MASTERARBEIT | MASTER'S THESIS

Titel | Title

Methodological Quality of Meta-Analytic Systematic Reviews in Clinical Child and Adolescent Psychology (2022-2024) – A Meta-Review

verfasst von | submitted by

Julia Maria Riedel BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien | Vienna, 2026

Studienkennzahl lt. Studienblatt | Degree programme code as it appears on the student record sheet:	UA 066 840
Studienrichtung lt. Studienblatt | Degree programme as it appears on the student record sheet:	Masterstudium Psychologie
Betreut von | Supervisor:	Univ.-Prof. Dr. Martina Zemp
Mitbetreut von | Co-Supervisor:	Dr. Magdalena Siegel BSc MSc

Table of Contents

Introduction

Meta-analytic systematic reviews occupy a central position in evidence-based clinical decision making in Clinical Child and Adolescent Psychology (CCAP), as they synthesize large bodies of empirical research into actionable conclusions (Bellato et al., 2023; Borenstein et al., 2021; Clarke & Chalmers, 2018; Murad et al., 2016; Weisz et al., 2023). However, the evidentiary strength of such reviews is not inherent to their design but depends on the methodological safeguards implemented during planning, execution, and analysis, as well as transparent reporting and statistical consistency (Borenstein et al., 2021; Cooper et al., 2019; Guyatt et al., 2008; Guyatt et al., 2011; Møller et al., 2018; Shea et al., 2017; Vazire et al., 2022). The present meta-review is the first large-scale investigation of the methodological quality of recent high-impact meta-analytic systematic reviews in the field of CCAP. Moreover, it is the first to explore the nuances of critical aspects of methodological quality in relation to relevant aspects of reporting transparency and to examine its association with statistical consistency. To situate this investigation, the present introduction first outlines the scope and relevance of CCAP and explains the role of meta-analytic systematic reviews within this discipline. It then discusses scientific credibility in light of ongoing concerns about replicability and reproducibility in psychological research, highlighting how these issues extend to research synthesis. Against this backdrop, three core concepts relevant to the trustworthiness of meta-analytic systematic reviews are introduced: methodological quality, its overlap with reporting transparency, and statistical consistency. Finally, the objectives and research questions guiding this meta-review are presented.

Clinical Child and Adolescent Psychology

Clinical Child and Adolescent Psychology (CCAP) addresses mental health conditions that commonly emerge early in the life course and often persist into adulthood (De Girolamo et al., 2012; Kessler et al., 2005; Solmi et al., 2022; World Health Organization [WHO], 2003; Polanczyk et al., 2015). As a subdiscipline of psychology, CCAP focuses on understanding the onset (Solmi et al., 2022), course, and treatment of mental disorders in young people (Dekkers et al., 2023; Weisz et al., 2023), as well as on developing and evaluating interventions that promote psychological well-being (Brown et al., 2023; Finch, 2012; Matson, 2023).

Epidemiological evidence indicates that a substantial proportion of mental disorders begin during childhood or adolescence, with early onset frequently associated with delayed intervention and less favorable long-term outcomes (De Girolamo et al., 2012; Kessler et al., 2005; Solmi et al., 2022; WHO, 2003; Polanczyk et al., 2015). Current prevalence estimates underscore the scale of this challenge, with mental disorders affecting a considerable proportion of young people worldwide and suicide ranking among the leading causes of death in late adolescence and early adulthood (Boonyasai et al., 2022; Sacco et al., 2024; WHO, 2024a; 2024b). These developmental trajectories underscore the importance of high-quality evidence to guide prevention, diagnosis, and intervention efforts in CCAP. Given the rapidly expanding volume of primary research, clinicians and researchers increasingly depend on meta-analytic systematic reviews to integrate findings and inform practice (Ioannidis, 2016).

The Role of Meta-Analytic Systematic Reviews in Clinical Child and Adolescent Psychology

Delivering evidence-based clinical care requires credible research and high-quality syntheses (Clarke & Chalmers, 2018). Meta-analytic systematic reviews play a key role in interpreting novel study results within the context of existing knowledge (Draborg et al., 2022; Lund et al., 2016). By systematically synthesizing, critically evaluating, and statistically summarizing results from multiple primary studies, these reviews serve as a lens for assessing the evidence and certainty of research (Borenstein et al., 2021; Cooper et al., 2019; Murad et al., 2016; Siddaway et al., 2019). Making these reviews vital for advancing knowledge and practice in CCAP (Bellato et al., 2023; Weisz et al., 2023). When conducted rigorously, they can clarify inconsistencies in the literature, estimate overall effect sizes, explore sources of heterogeneity, and identify gaps requiring further investigation (Møller et al., 2018; Weisz et al., 2023). For these reasons, such reviews are often regarded as providing the highest level of evidence in clinical research (Murad et al., 2016), helping to advance scientific theory, inform clinical practice, and influence policy (Clarke & Chalmers, 2018; Strawbridge et al., 2025). Taken together, this makes it essential that findings of meta-analytic systematic reviews are trustworthy.

The Status of Meta-Analytic Systematic Reviews in Light of the Replication Crisis

However, despite their influence and prestige, especially when published in reputable journals, the assumption that meta-analytic systematic reviews are inherently trustworthy warrants critical scrutiny (Clarke & Chalmers, 2018; Ioannidis, 2016). Since the early 2010s, large-scale replication efforts and methodological critiques have revealed substantial rates of non-reproducible findings (Bakker et al., 2012; McGee & Dawson, 2020; Munafò et al., 2017; Open Science Collaboration, 2015; Simmons et al., 2011). In response, psychological science has engaged in sustained critical discourse on the credibility, replicability, and reproducibility of its evidence base (Hardwicke et al., 2014; Klein et al., 2018; Pennington, 2023; Stevens, 2017). These failures were widely attributed to shortcomings in research and publishing practices, including publication bias, questionable research practices, and limited transparency, which contributed to an elevated rate of false-positive findings (Korbmacher et al., 2023) as well as undermined public trust in scientific results (Hendriks et al., 2020). This discussion became known as the *replication crisis* (Świątkowski & Dompnier, 2017).

Although initially centered on primary studies, the crisis necessarily extends to meta-analytic systematic reviews (Polanin et al., 2020). Because weak methodology, publication bias, and insufficient reporting of the included primary research increase the likelihood of introducing biased and spurious conclusions in the review, thereby further distorting the scientific record (Muka et al., 2020; Munafò et al., 2017; Polanin et al., 2020; Świątkowski & Dompnier, 2017). Meta-analyses are often used to clarify effects when findings across studies are inconsistent and thus provide a “clearer picture” of the evidence base (Świątkowski & Dompnier, 2017). However, if negative or null results remain systematically unpublished, the resulting “clearer image” will inevitably be distorted (Świątkowski & Dompnier, 2017). Consequently, meta-analyses remain vulnerable to publication bias, variable study quality, and misleading precision when synthesizing heterogeneous or flawed studies (Carter et al., 2019; Nosek et al., 2022; Rothstein et al., 2005; Vosgerau et al., 2019). Because aggregated findings can create an appearance of robustness, and meta-analyses are often treated as authoritative evidence (Carlson & Ji, 2011), ensuring that their findings are reliable and methodologically rigorous is especially critical (Higgins et al., 2009; Moreau & Corballis, 2019). Yet, despite their status as

the strongest form of evidence for clinical decision-making, many practitioners are not fully trained in evaluating them (Garritty et al., 2019; OCEBM Levels of Evidence Working Group, 2011). At the same time, their rapid proliferation, driven partly by their ease of production and academic incentives, has reached “epidemic” levels according to Ioannidis (2016), resulting in many overlapping, suboptimal, or contradictory reviews (Ebrahim et al., 2015; Ioannidis, 2016).

Thus, a meta-analysis or systematic review label does not guarantee high-quality evidence (Blümle et al., 2014; Chapelle et al., 2023; Ioannidis, 2016). Confidence in synthesized evidence requires careful assessment of study limitations, consistency of results, precision of effect estimates, and potential publication bias (Borenstein et al., 2021; Cooper et al., 2019; Guyatt et al., 2008; Guyatt et al., 2011). Major threats to their validity include poor reporting transparency, inadequate methodological quality, unrecognized statistical heterogeneity, and failure to assess overall evidence quality (Møller et al., 2018). Consequently, the evidentiary strength of many (meta-analytic) systematic reviews remains poor due to publication bias, failure to replicate, and inconsistency (Sotola, 2022), with researchers calling into question their usefulness (Chevret et al., 2018; Møller et al., 2018).

