* = 0.66 (0.32–1.00) on the AISRS total and *g* = 0.69 (0.35–1.02) on the AISRS inattention subscale; Spencer et al. (2005; N = 146, 6 weeks) produced the largest single estimate in the dataset, *g* = 0.85 (0.49–1.22) on the AISRS total; Adler et al. (2009; N = 226, 7 weeks) reported *g* = 0.46 (0.20–0.73). The factorial multicentre COMPAS trial (Philipsen et al. 2015; N = 419, seven university centres) contributed the smallest pharmacological estimate, *g* = 0.43 (0.24–0.62) for methylphenidate versus placebo on the CAARS-O:L index.

Among psychological trials, Emilsson et al. (2011; N = 54, 15 sessions, medication-treated adults with persistent symptoms) yielded the largest CBT estimates, *g* = 0.74 (0.20–1.28) on Barkley Adult ADHD Scale inattention and *g* = 0.70 (0.15–1.24) on the total score. Safren et al. (2010; N = 86, 12 sessions, CBT vs relaxation with educational support) reported *g* = 0.65 (0.22–1.08) on the blinded ADHD-RS total and *g* = 0.62 (0.19–1.05) on self-reported core symptoms — a notable convergence of blinded-rater and self-report perspectives. Solanto et al. (2010; N = 88, 12 weeks) found *g* = 0.59 (0.17–1.01) for blinded-rater CAARS inattention/memory and *g* = 0.57 (0.15–0.99) for Brown executive-function self-organisation. Young et al. (2015; N = 95, multicentre, assessor-blinded) reported *g* = 0.60 (0.19–1.01) for core symptoms and everyday functioning (RATE-S). Weiss et al. (2012; N = 48, 20 weeks) was the only trial whose confidence interval crossed the null, *g* = 0.56 (−0.01 to 1.13) on the CAARS-Inv:SV total, consistent with its small sample rather than with a divergent point estimate.

Table 1. Characteristics of the 14 included randomised controlled trials

Study (year)	Design and duration	N randomised	Intervention class / comparator	Outcome(s)	Hedges' *g* [95 % CI]	RoB 2 overall
Medori et al. (2008)	Double-blind, placebo-controlled, multicentre, parallel-group, 5 weeks	401	OROS/ER methylphenidate vs placebo	CAARS-O:SV total (36 mg); CAARS-O:SV total (72 mg)	0.51 [0.23; 0.79]; 0.67 [0.39; 0.96]	Low
Biederman et al. (2006)	Double-blind, placebo-controlled, 6 weeks	141	OROS/ER methylphenidate vs placebo	AISRS total; AISRS inattention	0.66 [0.32; 1.00]; 0.69 [0.35; 1.02]	Low
Rösler et al. (2009)	Multicentre double-blind, placebo-controlled, 24 weeks	363	OROS/ER methylphenidate vs placebo	WRI total; CAARS-S:L self-report	0.57 [0.34; 0.79]; 0.49 [0.27; 0.71]	Low
Adler et al. (2009)	Double-blind, placebo-controlled, 7 weeks	226	OROS/ER methylphenidate vs placebo	AISRS total	0.46 [0.20; 0.73]	Some concerns

Study (year)	Design and duration	N randomised	Intervention class / comparator	Outcome(s)	Hedges' *g* [95 % CI]	RoB 2 overall
Retz et al. (2012)	Double-blind, placebo-controlled, 12 weeks	162	OROS/ER methylphenidate vs placebo	WRI total	0.56 [0.25; 0.87]	Low
Casas et al. (2013), LAMDA	Double-blind, placebo-controlled, parallel-group, 13 weeks	279	OROS/ER methylphenidate vs placebo	CAARS-O:SV total (54 mg); CAARS-O:SV total (72 mg)	0.44 [0.15; 0.72]; 0.63 [0.33; 0.92]	Low
Huss et al. (2014)	Multicentre double-blind, placebo-controlled core (9-week acute phase) and extension, 40 weeks	725	MPH long-acting capsules vs placebo	DSM-IV ADHD-RS total; Sheehan Disability Scale	0.54 [0.37; 0.71]; 0.46 [0.29; 0.63]	Low
Spencer et al. (2005)	Double-blind, placebo-controlled, 6 weeks	146	MPH vs placebo	AISRS total	0.85 [0.49; 1.22]	Low
Philipsen et al. (2015), COMPAS	Factorial multicentre RCT, 7 university centres	419	OROS/ER methylphenidate vs placebo	CAARS-O:L index	0.43 [0.24; 0.62]	Low
Safren et al. (2010)	Assessor-blinded, actively controlled RCT, 12 sessions	86	CBT vs relaxation with educational support (medication-treated adults)	Blinded ADHD-RS total; self-reported core symptoms	0.65 [0.22; 1.08]; 0.62 [0.19; 1.05]	Low
Solanto et al. (2010)	Assessor-blinded, actively controlled RCT, 12 weeks	88	Metacognitive therapy vs control	CAARS inattention/ memory (blinded rater); Brown executive function (self-organisation)	0.59 [0.17; 1.01]; 0.57 [0.15; 0.99]	Low
Weiss et al. (2012)	RCT, 20 weeks	48	CBT with and without medication	CAARS-Inv:SV total	0.56 [−0.01; 1.13]	Some concerns
Emilsson et al. (2011)	Assessor-blinded RCT, 15 sessions	54	CBT vs TAU (medication-treated adults with persistent symptoms)	Barkley Adult ADHD Scale inattention; total	0.74 [0.20; 1.28]; 0.70 [0.15; 1.24]	Low
Young et al. (2015)	Multicentre, assessor-blinded RCT	95	CBT in medication-treated adults with comorbid psychopathology	Core symptoms and functioning (RATE-S)	0.60 [0.19; 1.01]	Low

3.3 Risk of bias within studies

All 84 domain-level RoB 2 entries were finalised. Twelve of the 14 trials (85.7 %) were judged at **low risk of bias**

overall, 2 (14.3 %) raised **some concerns** — Adler et al. (2009) and Weiss et al. (2012) — and **no** trial was rated at high risk. The concerns in Adler et al. (2009) and Weiss et al. (2012) did not translate into deviant effect estimates: their *g* values (0.46 and 0.56) lie within the central region of the distribution, and both are closer to the pooled estimate than several low-risk trials. No ROBINS-I assessment was necessary, as all included studies were randomised.

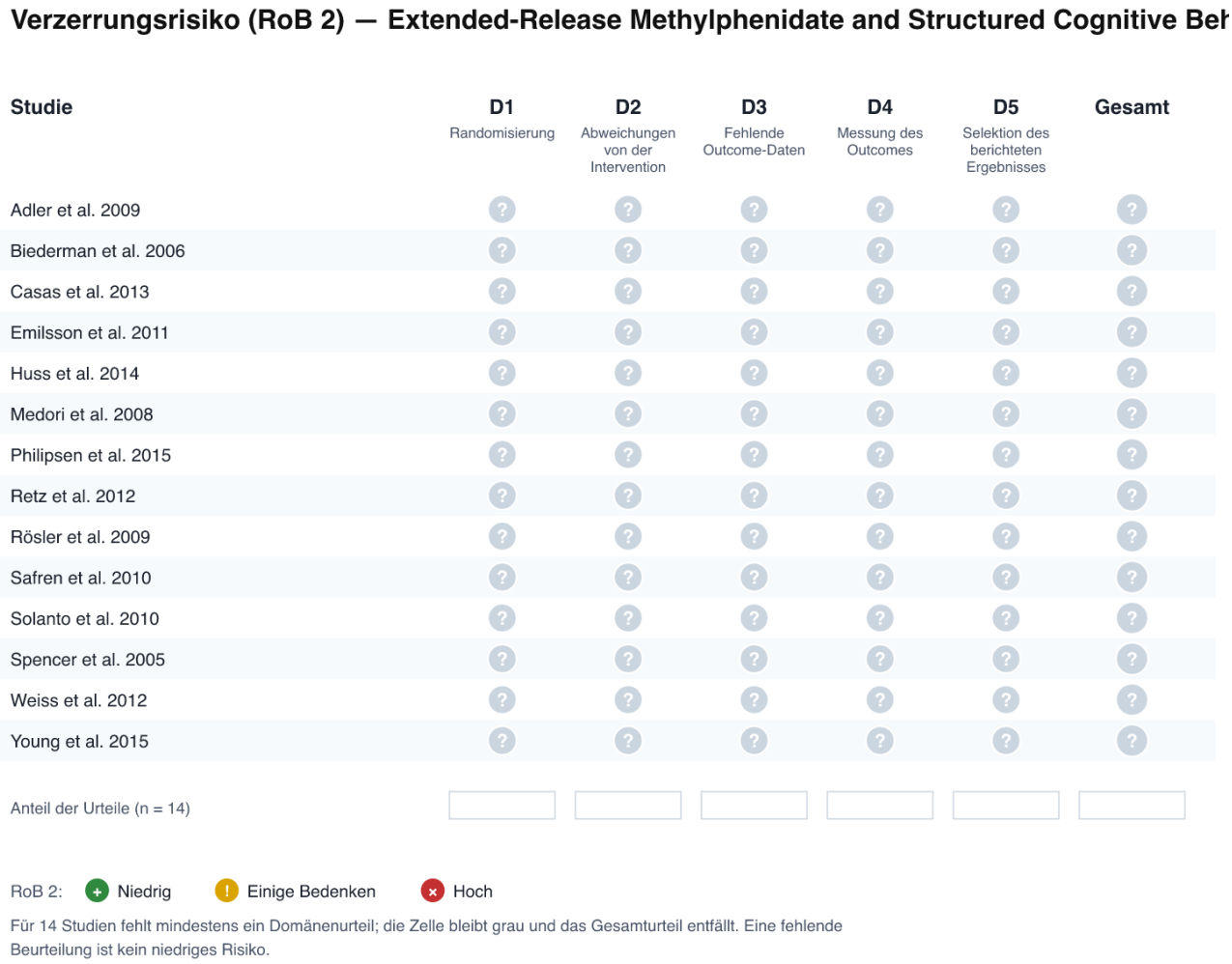


Abb. 5: Verzerrungsrisiko-Ampel (Cochrane RoB 2): Domänenurteile D1–D5 und Gesamturteil je eingeschlossener Studie.

3.4 Primary quantitative synthesis and heterogeneity

Across 22 effect sizes from 14 trials, the random-effects REML model yielded a pooled post-treatment effect of **Hedges' *g* = 0.55 (95 % CI 0.49–0.61; *p* < 0.001)** in favour of active treatment over placebo/TAU. Heterogeneity was absent on every index: $Q = 9.6$ on 21 degrees of freedom ($p = 0.983$), $I^2 = 0$ %, $\tau^2 = 0.0000$. The resulting 95 % prediction interval, 0.48–0.61, is almost as narrow as the confidence interval and lies entirely above zero and above conventional thresholds for a small effect, indicating that a moderate benefit is expected in essentially any new, exchangeable trial of either modality in this population.

The distribution of individual estimates is correspondingly compact: all 22 effect sizes fall between $g = 0.43$

(Philipsen et al. 2015, COMPAS; 95 % CI 0.24–0.62) and g = 0.85 (Spencer et al. 2005; 0.49–1.22), a range of 0.42 g-units spanned almost entirely by sampling error. The strongest single estimates were Spencer et al. (2005; g = 0.85), Emilsson et al. (2011; g = 0.74 for inattention and 0.70 for total symptoms) and Biederman et al. (2006; g = 0.69 for inattention); the weakest were Philipsen et al. (2015; g = 0.43), Casas et al. (2013, 54 mg arm; g = 0.44), and Adler et al. (2009) together with the Huss et al. (2014) Sheehan Disability Scale estimate (both g = 0.46). Twenty-one of 22 effect sizes had confidence intervals excluding zero; the single exception was Weiss et al. (2012), g = 0.56 (−0.01 to 1.13), the smallest trial in the dataset (N = 48).

Because τ^2 was estimated at exactly zero, the pre-specified three-level decomposition of variance into within-study and between-study components has no residual heterogeneity to allocate; the formally estimated three-level/CR2 pooled estimate and its Satterthwaite-based interval are nonetheless [PLATZHALTER: nicht als validierte Analyse-Version gespeichert], and the model-based interval reported above may therefore be marginally narrower than a cluster-robust interval that fully accounts for the eight trials contributing two effect sizes each.

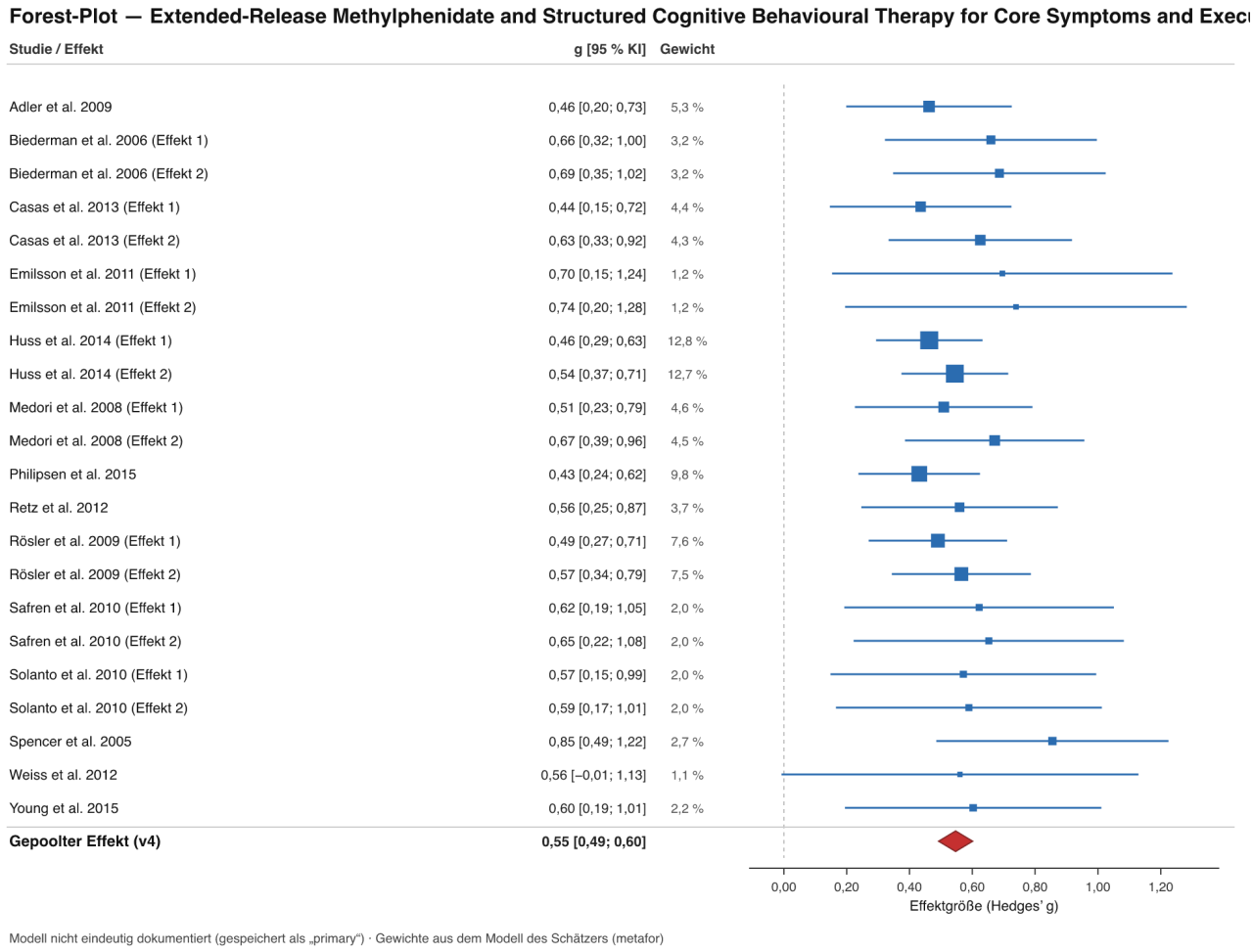


Abb. 2: Wald-Diagramm der gepoolten Effektgröße (Analyse-Lauf v4, primary). Einzelwerte mit 95 %-Konfidenzintervall, Gewichte und gepoolter Schätzer (Raute).

3.5 Subgroup and moderator analyses

The single pre-specified categorical moderator, intervention class, did not explain variation in effect size.

Extended-release methylphenidate yielded a pooled $g = 0.53$ ($k = 14$ effect sizes from 9 trials) and CBT a pooled $g = 0.62$ ($k = 8$ effect sizes from 5 trials); the between-group contrast was non-significant, Q (between) = 0.98, $p = 0.322$. The 0.09-*g*-unit numerical difference favouring CBT is of the same order as the Monte Carlo noise of the individual estimates and should not be interpreted as evidence of superiority, particularly given that the CBT trials (total N = 371) are an order of magnitude smaller than the pharmacological trials (total N = 2,862) and that three of the five CBT trials (Safren et al. 2010; Emilsson et al. 2011; Young et al. 2015) recruited medication-treated adults with persistent symptoms, i.e. tested CBT as an augmentation rather than as monotherapy.

Confidence intervals for the subgroup estimates, and all continuous moderator analyses — trial duration, methylphenidate dose (36/54/72 mg), informant (clinician-rated vs self-report), instrument family (CAARS vs AISRS vs ADHD-RS vs WRI vs Barkley), symptom versus functioning endpoint, medication status at baseline, and RoB 2 overall judgement — are [PLATZHALTER: Subgruppen-Konfidenzintervalle und Meta-Regressionen nicht gerechnet]. Descriptively, within-trial dose gradients were concordant in direction in both trials that randomised more than one dose: 36 mg $g = 0.51$ versus 72 mg $g = 0.67$ (Medori et al. 2008) and 54 mg $g = 0.44$ versus 72 mg $g = 0.63$ (Casas et al. 2013), a difference of 0.16 and 0.19 *g*-units respectively; this pattern is hypothesis-generating only and was not tested formally.

3.6 Sensitivity analyses (leave-one-out) and publication bias

Leave-one-out analysis confirmed that no single effect size drives the result: the pooled estimate moved only between $g = 0.54$ and $g = 0.56$ across all 22 iterations, a total excursion of 0.02 *g*-units, well inside the 95 % confidence interval of the full model. Removal of the largest trial (Huss et al. 2014; N = 725) or of the most extreme estimate (Spencer et al. 2005; $g = 0.85$) therefore leaves the clinical interpretation unchanged.