Ultimately, the conclusions of meta-analytic systematic reviews depend heavily on methodological choices regarding eligibility criteria, statistical aggregation methods, and bias assessment procedures (Møller et al., 2018). Evaluating such reviews from a methodological perspective, therefore, highlights the importance of systematically evaluating review quality, rather than assuming credibility based on scope, citation impact, or journal prestige (McKiernan et al., 2019; Świątkowski & Dompnier, 2017). Given the rapid growth of scientific publications and frequent inconsistencies across studies, the need for reliable, high-quality evidence synthesis to inform research and public health is increasingly urgent (Lozada-Martinez et al., 2025; Luchini et al., 2021).

Definition of Scientific Credibility

To further frame this debate, key concepts must be distinguished: trust, trustworthiness, reproducibility, replicability, and scientific credibility. Trust is the assumption that research was conducted with sound methods and reliable data, whereas trustworthiness implies that a finding *deserves* trust because it is founded

in high-quality research (Peels & Bouter, 2021). For (meta-analytic) systematic reviews, trustworthiness requires several methodological standards, including a scientifically as well as clinically meaningful question; compliance with established reporting guidelines; a preregistered, ideally published protocol with a predefined statistical analysis; an up-to-date, systematic and reproducible search strategy; explicit inclusion and exclusion criteria; advance consideration of methodological, statistical, and clinical heterogeneity; thorough evaluation as well as clear description of risk of bias (RoB) and random error; appropriate statistical analyses; rating of evidence quality; justified claims; and absence of conflicts of interests (Møller, et al. 2018).

For meta-analytic systematic reviews, credibility is closely tied to both reproducibility and replicability (Munafò et al., 2017; Nosek et al., 2022). Reproducibility refers to obtaining the same results when reanalyzing original data using the same analysis plan, while replicability involves collecting new data and applying the same analysis plan to determine whether the findings hold (Korbmacher et al., 2023; Peels & Bouter, 2021). In meta-analytic systematic reviews, reproducibility involves verifying results by implementing the same search strategies, input data, code, and analytic procedures (Maassen et al., 2020; Nguyen et al., 2025); replicability involves updating searches or reanalyzing data with novel statistical synthesis methods (Epskamp, 2019). Failures at either level can compromise the trustworthiness of synthesized evidence (Nosek et al., 2022).

High-quality (meta-analytic) systematic reviews are essential, yet the concept of ‘scientific credibility’ or ‘research quality’ remains difficult to define. In practice, partly due to academic career decisions such as funding allocations and recruitment processes, research quality is sometimes inferred from the prestige of the publishing outlet, based on the assumption that more prestigious journals publish higher-quality work (Świątkowski & Dompnier, 2017). However, the primary aim of empirical science is to generate knowledge through rigorous methodology; accordingly, research quality should be assessed based on methodological standards rather than journal reputation (Świątkowski & Dompnier, 2017). From a methodological perspective, then, research quality can be understood as the validity of claims and their contribution to cumulative knowledge (Świątkowski & Dompnier, 2017). It further depends on replicability and on the overall quality and

transparency of research practices (Ioannidis, 2012; Klein et al., 2018; Vazire et al., 2022). Improving credibility, therefore, requires continually improving methods and evaluative processes (Ioannidis et al., 2015), as well as systematic assessments capable of identifying weaknesses in evidence and enhancing the trustworthiness of findings (Ioannidis, 2010; 2012). Common criteria include aspects of internal, external, and construct validity, as well as imprecision, inconsistency, publication bias, and reporting completeness (Balshem et al., 2011; Brewer & Crano, 2014; Furuya-Kanamori et al., 2021; Lunny et al., 2025). This meta-review will focus on three core dimensions: methodological quality, reporting transparency, and statistical consistency, which are defined in the following sections.

Meta-Science and the Credibility Revolution

While the term *replication crisis* highlights the seriousness of these issues, particularly declining public trust in science and weakened support for evidence-based policy (Hendriks et al., 2020; Korbmacher et al., 2023), many now frame the response as a "credibility revolution" (Vazire, 2018). This reframing reflects a collective effort to strengthen scientific integrity through more rigorous, transparent, and high-quality practices (Korbmacher et al., 2023; Vazire et al., 2022). These developments have also solidified "meta-science" (or "meta-research") as a distinct field that turns the lens back onto science itself, focusing on evaluating and improving scientific practices (Ioannidis, 2018; Ioannidis et al., 2015).

Meta-science examines the status of psychological research and the transparency, integrity, and potential biases shaping its evidence (Cristea & Naudet, 2019). It provides high-level analyses aimed at identifying and prioritizing rigorous, impactful studies (Burns et al., 2011; Luchini et al., 2021; McGee & Dawson, 2020). Its scope includes methodological practices, reporting standards, reproducibility, and incentive structures (Fabiano et al., 2024; Goodman & Dickersin, 2011; Ioannidis, 2018). By analyzing how studies are designed, conducted, reported, and evaluated, and whose perspectives are included, meta-science seeks to identify factors that strengthen or undermine methodological rigor, credibility, transparency, reproducibility, and equity in research (Fabiano et al., 2024; Ghai et al., 2025; Hardwicke et al., 2020). In essence, it addresses how science should be performed, communicated, corrected, and rewarded (Ioannidis et al., 2015).

Meta-science also aims to provide corrective mechanisms that operate faster than the traditional, slow self-correcting process of science, promoting robust practices and reducing research waste (Ioannidis, 2018). Wasteful practices may include unnecessary duplication of studies, lack of preregistered protocols, and failure to share data (Cristea & Naudet, 2019). In contrast, efficient practices emphasize reporting transparency, comprehensive documentation of all research stages, alongside data sharing, checks for reproducibility, and objective replication efforts, particularly for results with clinical relevance (Cristea & Naudet, 2019). A key tool in this effort is the meta-review, a methodological review of reviews, which evaluates aspects of scientific trustworthiness such as transparency, reproducibility, or methodological quality of meta-analytic systematic reviews within a specific field (Biondi-Zoccai, 2016; López-Nicolàs et al., 2022; Weissgerber, 2021). Evaluating a (meta-analytic) systematic review requires assessing the exhaustiveness of the literature search, the reproducibility of study selection, and the clarity of reporting, as well as judging the quality of evidence at the outcome level: assessing RoB, precision, statistical consistency, and reporting bias (Murad et al., 2014).

Meta-research has repeatedly raised concerns about the methodological quality of (meta-analytic) systematic reviews often being at a critically low level (Bojcic et al., 2024; Leclercq et al., 2020a; Lorenz et al., 2020; Matthias et al., 2020), gaps in reporting transparency which hinder reproducibility (Hohn et al., 2020; Lakens et al., 2017; López-Nicolás et al., 2022; Maassen et al., 2020; Polanin et al., 2020; Steil et al., 2022), and persisting statistical inconsistencies (Nuijten & Polanin, 2020; Polanin & Nuijten, 2018). While trust is high in meta-analytic systematic reviews (Clarke & Chalmers, 2018; Ioannidis, 2016; Murad et al., 2016), the aforementioned shortcomings call into question the trustworthiness of meta-analytic reviews as a reliable synthesis tool. Despite these high stakes, no larger-scale systematic investigation has assessed the methodological quality, reporting transparency, and statistical consistency of meta-analytic systematic reviews in leading CCAP journals. Addressing this gap is crucial for strengthening trust in the evidence base of this field. Moreover, to investigate the rigorous standards required to support their role in research-based decision-making (Clarke & Chalmers, 2018).

The next section elaborates on three core components: methodological quality, reporting transparency, and statistical consistency.

Definition of Methodological Quality

From a meta-scientific perspective, methodological quality is a key indicator of the trustworthiness of meta-analytic systematic reviews. Sound methods help prevent bias and are central to a review's overall quality and the confidence that can be placed in its findings (Blümle et al., 2014; Pieper et al., 2018a). Broadly, methodological quality refers to how well a review is designed and conducted, and the extent to which its procedures are likely to yield accurate, unbiased findings consistent with best practice (Lunny et al., 2025; Murad et al., 2014; Reitsma et al., 2009; Strawbridge et al., 2025). In essence, it reflects whether appropriate safeguards were in place to minimize systematic error (Faggion, 2023). For (meta-analytic) systematic reviews, methodological quality also encompasses adherence to established standards for conduct and reporting, such as the Cochrane Handbook (Higgins et al., 2024) and PRISMA 2020 (Page et al., 2021a; Gates et al., 2018). Against this background, the present thesis examines the extent to which methodological quality overlaps with, or encompasses, aspects of reporting transparency.