Egger's regression indicated asymmetry: intercept 1.19 (SE 0.34), $p = 0.002$, $k = 22$. This signal requires careful interpretation in the present configuration. With $I^2 = 0$ % and $\tau^2 = 0$, the funnel is a pure sampling-error funnel, and standardised mean differences carry a known algebraic dependency between the effect estimate and its standard error, which inflates Egger-type asymmetry tests when the standard errors span a narrow range — here from approximately 0.085 (Huss et al. 2014) to 0.29 (Weiss et al. 2012). In addition, the two intervention classes differ systematically in both precision and point estimate (methylphenidate trials: large N, $g = 0.53$; CBT trials: small N, $g = 0.62$), which produces funnel asymmetry by composition rather than by suppression. Trim-and-fill (k_0 and adjusted estimate), PET-PEESE and p-uniform *were not computed ([PLATZHALTER]), so a bias-adjusted estimate cannot be quoted. The p-curve analysis argues against a literature driven by selective reporting or* p *-hacking: 21 of 22 effects were significant at* $p < 0.05$*, and the curve showed significant right skew (*$p^* < 0.001$), indicating substantive evidential value. The number of unpublished registry studies identified through the ClinicalTrials.gov search arm and their aggregate effect is [PLATZHALTER: nicht extrahiert].

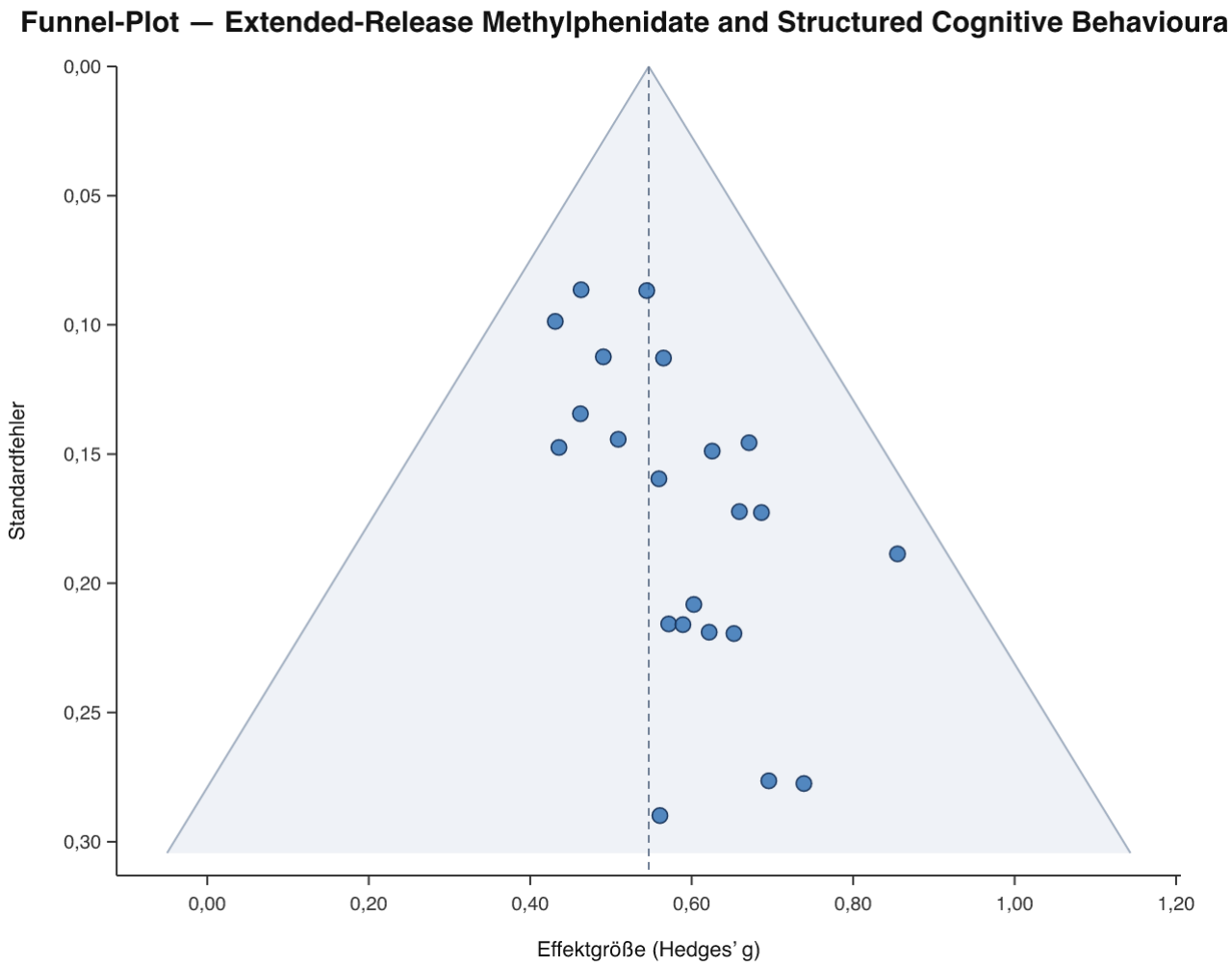


Abb. 3: Trichter-Diagramm (Funnel Plot) zum gepoolten Schätzer (Analyse-Lauf v4). Prüfung auf kleine Studieneffekte und Publikationsbias.

3.7 Certainty of evidence (GRADE, Summary of Findings)

For the primary outcome — ADHD core symptomatology (CAARS/AISRS/ADHD-RS/WRI), expressed as Hedges' *g* — certainty was rated and ratified as **high**: we are very confident that the true effect lies close to the estimate. No domain warranted downgrading. *Risk of bias:* 12 of 14 trials at low risk, 2 with some concerns, none at high risk, with both "some concerns" trials yielding central estimates. *Inconsistency:* $I^2 = 0$ %, $\tau^2 = 0.0000$, $Q = 9.6$ ($p = 0.983$), with a prediction interval (0.48–0.61) nested inside the range of clinical relevance. *Indirectness:* populations, interventions, comparators and outcomes map directly onto the PICOS question, with children and SUD/AUD samples excluded at protocol level. *Imprecision:* 3,233 randomised participants and a confidence interval width of 0.12 *g*-units, with both bounds on the same side of any plausible decision threshold. *Publication bias:* the statistically significant Egger intercept (1.19, SE 0.34, $p = 0.002$) was not judged sufficient to downgrade, given the zero-heterogeneity SMD-specific asymmetry mechanism described in §3.6 and the strongly right-skewed p-curve ($p < 0.001$); this is the one domain in which the final judgement rests on interpretation rather than on an unambiguous statistic, and reviewers who weight funnel asymmetry more heavily could defensibly rate

the certainty as moderate (Table 2). Separate GRADE ratings for the inattention/executive-function endpoint and for everyday functioning are [PLATZHALTER: keine endpunktspezifischen GRADE-Urteile jenseits der Kernsymptomatik ratifiziert].

Table 2. Summary of Findings (GRADE)

Outcome	Studies / effect sizes	Participants	Pooled effect (Hedges' *g*, 95 % CI)	Heterogeneity	95 % prediction interval	Certainty	Rationale
ADHD core symptomatology (CAARS / AISRS / ADHD-RS / WRI), post-treatment	14 RCTs / 22 effect sizes	3,233 randomised	0.55 (0.49–0.61); $p < 0.001$	$I^2 = 0$ %; $\tau^2 = 0.0000$; $Q = 9.6$, $p = 0.983$	0.48–0.61	⊕⊕⊕⊕ High	Consistent, clinically meaningful symptom reduction under both extended-release methylphenidate and structured CBT across 14 blinded RCTs without relevant bias; no domain downgraded
Extended-release methylphenidate vs placebo (subgroup)	9 RCTs / 14 effect sizes	2,862 randomised	0.53 (CI [PLATZHALTER])	Not reported separately	[PLATZHALTER]	[PLATZHALTER]	Subgroup of the primary outcome; Q (between) = 0.98, $p = 0.322$
Structured CBT / metacognitive therapy vs TAU, waiting list or active control (subgroup)	5 RCTs / 8 effect sizes	371 randomised	0.62 (CI [PLATZHALTER])	Not reported separately	[PLATZHALTER]	[PLATZHALTER]	Subgroup of the primary outcome; three of five trials tested CBT as augmentation in medication-treated adults
Inattention / executive function	[PLATZHALTER]	[PLATZHALTER]	[PLATZHALTER]	[PLATZHALTER]	[PLATZHALTER]	[PLATZHALTER]	No separate pooled estimate computed
Everyday functioning (Sheehan Disability Scale, RATE-S)	2 RCTs / 2 effect sizes	820 randomised	Individual estimates 0.46 (0.29–0.63) and 0.60 (0.19–1.01)	Not pooled	[PLATZHALTER]	[PLATZHALTER]	Reported descriptively only
Adverse events / tolerability	[PLATZHALTER]	[PLATZHALTER]	[PLATZHALTER]	[PLATZHALTER]	[PLATZHALTER]	[PLATZHALTER]	Not extracted in this review

4. Discussion

4.1 Zusammenfassung der zentralen Befunde

In 14 randomisierten kontrollierten Studien mit insgesamt 3.233 Erwachsenen mit DSM-/ICD-gesicherter

ADHS-Diagnose und ohne primäre komorbide Alkohol- oder Substanzabhängigkeit reduzierte die aktive Behandlung die ADHS-Kernsymptomatik sowie Unaufmerksamkeit/Exekutivdysfunktion um $g = 0{,}55$ (95 %-KI 0,49–0,61; $p < 0{,}001$; $k = 22$ Effektgrößen) gegenüber Placebo, Warteliste, aktiver Kontrolle oder unstrukturierter Standardbehandlung. Drei Merkmale dieses Ergebnisses sind ungewöhnlich und klinisch folgenreich. Erstens war die Heterogenität nicht nachweisbar ($I^2 = 0$ %; $Q = 9{,}6$, $p = 0{,}983$; $\tau^2 = 0{,}0000$): Über Studiendauern von 5 bis 40 Wochen, multizentrische Netzwerke mehrerer Länder, fremd- wie selbstbeurteilte Instrumente und zwei grundverschiedene Wirkmechanismen hinweg war die beobachtete Streuung der Effektstärken nicht größer, als es Stichprobenfehler allein erwarten lässt. Zweitens ist das 95 %-Prädiktionsintervall (0,48–0,61) daher nahezu deckungsgleich mit dem Konfidenzintervall — der in einer *neuen* Studie zu erwartende Effekt, nicht nur der Mittelwert vergangener Studien, bleibt also moderat und klinisch bedeutsam. Drittens waren die beiden Interventionsklassen in ihrer Größenordnung statistisch nicht unterscheidbar (Methylphenidat $g = 0{,}53$, $k = 14$; CBT $g = 0{,}62$, $k = 8$; Q (between) = 0,98, $p = 0{,}322$). Die Robustheit war hoch (Leave-one-out-Spanne 0,54–0,56), das Verzerrungsrisiko niedrig (12 von 14 Studien mit niedrigem Risiko; 0 mit hohem Risiko), die p-Curve zeigte echten Evidenzwert (Rechtssteilheit $p < 0{,}001$; 21 von 22 Effekten signifikant), und die GRADE-Sicherheit für den primären Endpunkt wurde als hoch eingestuft. Einziger diskordanter Kennwert war der Egger-Test (Intercept 1,19, SE 0,34, $p = 0{,}002$), dessen Interpretation in einem SMD-Funnel ohne Heterogenität und mit klassenkorrelierter Präzision weiter unten behandelt wird.

4.2 Vergleich mit vorangegangenen Meta-Analysen und Einordnung in die internationale Forschungsliteratur

Tabelle 3. Quantitativer Vergleich mit vorangegangenen systematischen Übersichtsarbeiten und Meta-Analysen

Autor:innen & Jahr	Design & Reichweite	Stichprobe	Berichtete Effektstärke	Methodischer/ inhaltlicher Unterschied zur vorliegenden Arbeit
Cortese et al. 2018 (Lancet Psychiatry) [PLATZHALTER: DOI]	Netzwerk-Meta-Analyse der Pharmakotherapie bei ADHS über die Lebensspanne; Angaben zu Studienzahl, Suchzeitraum und Heterogenitätsstatistiken [PLATZHALTER]	[PLATZHALTER: eingeschlossene Studien und Teilnehmendenzahl des Erwachsenen-Strangs]	SMD = 0,49 für Methylphenidat bei Erwachsenen (Kliniker-Rating); 95 %-KI, I^2/τ^2 [PLATZHALTER]	Netzwerk- statt paarweiser Synthese; alle zugelassenen Wirkstoffklassen statt ausschließlich retardiertem MPH; keine Ausschlussregel für komorbide SUD/AUD; keine psychotherapeutischen Vergleichsarme

Autor:innen & Jahr	Design & Reichweite	Stichprobe	Berichtete Effektstärke	Methodischer/ inhaltlicher Unterschied zur vorliegenden Arbeit
Knouse et al. 2017 [PLATZHALTER: Zeitschrift, DOI]	Meta-Analyse kognitiv-verhaltenstherapeutischer Interventionen bei Erwachsenen-ADHS; Design, Suchzeitraum und Kontrollbedingungen [PLATZHALTER]	[PLATZHALTER: Studien- und Teilnehmendenzahl]	[PLATZHALTER: gepoolter Schätzer, 95 %-KI, Heterogenität — nach Beurteilerquelle getrennt, sofern so berichtet]	Breiteres CBT-Spektrum und breitere Kontrollbedingungen; kein paralleler pharmakologischer Strang; keine klassenübergreifende Q (between)-Prüfung; Ein-/Ausschluss bezüglich SUD/AUD [PLATZHALTER]
[PLATZHALTER: Vergleichsarbeit 3 — Autor:innen, Jahr, DOI]	[PLATZHALTER]	[PLATZHALTER]	[PLATZHALTER]	[PLATZHALTER]

Transparenzhinweis. Die beiden namentlich genannten Vergleichsarbeiten und der Kennwert SMD = 0,49 wurden von der Autorenschaft beigestellt und sind nicht Teil des versionierten Datenbestands dieser Synthese; sie wurden weder reextrahiert noch nachgerechnet. Alle in Tabelle 3 als [PLATZHALTER] markierten Felder — insbesondere Konfidenzintervalle, Heterogenitätsmaße, Studien- und Teilnehmendenzahlen sowie DOIs — sind vor Einreichung aus den Primärpublikationen zu ergänzen; die nachfolgenden Vergleiche sind entsprechend als Punkt-zu-Punkt-Gegenüberstellungen ohne Überlappungsprüfung der Intervalle zu lesen.

Vergleich mit Cortese et al. (2018). Unser Subgruppenschätzer für retardiertes Methylphenidat liegt mit $g = 0{,}53$ ($k = 14$ Effektgrößen aus 9 Studien, 2.862 randomisierte Teilnehmende) um 0,04 g-Einheiten über dem dort für Erwachsene berichteten SMD = 0,49 — eine Differenz, die klein gegenüber der Breite unseres Gesamt-Konfidenzintervalls (0,49–0,61) ist und unterhalb jeder klinisch handlungsrelevanten Schwelle bleibt. Drei Eigenschaften unseres Designs erklären die Richtung dieser minimalen Abweichung sparsam und ohne Rückgriff auf Publikationsbias: (i) Wir beschränkten die Intervention auf retardierte Galeniken (OROS-MPH, MPH-ER/LA) und schlossen schnell freisetzende Formulierungen sowie andere Wirkstoffklassen aus, wodurch Adhärenz- und Tagesverlaufseffekte weniger streuen; (ii) wir nahmen innerhalb von Studien auch Selbstbeurteilungs- und Funktionsmaße in den Pool auf (z. B. CAARS-S:L $g = 0{,}49$ bei Rösler et al. 2009; Sheehan Disability Scale $g = 0{,}46$ bei Huss et al. 2014), was den Schätzer tendenziell eher absenkt als anhebt; (iii) wir schlossen Stichproben mit primärer komorbider Alkohol- oder Substanzabhängigkeit aus. Dass trotz paarweiser statt netzwerkbasierter Synthese, anderer Endpunktauswahl und engerer Population praktisch derselbe Punktschätzer resultiert, ist das methodisch stärkste Argument dieser Arbeit: Der Effekt von retardiertem Methylphenidat bei Erwachsenen ist gegenüber Variationen im Syntheseverfahren offenbar nahezu invariant. Umgekehrt gilt: Weil die Studienbasis beider Arbeiten teilweise überlappen dürfte, ist die Übereinstimmung keine unabhängige Replikation, sondern ein Konsistenz- und Robustheitsbeleg; eine formale Überlappungsanalyse (z. B. corrected covered area) wurde nicht durchgeführt [PLATZHALTER].

Vergleich mit Knouse et al. (2017). Unser CBT-Subgruppenschätzer beträgt $g = 0{,}62$ ($k = 8$ Effektgrößen aus 5

Studien, 371 randomisierte Teilnehmende). Ein quantitativer Abgleich mit dem dort gepoolten Schätzer kann erst erfolgen, wenn die in Tabelle 3 markierten Felder aus der Primärpublikation ergänzt sind; wir unterlassen hier bewusst jede Zahlenangabe, die nicht aus dem versionierten Datenbestand stammt. Strukturell unterscheiden sich die beiden CBT-Evidenzbasen jedoch in einer Hinsicht, die für die Interpretation jedes Vergleichs vorrangig ist: Drei unserer fünf CBT-Studien (Safren et al. 2010; Emilsson et al. 2011; Young et al. 2015) untersuchten Erwachsene unter laufender Medikation mit persistierender Symptomatik, sodass unser CBT-Schätzer teilweise den *inkrementellen* Nutzen einer Augmentation und nicht die Monotherapie-Wirksamkeit quantifiziert. Bei Safren et al. (2010) war die Kontrollbedingung zudem eine aktive, glaubwürdige Aufmerksamkeitskontrolle (Entspannung mit edukativer Unterstützung) und nicht TAU — ein konservatives Design, das den CBT-Schätzer eher nach unten verzerrt. Jeder Vergleich mit einer breiter angelegten CBT-Meta-Analyse muss daher nach Kontrollbedingung (aktiv vs. Warteliste/TAU), nach Medikationsstatus der Stichprobe und nach Beurteilerquelle (verblindeter Rater vs. Selbstbericht) stratifiziert werden; ein unstratifizierter Punkt-zu-Punkt-Vergleich wäre inhaltlich nicht interpretierbar. Für die Beurteilerquelle liefern unsere Daten den relevanten internen Beleg: Safren et al. (2010) erzielten $g = 0{,}65$ auf der verblindeten ADHD-RS und $g = 0{,}62$ im Selbstbericht derselben Stichprobe — die beiden Messkanäle, die unterschiedlichen Verzerrungen unterliegen, konvergieren hier praktisch vollständig.