Although the methodological quality of a (meta-analytic) systematic review is often equated with its RoB, the two are conceptually distinct (Gates et al., 2018; Lunny et al., 2025). While both relate to elements of internal validity, and quality assessments frequently include RoB assessments (Lunny et al., 2024), methodological quality has a broader remit (Gates et al., 2018; Lunny et al., 2024; Lunny et al., 2025; Reitsma et al., 2009; Stone et al., 2024). It is also important to differentiate between the RoB of a systematic review itself, sometimes termed "meta-bias", and the bias in its included primary research (Whiting et al., 2015). RoB of a systematic review focuses on the probability of errors in the review's design, conduct, or analysis that may have skewed the review's findings or produced misleading claims (Lunny et al., 2025; Whiting et al., 2015), whereas methodological quality evaluates whether the review was planned, executed, analysed, interpreted, and reported in ways that aim to prevent bias (Furuya-Kanamori et al., 2021; Lunny et al., 2025).

Importantly, a low-quality review does not necessarily imply a high RoB. Some quality indicators (e.g., reporting reasons for study exclusion) influence interpretability and trust without directly biasing results (Gates et al., 2018; Viswanathan et al., 2017). Moreover, assessing methodological quality generally checks whether such safeguards are present, not whether they ultimately prevented bias (Faggion, 2023; Shea et al., 2017). Conversely, even rigorously conducted reviews may still be subject to biases, such as selection bias (Muka et al., 2020; Nordström et al., 2023). Therefore, dedicated tools such as the Risk of Bias in Systematic Reviews (ROBIS; Whiting et al., 2015) are required to evaluate the RoB in systematic reviews (Lunny et al., 2024; 2025), focusing on whether review methods adequately reduced bias estimates and claims (Whiting et al., 2015). In overviews of systematic reviews, evaluating both methodological quality and RoB of the included (meta-analytic) systematic reviews is necessary before review findings are used in evidence-based practice (Furuya-Kanamori et al., 2021; Lunny et al., 2024).

Assessment of Methodological Quality

Since the 1980s, various tools have been developed to assess the methodological quality of systematic reviews, beginning with the Overview Quality Assessment Questionnaire (OQAQ; Oxman & Guyatt, 1991). This was followed by A MeaSurement Tool to Assess Systematic Reviews (AMSTAR), which addressed limitations of OQAQ and adopted a clinically focused approach (Shea et al., 2017; Strawbridge et al., 2025). Subsequent developments included R-AMSTAR, which introduced scoring (Kung et al., 2010), and AMSTAR 2, which expanded coverage to reviews of both randomized and non-randomized studies (Pieper et al., 2018a; Shea et al., 2017).

Although several tools are available (Whiting et al., 2015), AMSTAR 2 is currently the preferred tool in psychology for assessing methodological quality (Kolaski et al., 2023; Pollock et al., 2020; Lunny et al., 2024). AMSTAR 2 outlines key domains of high-quality reviews, including preregistration of the review protocol; clear eligibility criteria, rigorous searches of the literature, data extraction, and statistical synthesis procedures; justification of excluded studies; assessment of RoB and publication bias; and disclosure of conflicts of interest in both primary studies and the review itself (Shea et al., 2017). AMSTAR 2 comprises 16 items

and provides a quality rating of either *High*, *Moderate*, *Low*, or *Critically Low* (Shea et al., 2017). Three items focus on the conduct of the review's RoB assessment of its included primary studies (Kolaski et al., 2024b). Compared with its predecessor AMSTAR, it is more detailed and user-friendly, with items covering all stages of the review process (Perry et al., 2021; Pieper et al., 2018a). By providing structured guidance to identify strengths and weaknesses in methodology as well as potential sources of bias, AMSTAR 2 aims to enhance the credibility and reliability of evidence syntheses (Lunny et al., 2024).

Methodological Quality of Meta-Analytic Systematic Reviews in CCAP

Across several psychological research fields, methodological resources that could enhance credibility are not consistently implemented (Leclercq et al., 2020a). Substantial variability in methodological quality has led to concerns that heterogeneity in findings may partly reflect low study quality (Botella & Durán, 2019). Several overviews of reviews have examined methodological quality across CCAP-related domains, including interventions for depressive symptoms (Espada et al., 2023; Zbukvic et al., 2024), anxiety (Byrne et al., 2023; Galán-Luque et al., 2023; Zbukvic et al., 2024), and disordered eating (García-Fernández et al., 2023); responses to sexual abuse (Sánchez De Ribera et al., 2020); conduct issues (Romero et al., 2023); psychological distress among sexual and gender minority youth (O'Shea et al., 2025); psychological effects of cyberbullying (Kwan et al., 2020); and the effect of fathers' mental illness on child development (Scarlett et al., 2023).

As measured via AMSTAR 2 (Shea et al., 2017), most of these overviews reported *Low* to *Critically Low* methodological quality (Byrne et al., 2023; Cortese et al., 2019; Kwan et al., 2020; Romero et al., 2023; Sánchez De Ribera et al., 2020), while others even found *Low* to *Critically Low* quality across all included reviews (Espada et al., 2023; Galán-Luque et al., 2023; García-Fernández et al., 2023; O'Shea et al., 2025; Scarlett et al., 2023). The exception is one study (Zbukvic et al., 2024) that found *Moderate* to *High* quality in the examined systematic reviews. Key methodological shortcomings include failure to justify study exclusions, lack of preregistration, inadequate assessment of RoB and publication bias, and insufficient reporting of funding sources for primary studies (Byrne et al., 2023; Cortese et al., 2019; Espada et al., 2023; Galán-Luque et al., 2023; García-

Fernández et al., 2023; Kwan et al., 2020; O'Shea et al., 2025; Romero et al., 2023; Sala et al., 2024; Sánchez De Ribera et al., 2020; Scarlett et al., 2023).

Overall, these findings indicate substantial methodological weaknesses in (meta-analytic) systematic reviews across several CCAP domains. Notably, most overviews included fewer than 15 systematic reviews, highlighting the need for larger-scale investigations and clearer guidance on improving methodological quality in this field (Siegel et al., 2025). This need is particularly pressing, given evidence that studies published in high-impact scientific journals, contrary to the common assumption, may not necessarily be more methodologically sound and may even be less so than those published elsewhere (Brembs, 2018).

Reporting Quality and Methodological Quality

Methodological quality assessment depends, in part, on reporting completeness (Matthias et al., 2020), as thorough reporting enables sufficient critical appraisal (Lorenz et al., 2020). Accordingly, reporting transparency is crucial for evaluating the trustworthiness of meta-analytic systematic reviews in CCAP. It entails clear documentation of all theoretical, methodological, and analytical decisions, thereby preventing undisclosed selective reporting (Blümle et al., 2014; Korbmacher et al., 2023; Vazire, 2018).

Although closely related, reporting transparency and methodological quality are distinct and should thus be assessed separately (Blümle et al., 2014; Lunny et al., 2025). A review may be methodologically rigorous but poorly reported if key details (e.g., reporting reasons for study exclusions) are omitted (Blümle et al., 2014; Gates et al., 2018; Viswanathan et al., 2017). Conversely, a methodologically weak review may be transparently reported and fully disclose its limitations (Blümle et al., 2014), but its conclusions would be flawed. Without adequate reporting, it is impossible to distinguish between well-conducted but incompletely reported reviews and poorly conducted yet thoroughly reported ones, creating uncertainty about review quality (Vazire, 2017). Thus, adequate transparency can increase confidence by enabling error detection, evaluation of analytic decisions, and scrutiny of results (Vazire, 2017), but it does not guarantee methodological rigor - and vice versa (Sandoval-Lentisco et al., 2024).

Assessment of Reporting Transparency

Reporting guidelines play a critical role in improving transparency and completeness of reporting, thereby facilitating more accurate assessments of methodological quality (Blümle et al., 2014; Lorenz et al., 2020). The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021a) are the recommended tool for assessing reporting transparency (Kolaski et al., 2023). PRISMA emphasizes detailed documentation of methods and results to ensure trustworthiness and applicability (Page et al., 2021a). Several key PRISMA domains align with AMSTAR 2 critical domains, including protocol preregistration, transparent literature search procedures, clear documentation of study exclusions, specification of meta-analytic methods, and assessment of RoB and publication bias (Lunny et al., 2024; Page et al., 2021a; Page et al., 2021b; Shea et al., 2017; see Section RQ3: Transparency Checklist and Methodological Quality for more detailed examples). This overlap reflects shared concerns but does not imply equivalence, as reporting guidelines are not designed to evaluate methodological quality per se (Blümle et al., 2014). PRISMA focuses on what is reported, not methodological rigor (Kolaski et al., 2023). For instance, while it recommends reporting how many authors conducted study screening and coding (Page et al., 2021a), it does not require double coding or specify interrater agreement thresholds, core methodological safeguards assessed by AMSTAR 2 (Shea et al., 2017). Nonetheless, PRISMA scores significantly correlate with AMSTAR 2 scores (r = .56; Peña et al., 2022), supporting further investigation into transparency domains relevant to methodological quality, as they are essential for reproducibility and bias-free validation of results (O'Kelly et al., 2020). This has led to increasing calls for their joint assessment (Baveye, 2025).