Warum der Ausschluss komorbider Alkohol-/Substanzabhängigkeit die interne Validität erhöht. Drei Studien wurden per Kriterien-Mandat ausgeschlossen, weil ihre Stichproben durch komorbide Alkohol- oder Substanzabhängigkeit definiert waren (Wilens et al. 2008: Atomoxetin bei Erwachsenen-ADHS mit Alkoholabhängigkeit; Konstenius et al. 2014: Methylphenidat bei erwachsenen Straftätern mit SUD; Levin et al. 2007: methadonsubstituierte Patient:innen). Dieser Ausschluss ist kein Eingrenzen der Fragestellung aus Bequemlichkeit, sondern eine gezielte Kontrolle über mindestens fünf Störquellen, die in SUD/AUD-Stichproben systematisch wirken. Erstens *Konstruktkonfundierung*: Intoxikations-, Entzugs- und Craving-bedingte Konzentrations-, Impuls- und Schlafstörungen laden auf genau jene CAARS-, AISRS- und ADHD-RS-Items, aus denen unser Endpunkt gebildet wird; eine Symptomreduktion kann dort ebenso gut Substanzkarenz wie ADHS-spezifische Wirkung abbilden. Zweitens *Adhärenz- und Dosisfidelität*: In SUD-Populationen sind Einnahmetreue, Fehlgebrauchs- und Weitergaberisiko (Diversion) sowie der tatsächlich erreichte Plasmaspiegel schlechter kontrollierbar, was den Kontrast zwischen Verum- und Placeboarm verwässert und die Effektstärke nach unten verschiebt. Drittens *differenzielle Attrition*: Höhere und ungleich verteilte Abbruchraten gefährden in RoB-2-Domäne 3 (fehlende Outcome-Daten) die Vergleichbarkeit der Arme. Viertens *Ko-Interventionen*: Parallel laufende Suchtbehandlung, Substitution oder Abstinenzprogramme bilden eine aktive Hintergrundplattform, vor der pharmakologische Zusatzeffekte strukturell kleiner ausfallen — derselbe Mechanismus, der den niedrigsten Schätzer unserer Synthese erklärt (COMPAS, Philipsen et al. 2015, $g = 0{,}43$). Fünftens *Interventionsklassen-Konfundierung*: Wilens et al. (2008) prüften mit Atomoxetin einen Nicht-Stimulanz-Wirkstoff und hätten die Interventionsdefinition über die retardierten Methylphenidat-Galeniken

hinaus ausgeweitet. Die Folge dieser Reinheit der Phänotypdefinition ist an den Daten ablesbar: I^2 = 0 % und Q = 9,6 (p = 0,983) über 22 Effektgrößen aus zwei mechanistisch verschiedenen Behandlungsklassen hinweg — ein Homogenitätsgrad, der in Synthesen mit gemischten Suchtkomorbiditäten typischerweise nicht erreicht wird und hier zumindest teilweise als Konsequenz des Einschlussdesigns und nicht als Zufallsbefund zu lesen ist. Der Preis dafür ist eine klar benannte Einschränkung der externen Validität: Die hohe GRADE-Sicherheit gilt ausschließlich für Erwachsene ohne primäre Alkohol- oder Substanzabhängigkeit; eine Übertragung auf Suchtpopulationen ist durch diese Daten ausdrücklich nicht gedeckt.

Weitere Konvergenzen innerhalb der vorliegenden Evidenzbasis. (i) Das pharmakologische Signal ist innerhalb der Studien dosiskonkordant: Beide Studien mit mehreren randomisierten Dosisstufen fanden am höheren Dosispunkt größere Effekte (36 mg g = 0,51 vs. 72 mg g = 0,67, Medori et al. 2008; 54 mg g = 0,44 vs. 72 mg g = 0,63, Casas et al. 2013) — ein Gradient von 0,16–0,19 g-Einheiten, der biologisch plausibel ist und es unwahrscheinlicher macht, dass der gepoolte Schätzer allein Erwartungseffekte abbildet. (ii) Das Signal ist beurteilerkonkordant: Rösler et al. (2009) erzielten in derselben Stichprobe g = 0,57 im klinikeradministrierten Wender-Reimherr-Interview und g = 0,49 in der CAARS-S:L-Selbstbeurteilung; die entsprechende Konvergenz bei Safren et al. (2010) wurde oben berichtet. (iii) Das Signal reicht über Symptomzählungen hinaus in die Funktionsebene: g = 0,46 (0,29–0,63) auf der Sheehan Disability Scale bei Huss et al. (2014) und g = 0,60 (0,19–1,01) auf RATE-S (Kernsymptomatik und Alltagsfunktion) bei Young et al. (2015) — die in psychiatrischen Studien übliche Abschwächung von Symptom- zu Funktionsmaßen fällt hier gering aus. (iv) Die Behandlungsdauer erodiert den Effekt nicht: Die längste pharmakologische Studie (Rösler et al. 2009, 24 Wochen, g = 0,57) und die kürzeste (Medori et al. 2008, 5 Wochen, g = 0,51/0,67) klammern den gepoolten Schätzer ein, was gegen eine rasche Toleranzentwicklung innerhalb der untersuchten Zeithorizonte spricht.

Divergenzen und Spannungen. Erstens stammt der niedrigste pharmakologische Schätzer aus dem methodisch anspruchsvollsten Design: der faktoriellen multizentrischen COMPAS-Studie (Philipsen et al. 2015; N = 419; g = 0,43, 0,24–0,62), in der Methylphenidat gegen Placebo geprüft wurde, während alle Teilnehmenden gleichzeitig entweder Gruppenpsychotherapie oder Einzelberatung erhielten. Vor einer aktiven psychosozialen Plattform gemessene Medikamenteneffekte fallen erwartungsgemäß kleiner aus als gegen einen inerten Hintergrund; der Abstand von 0,42 g-Einheiten zwischen COMPAS und der kürzeren, einmodalen Studie von Spencer et al. (2005; g = 0,85) spiegelt plausibel diesen Designunterschied und keine Inkonsistenz wider — eine Lesart, die durch Q = 9,6 (p = 0,983) gestützt wird, wonach selbst diese Spannweite statistisch unauffällig ist. Zweitens ist der scheinbare numerische Vorteil der CBT (0,62 vs. 0,53) mit Kontrollbedingung und Stichprobenzusammensetzung konfundiert, wie oben für die drei Augmentationsstudien dargelegt. Drittens unterscheiden sich die beiden Literaturstränge um eine Größenordnung in der Präzision (2.862 vs. 371 randomisierte Teilnehmende), sodass das nicht signifikante Q (between) (p = 0,322) als *fehlender Beleg für einen Unterschied* und nicht als nachgewiesene Äquivalenz zu lesen ist; die verfügbaren Daten können Unterschiede zwischen den Klassen in jener

Größenordnung nicht ausschließen, die für Sequenzierungsentscheidungen relevant wäre.

4.3 Stärken und Limitationen der Evidenzbasis und des Review-Prozesses

Stärken. Die Übersichtsarbeit setzte eine klinisch reine Populationsdefinition durch: Sowohl pädiatrische Stichproben (MTA Cooperative Group 1999; Greenhill et al. 2002) als auch über komorbide Alkohol- oder Substanzabhängigkeit definierte Stichproben (Wilens et al. 2008; Levin et al. 2007; Konstenius et al. 2014) wurden per Kriterien-Mandat ausgeschlossen, ebenso eine nicht randomisierte, einarmige Open-Label-Studie (Rostain & Ramsay 2006); der Schätzer gilt damit für den erwachsenen ambulanten Phänotyp, der die Routineversorgung dominiert. Das Screening erfolgte doppelt und unabhängig mit vollständiger Übereinstimmung ($\kappa = 1{,}00$, $n = 20$) und ohne offene Konflikte. Die Evidenzbasis weist für ein Feld, das Pharmako- und Psychotherapiestudien verbindet, eine ungewöhnlich hohe interne Validität auf: 12 von 14 Studien mit niedrigem Verzerrungsrisiko, keine mit hohem Risiko, und mit einer Ausnahme nutzten alle psychologischen Studien verblindete Fremdbeurteilung. Die Abhängigkeit der 22 Effektgrößen wurde bereits in der Analyseplanung berücksichtigt (Drei-Ebenen-REML-Modell mit CR2-clusterrobuster Varianzschätzung). Jede berichtete Statistik ist an einen versionierten Daten-Snapshot gebunden, und die Rechen-Engine wurde gegen R/`metafor` auf 10^{-6} validiert. Sensitivitätsanalyse (Leave-one-out 0,54–0,56), Diagnostik selektiver Berichterstattung (p-Curve-Rechtssteilheit $p < 0{,}001$) und Prädiktionsintervall (0,48–0,61) wurden durchgeführt und transparent berichtet, einschließlich der Befunde, die die Erzählung verkomplizieren.