Reporting Transparency and Methodological Quality in Meta-Analytic Systematic Reviews in Social Science and Psychology

Ideally, (meta-analytic) systematic reviews would demonstrate both high methodological quality and high reporting transparency. Meta-scientific evidence suggests that these dimensions are positively related. For example, a large-scale meta-review in Chinese social sciences found that more complete reporting was associated with greater methodological quality (Guo et al., 2025). Similarly, a meta-epidemiological study of psychological meta-analyses showed that reviews explicitly stating their use of PRISMA achieved higher AMSTAR 2 scores and were

more likely to report protocols, assess RoB, and incorporate these assessments into their conclusions (Leclercq et al., 2020a). Consistent findings were also reported in clinical meta-scientific research, where reporting deficiencies and methodological weaknesses systematically co-occurred in domains such as RoB assessment, protocol preregistration, and inadequate reporting of publication bias and search strategy (Qi et al., 2024). At the same time, transparency and methodological quality are not interchangeable. Meta-research on family-centered interventions for perinatal depression showed that, while reviews were partially transparently reported, they were methodologically weak, resulting in unreliable or misleading conclusions (He et al., 2023).

Taken together, these findings suggest that examining such relevant domains of reporting transparency on which an assessment of methodological quality is dependent (Lorenz et al., 2020) may provide additional nuance beyond overall AMSTAR 2 ratings. Specifically, PRISMA-based domains may help identify *where* and *how* methodological weaknesses manifest within AMSTAR 2 critical domains, thereby offering a more fine-grained understanding of review quality. Investigating this relationship is particularly relevant in the context of contemporary meta-analytic systematic reviews published in leading CCAP journals, where no comprehensive meta-scientific assessment currently exists. Therefore, the present study aims to explore the extent to which transparency-related domains can be used to capture additional variability in methodological quality within AMSTAR 2 critical domains.

Definition of Statistical Consistency

Statistical conclusion validity is another key indicator of the trustworthiness of meta-analytic systematic reviews (Vazire et al., 2022). Accordingly, statistical consistency refers to the coherence and accuracy of statistical reporting within null hypothesis significance testing (NHST; Bakker & Wicherts, 2011; Nuijten et al., 2016), in such that test statistics (e.g., t, F, χ^2), their degrees of freedom, as well as the corresponding p-values should align (Nuijten & Polanin, 2020). Conversely, statistical inconsistencies may produce incorrect conclusions about the (non)-existence of effects, hinder reproducibility, and compromise accurate effect size calculations (Nuijten et al., 2016; Nuijten & Polanin, 2020; Polanin & Nuijten, 2018).

Although effect sizes are the primary focus in meta-analyses, NHST remains integral; for example, it is frequently used to test for moderator effects and statistical heterogeneity or in evaluating publication bias (Borenstein et al., 2021; Nuijten & Polanin, 2020; Polanin & Nuijten, 2018; Polanin & Pigott, 2015). Because methodological quality assessments of meta-analytic systematic reviews include evaluation of publication bias and heterogeneity (Shea et al., 2017), the continued reliance on NHST within meta-analyses warrants careful consideration of their results (Nuijten et al., 2016). In principle, experienced researchers conducting rigorous analyses should rarely yield statistical inconsistencies (Francis, 2013; Olsen et al., 2019). Indeed, such inconsistencies are more common in low-impact than in high-impact journals (Bakker & Wicherts, 2011). As methodologically rigorous reviews apply procedures intended to yield accurate and unbiased findings (Lunny et al., 2025; Murad et al., 2014; Reitsma et al., 2009; Strawbridge et al., 2025), inconsistencies typically signal errors in data collection, analysis, or reporting (Francis, 2013). This issue is especially salient in cases of publication bias or selective reporting, which can inflate the apparent significance of findings without strengthening true evidential value (Francis, 2013). As both statistical consistency and methodological quality contribute to overall research quality (Dougherty & Horne, 2022; Nuijten et al., 2016; Pieper et al., 2018a), they may be expected to co-occur. Examining this relationship can clarify whether methodological rigor corresponds to accurate statistical reporting in high-impact CCAP meta-analytic systematic reviews.

Assessment of Statistical Consistency

To detect statistical inconsistencies, tools like *statcheck* automatically scan academic articles for errors in reported NHST results (Nuijten & Polanin, 2020). The American Psychological Association (APA) recommends a standardized format for reporting NHSTs: for example, $t(28) = 2.14$, $p = .04$ (APA, 2020; Nuijten & Polanin, 2020). Because the test statistics, degrees of freedom, and *p*-value are mathematically connected, knowing any two allows computation of the third. *Statcheck* identifies reported NHST results in articles, recalculates them, and flags any mismatches, deeming them as either inconsistent or grossly inconsistent (Nuijten & Polanin, 2020). An inconsistency refers to a situation in which the reported *p*-value and the recalculated *p*-value are not the same, but the difference

is not large enough to alter the overall interpretation. Gross inconsistencies arise when the recalculated *p*-value suggests a different conclusion, such as a previously significant result becoming non-significant or vice versa (Nuijten & Polanin, 2020).

Statistical Consistency of Meta-Analytic Systematic Reviews in Social Science

Meta-research on statistical consistency in the social sciences has found that 39% of meta-analyses contain at least one inconsistency, defined as a discrepancy between reported and recalculated *p*-values (Polanin & Nuijten, 2018). Alarmingly, 8% of meta-analyses include gross inconsistencies where drawn conclusions may need revision due to changes in statistical (non)significance (Polanin & Nuijten, 2018). Despite the importance of accurate reporting, statistical consistency in CCAP meta-analytic systematic reviews published in high-impact journals has not yet been systematically investigated. Exploring its relationship with methodological quality can help determine whether accurate statistical reporting is associated with higher-quality meta-analytic systematic reviews, an important question given that both components are essential for producing trustworthy findings in CCAP.

Study Objective

Taken together, the preceding sections establish four points. First, methodological quality, reporting transparency, and statistical consistency are key pillars of scientific credibility. Second, meta-analytic systematic reviews occupy a central role in evidence-based decision-making in CCAP. Third, meta-scientific research demonstrates that neither methodological quality nor reporting transparency can be assumed, even in high-impact outlets. Fourth, existing meta-scientific investigations in CCAP remain fragmented, small in scale, or limited to single dimensions of credibility. The present meta-review, as part of a larger investigation into these three pillars within meta-analytic systematic reviews in CCAP (Siegel et al., 2025), addresses these gaps by systematically examining methodological quality, its transparency-related nuances, and its association with statistical consistency in leading CCAP meta-analytic systematic reviews from 2022 to 2024. For this, the current state of methodological quality will be examined by providing a cross-sectional overview of meta-analytic systematic reviews published in leading scientific journals in the field of CCAP from 2022 to 2024. Second, additional nuances of methodological quality will be highlighted by narratively

assessing relevant and related aspects of reporting transparency. Third, methodological quality and its association with statistical consistency will be examined. This assessment may guide applied meta-analysts and inform future research aimed at improving these core aspects, possibly strengthening the trustworthiness of meta-analytic systematic reviews as a research synthesis tool in CCAP.

Research Questions

The following research questions are addressed within this thesis:

RQ1: Overall Methodological Quality

How high is the overall methodological quality of current meta-analytic systematic reviews (2022–2024) published in leading scientific journals in the field of CCAP?

RQ 2: Methodological Quality Within AMSTAR 2 Critical Domains

How high is the methodological quality of current meta-analytic systematic reviews (2022-2024) published in leading scientific journals in the field of CCAP within the critical domains of AMSTAR 2?

RQ 3: Transparency-Related Domains as Sources of Additional Nuance in Methodological Quality Within AMSTAR 2 Critical Domains

To what extent can relevant transparency domains (based on PRISMA) be used to identify additional nuances in methodological quality within the critical domains of AMSTAR 2 in current meta-analytic systematic reviews (2022-2024) published in leading scientific journals in the field of CCAP?

RQ 4: Association Between Methodological Quality and Statistical Consistency

What is the association between methodological quality and statistical consistency of current meta-analytic systematic reviews (2022-2024) published in leading scientific journals in the field of CCAP?

Methods

Statements and Declarations

This meta-review was carried out as part of a larger meta-scientific research project on meta-analytic systematic reviews in the field of CCAP (for the study protocol, see Siegel et al., 2025). As part of the overarching project, one adjacent master's thesis was already completed by Selina Fanninger (title: "Open Science Practice Adoption in Clinical Child and Adolescent Psychology: A Meta-Review of Meta-Analytic Systematic Reviews (2022–2024)") and two more are currently being conducted by Hannah Maria Moser (working title: "Transparenz in metaanalytischen systematischen Reviews der Klinischen Psychologie des Kindes- und Jugendalters (2022– 2024): Ein Meta-Review zu allgemeinen sowie kind- und jugendzentrierten Berichtsaspekten [Transparency in meta-analytic systemic reviews in clinical child and adolescent psychology (2022–2024): A meta-review of general and child- and adolescent-centered aspects of reporting]"), and Marie-Charlotte Ulrich (working title: "Transparency of Intervention Reviews in Clinical Child and Adolescent Psychology: A Meta-Review of Meta-Analytic Systematic Reviews"). For this reason, the methodological description in this master's thesis is likely to overlap with these theses and the study protocol of the larger meta-scientific research project (Siegel et al., 2025). Critically, neither the research questions nor the resulting analyses overlap between these theses.

The overarching research project was preregistered on November 2, 2024, before the systematic literature search; the protocol can be accessed via https://osf.io/ah5m4, with the final study protocol available via https://doi.org/10.3389/fpsyg.2025.1535606. The present master's thesis was not preregistered separately, but the protocol was submitted as part of the course requirements for the master's seminar and is available upon request by contacting the author via email (juliariedel2@gmail.com). As no distinct reporting guidelines exist for methodological meta-reviews (Dal Santo et al., 2023), relevant aspects of the PRIOR (Gates et al., 2022; see Appendix B), PRISMA (Page et al., 2021a), PRISMA-P (Moher et al., 2015), and PRISMA S (Rethlefsen et al., 2021) guidelines were implemented. Materials involving screening, coding, and analysis (i.e., scripts and data) were shared within the research team, and some are included in the Appendix (e.g., study characteristics in Appendix H). Additional materials, data, and

information are available upon request by contacting Julia Riedel (juliariedel2@gmail.com), Magdalena Siegel (magdalena.siegel@univie.ac.at), or Martina Zemp (martina.zemp@univie.ac.at). This research project obtained no financial compensation, and the author declares no competing interests. All sections of this thesis: introduction, methods, results, and discussion were written by the author, with ChatGPT (version 4, version 5; OpenAI, 2024; 2025) assisting in improving clarity, flow, and brevity, as recommended for non-native English speakers (Nakagawa & Lagisz, 2024). Moreover, ChatGPT (version 5; OpenAI, 2025) aided in writing and revising R code. All AI-generated suggestions were examined by the author, who takes full responsibility for the content of this master's thesis as well as all supplementary information.

Deviations from Protocol

Coding Procedures. To ensure consistency with the thesis's scope and to account for the variety of meta-analytic systematic reviews included, several minor adaptations and clarifications were made to the AMSTAR 2 checklist (Shea et al., 2017). This involved item 2, in which protocols were not compared with published articles, as this fell outside the scope of the study. Furthermore, item 3, regarding the rationale for choosing the included study designs, was adjusted for non-intervention meta-analytic systematic reviews, such that specifying eligible study types was considered sufficient. Item 4, concerning searches within trials or registries, was coded only for meta-analytic systematic reviews involving intervention studies. For item 9, concerning RoB assessment in primary studies, it was deemed acceptable for non-intervention reviews when authors used either a standardized tool or a well-reasoned, self-developed approach. For intervention-focused reviews, the original AMSTAR 2 criteria for RoB assessment in primary studies remained unchanged. Regarding item 15, graphical assessments of publication bias alone were considered insufficient, as this has been noted to be prone to subjectivity bias (Siegel et al., 2022); statistical testing was required. These refinements did not alter the overall structure or intent of the AMSTAR 2 tool and were developed consensually by the research team.

Statistical Analysis. The initial protocol for this thesis specified that overall methodological quality would be assessed using the AMSTAR 2 confidence rating, categorized as high, moderate, low, or critically low. This variable was then to be

dichotomized into high or moderate (coded as 1) vs. low or critically low (coded as 0) for the dichotomized approach of RQ4. However, this analysis was not carried out because too few meta-analytic systematic reviews were rated as high or moderate, and chi-square or odds ratio analyses were therefore not applicable. Consequently, only the continuous approach, using *t*-tests and effect sizes reported as Hedges' *g*, was implemented to examine associations with statistical consistency. For this purpose, the *statcheck* output was planned to be dichotomized into reports classified as either containing (1) or not containing (0) any inconsistencies or gross inconsistencies. However, no gross statistical inconsistencies were detected in the dataset. Thus, the *statcheck* output was contrasted in meta-analytic systematic reviews flagged vs. not flagged for any statistical inconsistency. This analysis was not specified in the protocol.

Software and Computational Tools. Several R packages not mentioned in the preregistration were used for data processing, statistical analysis, and visualization; their use is detailed in the "Data Analysis" section. ChatGPT (version 5; OpenAI, 2025) supported code development and further research. This support was restricted to implementation assistance and did not affect any analytic choices. All AI-generated content was reviewed and validated by the author for accuracy and suitability.

Inclusion and Exclusion Criteria

Publication type and language

For additional details regarding eligibility criteria, see the protocol of the overarching project via https://doi.org/10.3389/fpsyg.2025.1535606. Because no universally accepted definition of (meta-analytic) systematic reviews exists (Faggion & Diaz, 2019), study and publication types were defined according to current PRISMA guidelines. PRISMA describes a systematic review as "a review that uses explicit, systematic methods to collate and synthesize findings of studies that address a clearly formulated research question," and a meta-analysis as "a statistical technique used to synthesize results when study effect estimates and their variances are available, yielding a quantitative summary of results" (Page et al., 2021a, p. 3). In line with these definitions, a meta-analytic systematic review was required to include a frequentist meta-analysis based on effect sizes from at least two primary studies.

Eligible reviews had to be published between 2022 and 2024 in one of the following journals (see Section *Information Sources and Search Strategy* for journal justification): *Child Development*, *Clinical Psychology Review*, *Development and Psychopathology*, *European Child & Adolescent Psychiatry*, *Journal of Child Psychology & Psychiatry*, *Journal of the American Academy of Child and Adolescent Psychiatry*, or *Psychological Bulletin*. Publication year was defined according to *Web of Science* (*Core Collection*; category: early access). Consequently, articles later assigned to a 2025 volume and issue were eligible if their early access publication date was 2024. When a volume and issue are assigned immediately, the early access and final publication dates are identical in *Web of Science*. Reviews first submitted to the journals before March 2021 were excluded, as the updated PRISMA 2020 (Page et al., 2021a) guidelines were published at that time, and authors were therefore unlikely to have been asked to revise their reporting from PRISMA 2009 (Liberati et al., 2009) during peer review. Consistent with the language policies of the target journals, all publications were in English.

Study type

Using this framework, eligible reviews had to (1) articulate a clearly defined research question or objective, (2) report explicit inclusion and exclusion criteria, (3) describe a systematic literature search naming at least one database and providing example keywords, and (4) include a systematic data extraction yielding at least two effect sizes from two or more primary studies that were (5) subsequently synthesized using frequentist meta-analytic methods. The following publication types were excluded due to distinct purposes or reporting standards: scoping reviews, overviews of reviews, umbrella reviews, meta-meta-analyses, meta-reviews, second-order meta-analyses, and systematic reviews focused on diagnostic test accuracy or psychometric properties of assessment tools.

Population

The target population consisted of samples with a mean age between 0 and 20 years; studies were also included if more than 90% of participants had a mean age below 20. Reviews were excluded if this information was not reported, without further examination of the primary studies. This criterion does not require that the

meta-analytic review include self-report outcomes for this population (see Section *Outcome*).

Intervention

Meta-analytic systematic reviews did not need to analyze intervention effects to be eligible; however, those that did had to examine at least one clinical psychological or psychotherapeutic intervention. Reviews focused solely on non-psychological interventions (e.g., psychopharmacology) were excluded.

Comparator

No inclusion criteria were imposed regarding comparators.

Outcome

Eligible outcomes were synthesized effect sizes addressing child and/or adolescent adjustment problems within the scope of CCAP. These included externalizing (behavioral) problems (e.g., hyperactivity), internalizing (emotional) problems (e.g., anxiety), and associated mental health disorders classified in established systems such as DSM-5 (Falkai & Wittchen, 2018) or ICD-10 (Dilling et al., 2015; e.g., social phobia, attention-deficit/hyperactivity disorder). Meta-analytic systematic reviews examining other clearly defined mental disorders (e.g., posttraumatic stress disorder, autism spectrum disorder, bipolar disorder, or schizophrenia), or broader indicators of mental health (e.g., psychological distress, overall functioning, suicidal behavior) were also eligible. Reviews dealing exclusively with non-psychological domains (e.g., academic performance, physical health) or biological measures (e.g., genetic or biochemical markers) were excluded.