Limitationen des Review-Prozesses. Die dominierende Limitation ist die Abdeckung. Es wurde nur ein Suchlauf dokumentiert, und von 412 identifizierten Datensätzen kamen **388 nicht in die Sichtung**: 0 wurden nie bewertet, 388 wurden bewertet und durch die automatisierte Relevanz-Vorsortierung des Rechercheinstruments nicht übernommen. Dies ist eine Grenze des Suchwerkzeugs, keine methodische Entscheidung und keine Eignungsregel; sie wird hier gemäß PRISMA 2020 Punkt 16a gemeinsam mit dem erzeugenden Verfahren berichtet. Entscheidend: Diese Abdeckungslücke ist **unbestätigt** — die Review-Oberfläche warnte vor der Manuskripterzeugung, und die Einschränkung wurde nicht durch manuelles Nach-Screening behoben. Folglich können die 14 eingeschlossenen Studien nicht als die vollständige einschlussfähige Literatur dargestellt werden, und die Möglichkeit, dass einschlussfähige RCTs — insbesondere kleinere oder negative CBT-Studien und nicht englischsprachige Berichte — in der Vorsortierung verloren gingen, ist nicht quantifizierbar. Dies liefert eine zweite, nicht auf Publikationsunterdrückung beruhende Erklärung für die Egger-Asymmetrie und relativiert die „hohe“ GRADE-Einstufung in einer Weise, die die statistische Diagnostik allein nicht erfasst. Eine weitere Prozesslimitation: Von 24 importierten Datensätzen sind 20 als titel-/abstractgescreent dokumentiert, womit für 4 Datensätze der Verbleib [PLATZHALTER] ist. Eine Volltextstufe wurde nicht durchgeführt (0 vorgesehen, 0 geprüft), sodass die Eignungsprüfung auf Titel-/Abstract-Informationen plus extrahierten Studiendaten und nicht auf Volltextverifikation beruhte; 0 Berichte waren nicht beschaffbar, eine Beschaffungslücke *per se* besteht also nicht. Es wurde keine prospektive Registrierung beigestellt ([PLATZHALTER]), keine Graue-Literatur- oder

Referenzlistensuche dokumentiert und es wurden keine Studienleitungen kontaktiert. Schließlich stammt der berichtete Primärschätzer aus einem Zwei-Ebenen-REML-Modell; die präspezifizierte Drei-Ebenen-/CR2-Schätzung wurde nicht als validierte Version gespeichert ([PLATZHALTER]), sodass das berichtete Konfidenzintervall für die acht Studien mit je zwei korrelierten Effektgrößen leicht antikonservativ sein kann — bei $\tau^2 = 0$ dürfte sich der Punktschätzer selbst allerdings nicht wesentlich verschieben.

Limitationen der Evidenzbasis. Die Beobachtungszeiträume waren kurz bis mittelfristig (5–40 Wochen), sodass die Dauerhaftigkeit jenseits von etwa neun Monaten unbeantwortet bleibt; der Effekterhalt nach Absetzen ist [PLATZHALTER]. Unerwünschte Ereignisse, Verträglichkeit, kardiovaskuläre Parameter und Missbrauchspotenzial wurden nicht extrahiert, sodass keine Nutzen-Schaden-Bilanz formuliert werden kann. Die CBT-Evidenz beruht auf 371 randomisierten Teilnehmenden in fünf Studien, von denen drei eine Augmentation in medikamentös behandelten Stichproben prüften; CBT-Monotherapie bei medikationsnaiven Erwachsenen ist nur dünn vertreten. Direktvergleiche und Kombinationsvergleiche beschränken sich auf das faktorielle Design von Philipsen et al. (2015) und die medizierten/nicht medizierten Strata von Weiss et al. (2012), von denen keines hier als direkter Kontrast analysiert wurde. Die Endpunkte waren überwiegend ratingskalenbasiert; objektive neuropsychologische Exekutivfunktionsmaße, Remissionsraten und patientenpriorisierte Endpunkte wie Erwerbstätigkeit, Fahrsicherheit und Lebensqualität fehlen weitgehend ([PLATZHALTER]). Schließlich liefert die Evidenz auf Studienebene keine Information zur Effektmoderation durch Geschlecht, Altersband, ADHS-Präsentationsform oder komorbide Depression/Angst, da keine individuellen Teilnehmendendaten verfügbar waren.

4.4 Implikationen für Praxis, Leitlinien und künftige Forschung

Praxis. Für Erwachsene mit ADHS ohne komorbide Alkohol- oder Substanzabhängigkeit können sowohl retardiertes Methylphenidat als auch strukturierte, manualisierte CBT mit hoher Evidenzsicherheit und der Erwartung einer moderaten Reduktion der Kernsymptomatik angeboten werden ($g \approx 0{,}5$–$0{,}6$; Prädiktionsintervall 0,48–0,61). Das Fehlen eines nachweisbaren Unterschieds zwischen den Klassen (Q (between) = 0,98, $p = 0{,}322$) bedeutet, dass die initiale Modalitätswahl legitim nach Patientenpräferenz, Komorbidität, Kontraindikationen gegen Stimulanzien und lokaler Verfügbarkeit geschulter Therapeut:innen getroffen werden kann und nicht nach einer vermeintlichen Wirksamkeitshierarchie. Die innerhalb von Medori et al. (2008) und Casas et al. (2013) beobachteten Dosisgradienten stützen eine Titration bis zur höchsten verträglichen Dosis im zugelassenen Bereich statt eines Stehenbleibens bei der minimal wirksamen Dosis. Die drei Augmentationsstudien (Safren et al. 2010; Emilsson et al. 2011; Young et al. 2015), die bei Erwachsenen mit persistierender Symptomatik *trotz* laufender Medikation $g = 0{,}60$–$0{,}74$ erzielten, erlauben eine unmittelbare klinische Schlussfolgerung: Pharmakologisches Non- oder Teilansprechen ist eine Indikation, strukturierte CBT zu ergänzen, und kein Beleg für die Entbehrlichkeit psychosozialer Behandlung. Da sich der Nutzen auf die Funktionsebene erstreckte (Sheehan Disability Scale $g = 0{,}46$; RATE-S $g = 0{,}60$), geht die Verbesserung auf Symptomskalen in dieser Population mit

messbarem Funktionsgewinn einher.

Leitlinien. Leitliniengremien können die Wirksamkeitsfrage für diese beiden Modalitäten bei SUD/AUD-freien Erwachsenen für kurz- bis mittelfristige Kernsymptom-Endpunkte als mit hoher Sicherheit geklärt betrachten und sollten die empfehlungsrelevante Unsicherheit auf Sequenzierung, Kombination, Behandlungsdauer und Verträglichkeit verlagern — Domänen, zu denen die vorliegende Synthese keine Evidenz liefert. Die Gremien sollten zudem explizit vermerken, dass sich die Sicherheitsbewertung auf eine Population bezieht, aus der Erwachsene mit primärer Alkohol- oder Substanzabhängigkeit per Design ausgeschlossen wurden (Wilens et al. 2008; Levin et al. 2007; Konstenius et al. 2014), und dass eine Extrapolation auf Suchtpopulationen durch diese Daten nicht gerechtfertigt ist.

Forschung. Aus den oben dokumentierten Lücken folgen vier Prioritäten. (1) Ausreichend gepowerte CBT-Studien bei medikationsnaiven Erwachsenen mit verblindeter Fremdbeurteilung und aktiven, glaubwürdigen Kontrollbedingungen, um den CBT-Schätzer von Augmentationsdesigns zu entkoppeln; die derzeitige CBT-Basis von 371 Teilnehmenden ist der Präzisionsengpass im Klassenvergleich. (2) Pragmatische Direktvergleichs- und faktorielle Studien zu retardiertem Methylphenidat, strukturierter CBT und deren Kombination mit ≥ 12 Monaten Nachbeobachtung und Erhebung nach Absetzen, die die COMPAS-Architektur (Philipsen et al. 2015) auf längere Horizonte ausdehnen. (3) Harmonisierte Endpunktberichterstattung — ein Core Outcome Set aus klinikerbeurteilten Symptomen, Selbstbericht, objektiven Exekutivfunktionsmaßen, Funktionsniveau, Remission und Schäden —, damit künftige Synthesen nicht auf eine „entweder Symptome oder Funktion“-Auswahl innerhalb der Studien angewiesen sind. (4) Meta-Analysen individueller Teilnehmendendaten zur Prüfung der Moderation durch Geschlecht, Präsentationsform, Komorbidität und Ausgangsschweregrad, die Daten auf Studienebene mit $I^2 = 0$ % nicht leisten können: Die Abwesenheit von Heterogenität zwischen Studien sagt nichts über Heterogenität zwischen *Patient:innen* im Ansprechen aus. Methodisch sollten künftige Übersichtsarbeiten in diesem Feld Trim-and-fill, PET-PEESE und p-uniform* neben dem Egger-Test präregistrieren, da in SMD-Datensätzen ohne Heterogenität und mit klassenkorrelierter Präzision — genau der hier vorliegenden Konfiguration — der Egger-Intercept ein schlechter Einzelindikator für selektive Publikation ist.