Information Sources and Search Strategy

When this meta-review was conceptualized, there was no single, open-submission, non-invitation-only flagship CCAP journal devoted specifically to research synthesis. Accordingly, seven journals were selected as top-ranking outlets in CCAP that publish meta-analytic systematic reviews, representing the field's most impactful research. These journals were identified through a structured three-step procedure drawing on expert input and journal metrics. For details, see https://osf.io/4v3zr/. The final set included: (1) *Child Development*, (2) *Clinical Psychology Review*, (3) *Development and Psychopathology*, (4) *European Child &*

Adolescent Psychiatry, (5) *Journal of Child Psychology & Psychiatry*, (6) *Journal of the American Academy of Child and Adolescent Psychiatry*, and (7) *Psychological Bulletin*. Details on the piloted search strategy are presented in Siegel et al. (2025), with the complete search string provided in Table 1. Meta-analytic systematic reviews were retrieved through the *Web of Science Core Collection*, as available via the University of Vienna. At the time of searching, which was February 3, 2025, all seven journals were fully indexed in this database. The search yielded 614 hits.

Table 1

Search String within Web of Science Core Collection

Set	Content
1	IS=("0009-3920" OR "0272-7358" OR "0954-5794" OR "1018-8827" OR "0021- 9630" OR "0890-8567" OR "0033-2909")
2	TS=("systematic review" OR "meta-analy*" OR metaanaly* OR "meta-regress*" OR metaregress*)
3	EAY=(2022-2024)
4	1 AND 2 AND 3

Note. Reprint from Siegel et al. (2025). Web of Science Core Collection as subscribed by the University of Vienna, including Science Citation Index Expanded (SCI-EXPANDED)—1900-present; Social Sciences Citation Index (SSCI)—1900-present; Arts & Humanities Citation Index (AHCI)—1975-present; Conference Proceedings Citation Index – Science (CPCI-S)—1990-present; Conference Proceedings Citation Index – Social Science & Humanities (CPCI-SSH)—1990-present; Book Citation Index – Science; (BKCI-S)—2005-present; Book Citation Index – Social Sciences & Humanities (BKCI-SSH)—2005-present; Emerging Sources Citation Index; (ESCI)— 2005-present; Current Chemical Reactions; (CCR-EXPANDED)—1985-present; Index Chemicus (IC)—1993-present. "Exact search" feature of Web of Science turned on. Search performed on February 3, 2025, 614 hits. International Standard Serial Numbers (ISSNs) were used instead of journal titles to ensure unique source identification consistent with the approach of Chin et al. (2022).

Study Selection

The study selection process followed a two-stage screening procedure (Page et al., 2021a), with each meta-analytic systematic review being coded twice and screened independently by three researchers: Magdalena Siegel (MS), SF, and Julia Riedel (JR). In the first stage, titles and abstracts of identified records were screened using liberal acceleration (Page et al., 2018), meaning that any record marked as eligible by at least one author proceeded to full-text review. In the

second stage, full texts were assessed, and a meta-analytic systematic review was included only when all three researchers reached consensus. Cases of disagreement were resolved by the senior author, Martina Zemp (MZ). Overall inter-rater reliability was strong for both stages: screening agreement was 90% (pairwise range: 92–95%), and full-text agreement was 89% (pairwise range: 90–97%). Two articles (Shephard et al., 2022; Wang et al., 2022) included more than one meta-analytic systematic review. Subsequently, a random selection was done per study by performing a reproducible procedure in R. To guarantee sufficient replicability, the sample() function was implemented with a fixed random seed (R Core Team, 2021). Appendices D and F list, respectively, all included and excluded studies along with their reasons for exclusion after the full-text review.

Data Collection Process

Data collection for descriptive information and methodological quality was conducted by three researchers: SF and JR each coded 50% of the included records after full-text assessment (randomly sampled), and MS coded 10%. The inter-rater agreement on methodological quality was 80.2%, which is above the standard recommended by AMSTAR 2 (Shea et al., 2017). Reporting transparency was double-coded. JR coded 100% of the meta-analytic systematic reviews. HM and MU each coded 50%. Due to time-constraints calculation of inter-rater reliability was not feasible. Descriptive information was single-coded by SF and JR (50% each), thus no IRR was calculated. Authors of eligible systematic reviews were not contacted for further information, as the goal was to evaluate the reports based solely on the information provided therein. No automation tools were used for data extraction of methodological quality and reporting transparency. However, the data for evaluating statistical consistency were generated using the *statcheck* R package (see section "Data Items" below**;** Nuijten & Epskamp, 2024).

Data Items

In total, three item groups were coded, and a fourth group was generated via *statcheck*: (1) Descriptive study information, (2) methodological quality, (3) items of reporting transparency relevant to methodological quality, and (4) statistical consistency. For group three items, see Appendix E (Tables 4-9), and for the full codebook of the overarching project, see https://osf.io/qz8dr. Descriptive information about each meta-analytic systematic review and its journal

outlet was coded (see Section *Descriptive information about the review* in the codebook). This included the journal in which the review appeared, the specific (sub-)clinical outcomes examined in children or adolescents (recorded in free text and later grouped). Additionally, it was coded whether the review evaluated intervention effects (yes/no), the age group or groups targeted by the included studies (infancy [0–24 months], childhood [2–10 years], adolescence [11–20 years], with multiple assignments allowed), the number of primary studies included, and the number of participants represented (with separate counts for children and parents).

RQ 1 and 2: Methodological Quality

To examine the methodological quality of meta-analytic systematic reviews across domains, an adapted version of AMSTAR 2 (Shea et al., 2017; see "Deviations from Protocol"), allowing for comparisons with other meta-scientific studies on methodological quality, was used. AMSTAR 2 is a widely used, valid, and reliable 16-item checklist with seven critical domains (Kolaski et al., 2023; Leclercq et al., 2020a; Lorenz et al., 2019; Shea et al., 2017). The critical domains encompass the presence of a pre-registered protocol (item 2); a sufficient literature search (item 4); justification for exclusion of studies (item 7); assessments of RoB and publication bias (item 9, 13, 15); and the use of appropriate meta-analytical methods (item 11). Each item is rated "yes" or "no", with items 2, 4, 7, 8, 9 allowing a "partially yes" rating based on available information in the systematic review (Pieper et al., 2018a; Shea et al., 2017). Notably, items 11, 12, and 15 also allow an answer of "not applicable" (Pieper et al., 2018a; Shea et al., 2017), though this answer option was not relevant for the scope of this thesis, as solely reviews with a meta-analysis were considered eligible (see Section *Inclusion and Exclusion Criteria*). Overall confidence in the results of a review is categorized into four levels: high (0-1 item from a non-critical domain rated "no"); moderate (>1 item from a non-critical domain rated "no"); low (1 item from a critical domain rated "no"); and critically low (>1 item from a critical domain rated "no"). Critical weaknesses severely lower overall confidence, while multiple non-critical weaknesses may reduce confidence from moderate to low (Shea et al., 2017). For this meta-review, both the frequency of overall AMSTAR 2 confidence ratings and response options ("*Yes*", "*Partially Yes*", and "*No*") of critical domains were assessed. Additionally, an

AMSTAR 2 confidence sum score was computed across all items, with higher scores reflecting higher methodological quality. Each item was scored as 1 point for “yes”, 0.5 points for “partial yes”, and 0 points for “no”, resulting in a continuous AMSTAR 2 sum score ranging from 0 to 16. A maximum score of 16 represents perfect adherence across all AMSTAR 2 items.