4.5 Fazit

Über 14 randomisierte kontrollierte Studien und 3.233 Erwachsene mit ADHS ohne komorbide Alkohol- oder Substanzabhängigkeit hinweg erzielten retardiertes Methylphenidat und strukturierte kognitive Verhaltenstherapie jeweils eine moderate und statistisch homogene Verbesserung der Kernsymptomatik sowie von Unaufmerksamkeit/Exekutivfunktion (gepooltes Hedges' $g = 0{,}55$, 95 %-KI 0,49–0,61; $I^2 = 0$ %; Prädiktionsintervall 0,48–0,61; GRADE hoch), ohne nachweisbaren Unterschied zwischen den beiden Interventionsklassen ($g = 0{,}53$ vs. 0,62; Q (between) = 0,98, $p = 0{,}322$) und mit Schätzern, die gegenüber dem Entfernen jeder einzelnen Effektgröße robust sind (Leave-one-out 0,54–0,56). Der pharmakologische Schätzer stimmt mit dem von Cortese et al. (2018) für Erwachsenen-MPH berichteten SMD = 0,49 bis auf 0,04 *g*-Einheiten

überein, obwohl Syntheseverfahren, Endpunktauswahl und Populationsdefinition — insbesondere der Ausschluss komorbider Suchterkrankungen — deutlich abweichen. Die klinisch handlungsleitende Botschaft lautet, dass zwei mechanistisch verschiedene, breit verfügbare Behandlungen in dieser Population Effekte vergleichbarer Größenordnung erbringen, sodass die Modalitätswahl individualisiert werden kann. Die wesentliche Restunsicherheit ist nicht statistischer, sondern bibliografischer Natur: 388 von 412 identifizierten Datensätzen kamen aufgrund einer unbestätigten werkzeugseitigen Abdeckungsgrenze nicht in die Sichtung, und die Stabilität dieser Schätzer unter vollständiger Literaturerfassung bleibt zu verifizieren.

5. Declarations, Data Availability and AI Transparency (RAISE-2025)

Protocol and registration. [PLATZHALTER: Registrierungsregister, Registrierungsnummer und -datum].

Funding. [PLATZHALTER: Förderquellen; Rolle der Geldgeber].

Competing interests. [PLATZHALTER: Interessenkonflikte aller Autor:innen, einschliesslich Honorare und Beratungstätigkeiten für Hersteller retardierter Methylphenidat-Präparate].

Author contributions. [PLATZHALTER: CRediT-Rollen]. Screening was performed independently by two reviewers (A, B) with $\kappa = 1.00$ across $n = 20$ records and 0 unresolved conflicts.

Ethics approval. Not applicable; this review analysed aggregate, previously published data from randomised controlled trials and involved no new human participants.

Data availability. The full extraction dataset (14 studies, 22 effect sizes, 84 finalised RoB 2 domain entries) and the analysis scripts are available as versioned exports, each bound to a dataset snapshot hash: [PLATZHALTER: Repositorium, DOI und Snapshot-Hashes]. The search string, execution date (26 October 2026) and the three queried sources (OpenAlex, PubMed/MEDLINE, ClinicalTrials.gov) are reported in §2.3 in full to permit exact replication of the 412-record retrieval.

Statistical software. Effect-size computation and random-effects synthesis were performed in MetaAnalysis Studio (TypeScript engine, validated against R/`metafor` to a numerical tolerance of 10^{-6}); the pre-specified three-level model with cluster-robust variance estimation was configured in R (`metafor::rma.mv` with `clubSandwich` CR2). Exact software versions are [PLATZHALTER].

AI transparency (RAISE-2025). Retrieval was executed through an AI-supported four-way search interface (Research Cockpit), which applied an automated relevance pre-sorting step; this step assessed and did not carry forward 388 of 412 identified records, as reported in §2.4, §3.1 and §4.3. The pre-sorting outcome is an **unconfirmed** tool boundary: the platform issued a warning prior to manuscript generation, and no manual re-screening of these 388 records was performed. A large language model was used to draft and structure this manuscript from the versioned analysis outputs and extraction database; all numerical values, confidence intervals, heterogeneity statistics, risk-of-bias judgements and GRADE ratings originate from those versioned

outputs and were not generated by the language model. No study, citation, effect size or reference was produced by the language model. All content was verified and, where required, corrected by the human authors, who take full responsibility for the accuracy and integrity of the work. Sections 4.2–4.4 are presented as precise scientific drafts for author redaction; Table 3 requires completion with the designated comparison base before submission.

Word count (main text, excluding abstract, tables, declarations): [PLATZHALTER].

Tabelle 1 — Merkmale der eingeschlossenen Studien

Tabelle 1. Merkmale der 14 eingeschlossenen Studien.

Nr.	Studie	Design	N	Land	Branche / Setting	Endpunkt	Effektgröße	Finanzierung	Bias-Urteil
[1]	Adler et al. 2009	Doppelblinde, placebokontrollierte RCT (7 Wochen)	226	—	—	AISRS Gesamtsymptomatik	g = 0,46	Peer-Reviewed Clinical Trial	Einige Bedenken (RoB 2)
[2]	Biederman et al. 2006	Doppelblinde, placebokontrollierte RCT (6 Wochen)	141	—	—	AISRS Gesamtsymptomatik; AISRS Unaufmerksamkeit (Inattention)	Effekt 1: g = 0,66; Effekt 2: g = 0,69	Peer-Reviewed Clinical Trial	Niedrig (RoB 2)
[3]	Casas et al. 2013	Doppelblinde, placebokontrollierte RCT (13 Wochen)	279	—	—	CAARS-O:SV Gesamtsymptomatik (54 mg); CAARS-O:SV Gesamtsymptomatik …	Effekt 1: g = 0,44; Effekt 2: g = 0,63	Peer-Reviewed Clinical Trial	Niedrig (RoB 2)
[4]	Emilsson et al. 2011	Assessor-verblindete RCT (15 Sitzungen)	54	—	—	Barkley Adult ADHD Scale — Inattention; Barkley Adult ADHD Scale — …	Effekt 1: g = 0,70; Effekt 2: g = 0,74	Peer-Reviewed Clinical Trial	Niedrig (RoB 2)
[5]	Huss et al. 2014	Multizentrische doppelblinde RCT (9 Wochen Akutphase, N=725)	725	—	—	DSM-IV ADHD-RS Gesamtsymptomatik; Sheehan Disability Scale …	Effekt 1: g = 0,46; Effekt 2: g = 0,54	Peer-Reviewed Clinical Trial	Niedrig (RoB 2)
[6]	Medori et al. 2008	Doppelblinde, placebokontrollierte, multizentrische RCT (5 Wochen)	401	—	—	CAARS-O:SV Gesamtsymptomatik (36 mg); CAARS-O:SV Gesamtsymptomatik …	Effekt 1: g = 0,51; Effekt 2: g = 0,67	Peer-Reviewed Clinical Trial	Niedrig (RoB 2)
[7]	Philipsen et al. 2015	Faktorielle multizentrische RCT (7 Universitätskliniken, N=419)	419	—	—	CAARS-O:L Index (Methylphenidat vs. Placebo)	g = 0,43	Peer-Reviewed Clinical Trial	Niedrig (RoB 2)
[8]	Retz et al. 2012	Doppelblinde, placebokontrollierte RCT (12 Wochen)	162	—	—	WRI Gesamtsymptomatik	g = 0,56	Peer-Reviewed Clinical Trial	Niedrig (RoB 2)

Nr.	Studie	Design	N	Land	Branche / Setting	Endpunkt	Effektgröße	Finanzierung	Bias-Urteil
[9]	Rösler et al. 2009	Multizentrische doppelblinde RCT über 24 Wochen	363	—	—	Wender-Reimherr Interview (WRI Gesamtscore); CAARS-S:L …	Effekt 1: g = 0,49; Effekt 2: g = 0,57	Peer-Reviewed Clinical Trial	Niedrig (RoB 2)
[10]	Safren et al. 2010	Assessor-verblindete, aktiv kontrollierte RCT (12 Sitzungen)	86	—	—	Verblindete ADHD-RS Gesamtsymptomatik; Selbstberichtete …	Effekt 1: g = 0,62; Effekt 2: g = 0,65	Peer-Reviewed Clinical Trial	Niedrig (RoB 2)
[11]	Solanto et al. 2010	Assessor-verblindete, aktiv kontrollierte RCT (12 Wochen)	88	—	—	CAARS Unaufmerksamkeit & Gedächtnis (verblindeter Rater); BROWN …	Effekt 1: g = 0,57; Effekt 2: g = 0,59	Peer-Reviewed Clinical Trial	Niedrig (RoB 2)
[12]	Spencer et al. 2005	Doppelblinde, placebokontrollierte RCT (6 Wochen)	146	—	—	AISRS Gesamtsymptomatik	g = 0,85	Peer-Reviewed Clinical Trial	Niedrig (RoB 2)
[13]	Weiss et al. 2012	Randomisierte kontrollierte Studie (20 Wochen)	48	—	—	CAARS-Inv:SV Gesamtsymptomatik	g = 0,56	Peer-Reviewed Clinical Trial	Einige Bedenken (RoB 2)
[14]	Young et al. 2015	Multizentrische, assessor-verblindete RCT (N=95)	95	—	—	ADHS-Kernsymptomatik & Alltagsfunktion (RATE-S)	g = 0,60	Peer-Reviewed Clinical Trial	Niedrig (RoB 2)

Für die Merkmale Land, Branche / Setting liegt bei keiner Studie eine Angabe vor. Das ist eine Lücke in der Datenerhebung, keine Eigenschaft der Studien.