RQ 3: Transparency-Checklist and Methodological Quality

PRISMA 2020 is the standard tool for reporting transparency in psychology-related systematic reviews (Kolaski et al., 2023; Steil et al., 2022), providing detailed reporting guidelines for review authors (Page et al., 2021a; 2021b), while AMSTAR 2 is used for a critical assessment of methodological quality (Shea et al., 2017). These constructs may overlap, suggesting a potential association between the tools (Lorenz et al., 2020; Peña et al., 2022). For the meta-review reported within this thesis, PRISMA 2020 (Page et al., 2021a) and PRISMA-S (for search strategies; Rethlefsen et al., 2021) were adapted to create a 40-unique-item-ID appraisal tool for reporting transparency (see Appendix E, Tables 4-9), aiding in identifying additional nuances in methodological quality within AMSTAR 2 critical domains (for the full checklist, see Siegel et al., 2025). For instance, items were added to assess interrater reliability reporting in screening and coding (without evaluating their value, as AMSTAR 2 does), and reporting thresholds for statistical methods for bias detection were added (Siegel et al., 2022; Table E9). Additionally, an item adapted from López-Nicolás et al. (2022) was included to assess whether formulae for effect size calculation are reported (Table E8), an essential factor for reproducibility with meta-analyses not explicitly covered in PRISMA (Lakens et al., 2017; Maassen et al., 2020). The checklist incorporates transparency-related items relevant to methodological quality and, taken together, may provide a more comprehensive assessment. In this meta-review, relevant transparency checklist items will be narratively assessed for each AMSTAR 2 critical domain. By way of illustration, AMSTAR 2 item 15 (publication bias) requires a binary judgment of whether authors conducted and discussed graphical or statistical assessments of publication bias (Shea et al., 2017). The transparency checklist adds detail by specifying which methods were used and, for statistical approaches, whether decision thresholds were reported. For graphical assessments, it further captures whether authors reported the number of reviewers involved, whether assessments

were conducted independently, and how discrepancies were resolved (Table E9). Similarly, for AMSTAR 2 items 9 and 13 (RoB assessment and interpretation), the transparency checklist documents whether authors reported decision rules for non-standard or newly developed tools, the number and independence of reviewers, procedures for resolving disagreements, any adaptations to existing tools, and whether study authors were contacted for missing or unpublished information relevant to RoB judgments (Table E7).

RQ 4: Statistical Consistency

To assess the consistency of *p*-values with corresponding degrees of freedom and test statistics (Nuijten et al., 2016), this meta-review used the *statcheck* CRAN package (version 1.50; Nuijten & Epskamp, 2024) in its default settings (Nuijten & Epskamp, 2024). *Statcheck* is a free software tool that automatically scans academic articles for statistical inconsistencies in reported NHST results by recalculating *p*-values from degrees of freedom and test statistics and comparing them to author-reported values. The output classifies values as consistent, inconsistent, or grossly inconsistent (Nuijten & Polanin, 2020). An inconsistency refers to a situation in which the reported *p*-value and the recalculated *p*-value are not the same, but the difference is not large enough to alter the overall interpretation. Gross inconsistencies arise when the recalculated *p*-value suggests a different conclusion, such as a previously significant result becoming non-significant or vice versa (Nuijten & Polanin, 2020). For this meta-review, the full text .html file of each meta-analytic systematic review was used, and the *statcheck* output was dichotomized: reports were classified as either containing (1) or not containing (0) any inconsistencies.

Data Analysis

Descriptive statistics were computed, including means (*M*) and standard deviations (*SD*) for continuous variables, and counts and percentages for categorical and dichotomous variables. For research questions addressing frequencies (RQs 1-4), counts and percentages were reported. For RQ3, the fulfillment of relevant transparency items (based on PRISMA and its extensions) was assessed narratively by contrasting AMSTAR 2 critical domains and related transparency items to identify potential additional nuances of methodological quality in relation to reporting quality. Deviating from the protocol (see section “Deviations

from Protocol"), the exploratory analysis of RQ4 followed only the continuous approach. For this, the *statcheck* outcome was dichotomized into "no statistical inconsistency" vs. "any inconsistencies", and an AMSTAR 2 sum score (0-16) was calculated with higher scores reflecting higher methodological quality. *T*-tests were conducted, and effect sizes were reported using Hedges' *g* (Borenstein et al., 2021). Statistical analyses were performed in R (R Core Team, 2021) using RStudio, with the most recent CRAN versions of packages available at the time of analysis (November 19, 2025). R packages used included *tidyverse* (Wickham et al., 2019), *dplyr* (Wickham et al., 2025a), *readr* (Wickham et al., 2025b), *scales* (Wickham et al., 2025c)*, tidyr* (Wickham et al., 2025d), *here* (Müller, 2020), *janitor* (Firke, 2024), *purrr* (Wickham & Henry, 2025), *readxl* (Wickham & Bryan, 2023), *writexl* (Aden-Buie, 2023), *ggplot2* (Wickham, 2016), *and stringr* (Wickham, 2025).

Results

Identification and Selection of Studies

From the systematic literature search in *Web of Science Core Collection* (via University of Vienna) on February 3, 2025 (validated on February 4, 2025), a total of 614 potentially relevant references were retrieved. After title/abstract screening, 145 references remained for full text assessment (see Appendix F for the PRISMA flow diagram). In total, 66 meta-analytic systematic reviews met the final criteria for inclusion (a complete list of included reviews, as well as a list of justified study exclusions, see Appendix C and D, respectively).

Descriptive Characteristics of Included Meta-Analytic Systematic Reviews

Descriptive characteristics of the included meta-analytic systematic reviews are reported in Table 2. The sample (k = 66) encompasses more than 3,300 single publications, with data from over three million child participants and more than 150,000 parents. The meta-analytic systematic reviews differed substantially in the number of primary studies included (M = 50; range = 10 - 365). Most meta-analytic systematic reviews focused on school-aged children (75.8%) as well as adolescents (90.9%), with only a few concentrating on infants. A considerable proportion (40.9%) dealt with intervention effects. Content-wise, the reviews most often addressed externalizing (28.8%) and internalizing symptoms (27.3%), followed by specific mental health conditions such as depression, attention-deficit/hyperactivity disorder, anxiety, post-traumatic stress disorder, autism, and suicidality. Most meta-analytic systematic reviews were published in *European Child & Adolescent Psychiatry* (30.3%), *Clinical Psychology Review* (21.2%), and *Journal of Child Psychology and Psychiatry* (16.7%). Additional details are provided in Table 2; further study characteristics are presented in Appendix H.

Table 2

Descriptive Characteristics of Included Meta-Analytic Systematic Reviews (k = 66)

Meta-Analytic Systematic Review	Descriptive Statistics
Characteristic	
Included number of primary studies	3,323 (M = 50; range = 10-365)

Meta-Analytic Systematic Review	Descriptive Statistics
Included number of children	3,134,802 (*M* = 65,308; range = 436-932,229; *k* = 48)
Included number of parents	152,302 (*M* = 19,038; range = 1,006-52,203; *k* = 8)
Intervention focus	27 (40.9%)
Age group(s)*	
Infants (0-24 months)	13 (19.7%)
School-aged (2-10 years)	50 (75.8%)
Adolescents (11-18 years)	60 (90.9%)
Unclear	2 (3.0%)
Content domain(s)*	
Externalizing symptoms	19 (28.8%)
Internalizing symptoms	18 (27.3%)
Depressive symptoms	11 (16.7%)
Attention deficit hyperactivity disorder	9 (13.6%)
Anxiety	9 (13.6%)
Post-traumatic stress disorder	7 (10.6%)
Autism spectrum disorder	6 (9.1%)
General psychopathology	6 (9.1%)
Suicidality	5 (7.6%)
Publication outlet	
European Child & Adolescent Psychiatry	20 (30.3%)
Clinical Psychology Review	14 (21.2%)
Journal of Child Psychology and Psychiatry	11 (16.7%)
Journal of the American Academy of Child and Adolescent Psychiatry	11 (16.7%)
Development and Psychopathology	4 (6.1%)
Child Development	3 (4.6%)
Psychological Bulletin	3 (4.6%)

Note. Child and parent sample sizes refer only to meta-analytic systematic reviews reporting extractable data (*k* = 48 and *k* = 8, respectively). *Multiple assignments possible.