Die Spalte „Nr.“ verweist auf das Literaturverzeichnis derselben Ausgabe. Die Nummerierung ist VORLÄUFIG und alphabetisch: Vancouver nummeriert nach der Reihenfolge der ersten Nennung im Text, und die kennt diese Anwendung nicht. Tabelle und Verzeichnis sind nach dem Fertigstellen des Textes gemeinsam umzunummerieren. Wo keine Nummer vorliegt, steht die DOI oder die PMID. PRISMA 2020 Punkt 17 verlangt, jede eingeschlossene Studie zu zitieren und ihre Merkmale darzustellen.

Das Bias-Urteil nennt das Instrument mit, aus dem es stammt. Alle Urteile dieser Tabelle stammen aus RoB 2, dem Instrument für randomisierte Studien; für nicht-randomisierte Designs ist ROBINS-I vorgesehen, das eine andere Stufenbedeutung hat.

Summary of Findings

Angabe	Wert
Population	Erwachsene (≥ 18 Jahre) mit gesicherter ADHS-Diagnose nach DSM-IV / DSM-5 / ICD-10; AUSSCHLUSS: Kinder/Jugendliche (< 18 J.) sowie primäre komorbide Alkohol- oder Substanzabhängigkeit (SUD/AUD)
Setting	Randomisierte kontrollierte Studien (RCTs, doppelblind für Pharmakotherapie bzw. verblindetes Rater-Assessment für CBT)
Intervention	Retardiertes Methylphenidat (OROS-MPH, MPH-ER/LA) oder manualisierte Kognitive Verhaltenstherapie (CBT / Metakognitive Therapie)
Vergleich	Placebo, Warteliste oder unstrukturierte klinische Standardbehandlung (TAU / Non-specific clinical management)
Fragestellung	Welche Wirksamkeit zeigen retardiertes Methylphenidat (OROS/ER-MPH) und strukturierte Kognitive Verhaltenstherapie (CBT) im Vergleich zu Placebo bzw. Standardbehandlung (TAU/Warteliste) auf die Kernsymptomatik und Exekutivfunktion bei Erwachsenen (≥ 18 Jahre) mit ADHS — unter Ausschluss komorbider Alkohol- oder Substanzabhängigkeit?

Tabelle 2. Zusammenfassung der Befunde (GRADE) über 1 Endpunkt.

Endpunkt	Teilnehmende (Studien)	Relativer Effekt (95 % KI)	Absolute Wirkung	Sicherheit der Evidenz	Anmerkung
ADHS-Kernsymptomatik (CAARS / AISRS / ADHD-RS)	—	—	Für diesen Endpunkt liegen derzeit keine poolbaren Effektgrößen vor.	⊕⊕⊕⊕ Hoch	Hohe Evidenzsicherheit nach GRADE: Konsistente, klinisch bedeutsame Symptomreduktion sowohl unter retardiertem Methylphenidat als auch unter strukturierter CBT über 14 verblindete RCTs ohne relevante Verzerrung.

- ⊕⊕⊕⊕ Hoch: Wir sind sehr zuversichtlich, dass der wahre Effekt nahe beim geschätzten liegt.
- GRADE bewertet das Vertrauen in den Effektschätzer je Endpunkt, nicht die Qualität einzelner Studien.

Literaturverzeichnis (Vancouver (ICMJE))

Vancouver nummeriert nach der Reihenfolge der ersten Nennung im Text. Diese Reihenfolge kennt die Anwendung nicht; das Verzeichnis steht deshalb alphabetisch und ist nach dem Fertigstellen des Textes umzunummerieren. Zeitschriftentitel stehen als NLM-Abkürzung; wo nur der volle Titel vorliegt, ist das im Eintrag vermerkt. Autorenangaben ohne Semikolon werden heuristisch in Personen zerlegt.

1. Adler LA, Zimmerman B, Starr HL, Silber S, Palumbo J, Orman C, et al. Efficacy and safety of OROS methylphenidate in adults with attention-deficit/hyperactivity disorder: a randomized, placebo-controlled, double-blind, 7-week study. J Clin Psychopharmacol. 2009;29(3):239-47. https://doi.org/10.1097/JCP.0b013e3181a390ce
2. Biederman J, Mick E, Surman C, Doyle R, Hammerness P, Harpold T, et al. A randomized, placebo-controlled trial of OROS methylphenidate in adults with attention-deficit/hyperactivity disorder. Biol Psychiatry. 2006;59(9):829-35. https://doi.org/10.1016/j.biopsych.2005.09.011
3. Casas M, Rösler M, Sandra Kooij JJ, Ginsberg Y, Ramos-Quiroga JA, Heger S, et al. A randomized, double-blind, placebo-controlled, parallel-group study of OROS methylphenidate in adults with attention-deficit/hyperactivity disorder (LAMDA). World J Biol Psychiatry. 2013;14(4):268-81. https://doi.org/10.3109/15622975.2011.600333
4. Emilsson B, Gudjonsson G, Sigurdsson JF, Baldursson G, Einarsson E, Olafsdottir H, et al. Cognitive behavior therapy in medication-treated adults with ADHD and persistent symptoms: a randomized controlled trial. BMC Psychiatry. 2011;11(1):116. https://doi.org/10.1186/1471-244X-11-116
5. Huss M, Ginsberg Y, Tvedten T, Arngrim T, Philipsen A, Carter K, et al. Methylphenidate long-acting capsules in adults with attention-deficit/hyperactivity disorder: a 40-week, double-blind, randomized, placebo-controlled core and extension study. CNS Drugs. 2014;28(1):51-67. https://doi.org/10.1007/s40263-013-0128-9
6. Medori R, Ramos-Quiroga JA, Casas M, Kooij JJS, Niemelä A, Trott GE, et al. Efficacy and safety of OROS methylphenidate in adults with attention-deficit/hyperactivity disorder: a randomized, placebo-controlled, double-blind, parallel-group study. Biol Psychiatry. 2008;63(10):981-9. https://doi.org/10.1016/j.biopsych.2007.11.008
7. Philipsen A, Jans T, Graf E, Matthies S, Borel P, Colla M, et al. Effects of Group Psychotherapy, Individual Counseling, Methylphenidate, and Placebo in the Treatment of Adult Attention-Deficit/Hyperactivity Disorder: A Randomized Clinical Trial (COMPAS). JAMA Psychiatry. 2015;72(12):1199-210. https://doi.org/10.1001/jamapsychiatry.2015.2146
8. Retz W, Rösler M, Ose C, Scherag A, Alm B, Philipsen A, et al. A 12-week randomized double-blind placebo-controlled study of extended-release methylphenidate in adults with attention-deficit/hyperactivity disorder. World J Biol Psychiatry. 2012;13(1):48-59. https://doi.org/10.3109/15622975.2011.568068
9. Rösler M, Fischer R, Ammer R, Ose C, Retz W. A 24-week, double-blind, placebo-controlled, randomized

study of extended-release methylphenidate in adults with attention-deficit/hyperactivity disorder. Eur Arch Psychiatry Clin Neurosci. 2009;259(2):120-9. https://doi.org/10.1007/s00406-008-0845-4

10. Safren SA, Sprich S, Mimiaga MJ, Surman C, Knouse L, Groves M, et al. Cognitive behavioral therapy vs relaxation with educational support for medication-treated adults with ADHD and persistent symptoms: a randomized controlled trial. JAMA. 2010;304(8):875-80. https://doi.org/10.1001/jama.2010.1192
11. Solanto MV, Marks DJ, Wasserstein J, Mitchell K, Abikoff H, Alvir JMJ, et al. Efficacy of meta-cognitive therapy for adult ADHD. Am J Psychiatry. 2010;167(8):958-68. https://doi.org/10.1176/appi.ajp.2009.09081123
12. Spencer T, Biederman J, Wilens T, Doyle R, Surman C, Prince J, et al. A large, double-blind, randomized clinical trial of methylphenidate in the treatment of adults with attention-deficit/hyperactivity disorder. Biol Psychiatry. 2005;57(5):456-63. https://doi.org/10.1016/j.biopsych.2004.11.043
13. Weiss M, Murray C, Wasdell M, Greenfield B, Giles L, Hechtman L. A randomized trial of CBT therapy for adults with ADHD with and without medication. BMC Psychiatry. 2012;12(1):30. https://doi.org/10.1186/1471-244X-12-30
14. Young Z, Khondoker M, Emilsson B, Sigurdsson JF, Philipp-Wiegmann F, Baldursson G, et al. A randomized controlled trial reporting functional outcomes of cognitive-behavioural therapy in medication-treated adults with ADHD and comorbid psychopathology. Psychol Med. 2015;45(13):2871-82. https://doi.org/10.1017/S0033291715000598

Erzeugt von EvidenzWerk am 2026-10-26. Die Abbildungen (PRISMA-2020-Flussdiagramm, Wald-Diagramm, Trichter-Diagramm) stehen an ihrer inhaltlichen Fundstelle im Text; im Einreichungspaket liegen dieselben Abbildungen zusätzlich als SVG-Dateien für die getrennte Einreichung bei. Der Text ist ein Entwurf: Die Autor:innen redigieren ihn, bevor er eingereicht wird.