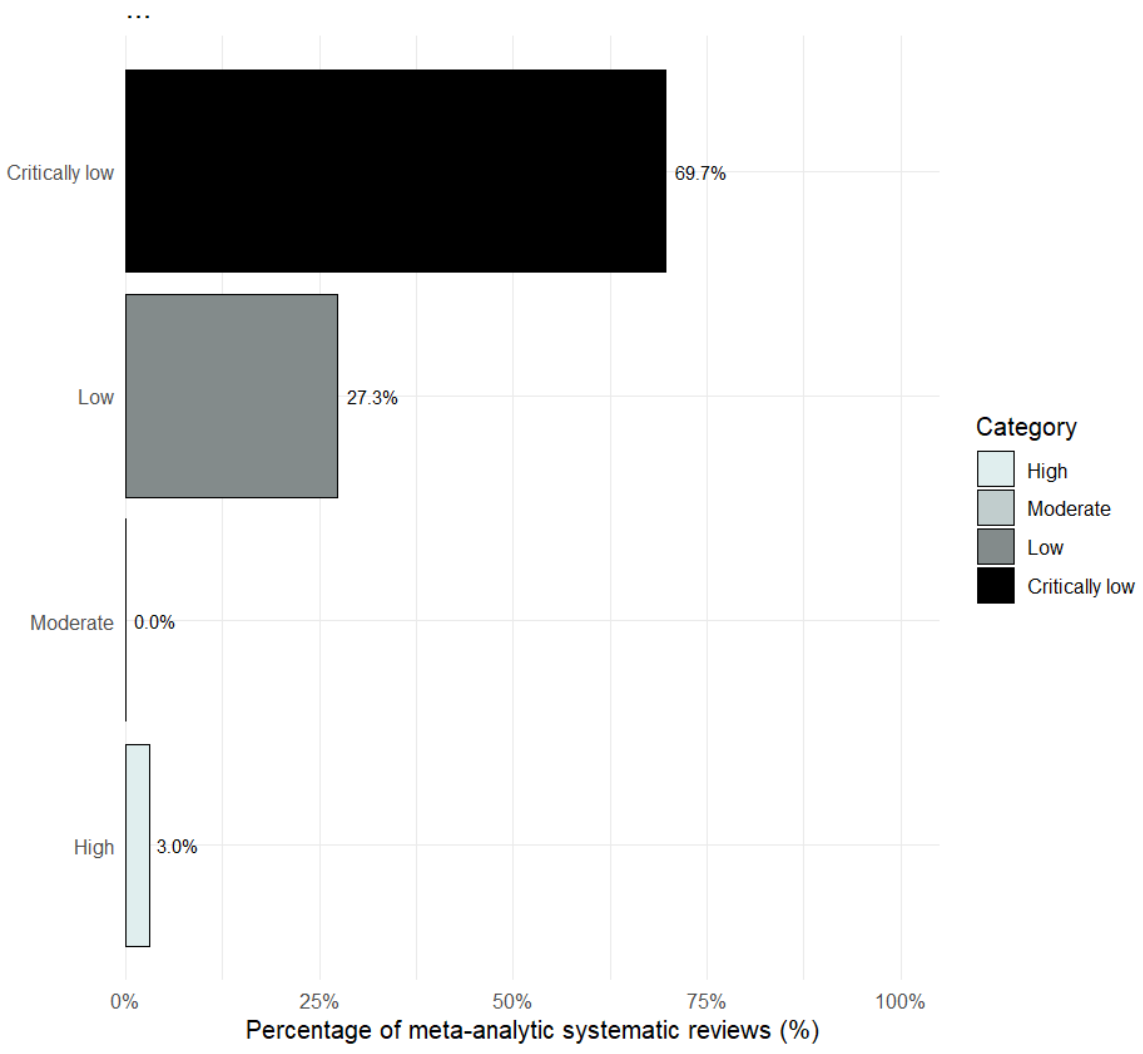
Methodological Quality Assessment

RQ 1: Overall Methodological Quality

Methodological quality was assessed using the AMSTAR 2 checklist (Shea et al., 2017). The mean quality sum score was 11.2 (*SD* = 2.34) out of 16 possible points. Most meta-analytic systematic reviews failed to meet at least one critical domain, substantially lowering overall confidence ratings. About two-thirds received a critically low rating (*k* = 46; 69.7%), followed by 18 (27.3%) with a low rating. Merely 2 reviews (3.0%), by Bellato et al. (2024) and Haza et al. (2024), achieved a high-quality rating. None of the meta-analytic systematic reviews were rated as being of moderate quality (Figure 1).

Figure 1

AMSTAR 2 Overall Confidence Rating, k=66

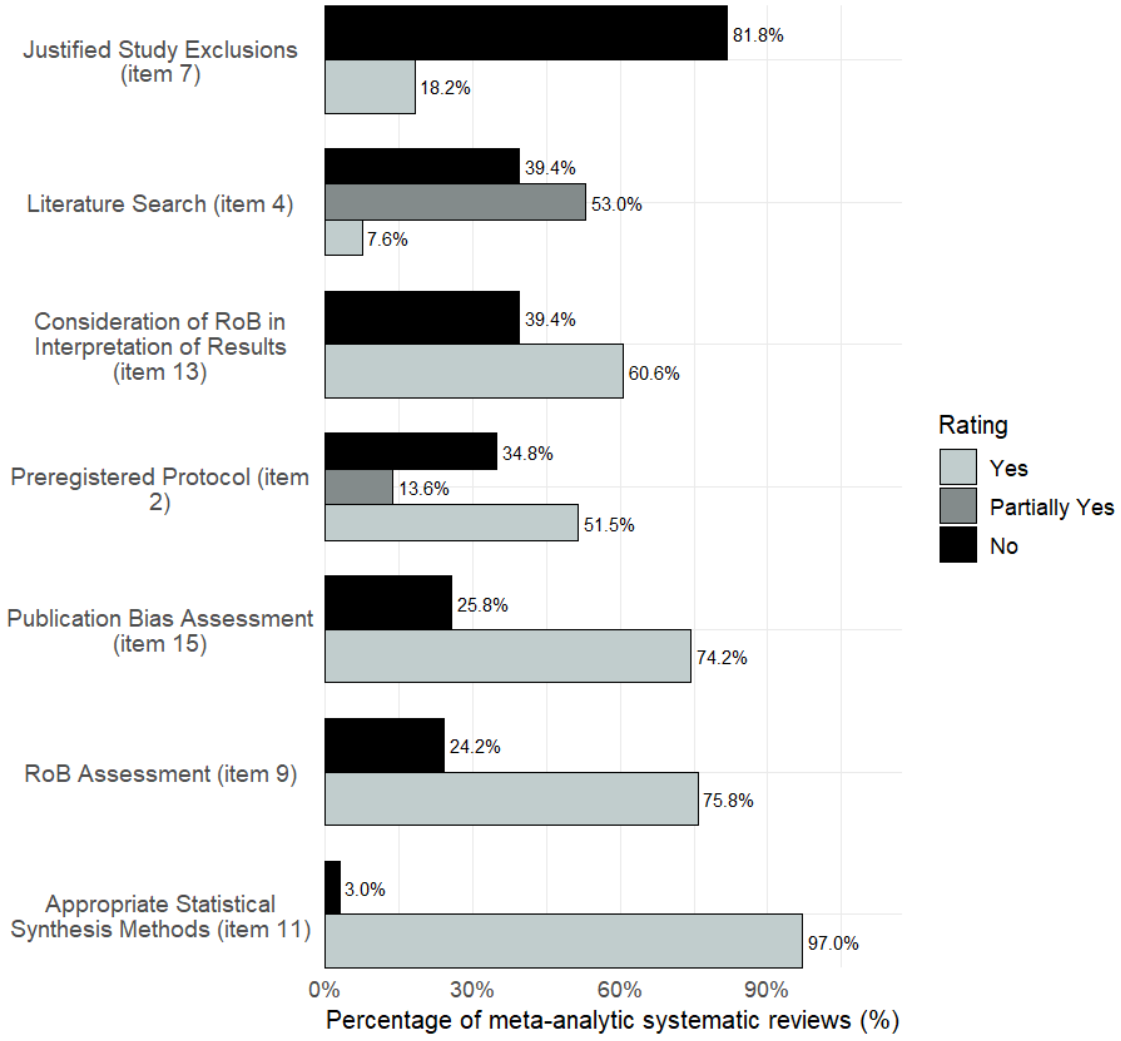


RQ 2: Methodological Quality Within AMSTAR 2 Critical Domains

As illustrated in Figure 2, adherence across the seven critical domains varied substantially. The highest level of compliance was observed for item 11 (using appropriate synthesis methods; *k* = 64; *Yes* = 97%); item 9 (assessment of RoB; *k* = 50; *Yes* = 75.8%), and item 15 (adequate investigation of and reporting of publication bias; *k* = 17; *Yes* = 74.2%). Moderate adherence was observed for item 2 (having a preregistered protocol), with 34 meta-analytic systematic reviews rated *Yes* (51.5%), nine received a *Partially Yes* rating (13.6%), and about a third received a *No* rating (*k* = 23; 34.8%). A *Partially Yes* rating requires stating the use of a protocol including review questions, search strategy, eligibility criteria, and RoB assessment, whereas a *Yes* rating requires protocol registration, a prespecified meta-analysis plan, plans to assess heterogeneity, and justification of any protocol amendments (Shea et al., 2017). Item 4 (conducting a comprehensive search) yielded a more varied pattern, with five of the reviews rated *Yes* (7.6%), 35 rated *Partially Yes* (53.0%), and 26 rated *No* (39.4%). A *Partially Yes* rating reflects use of at least two databases, reporting of keywords or search strategies, and justifying publication restrictions; a *Yes* rating additionally requires backward searching, consultation with content experts, searches of grey literature where relevant, and a search performed within 24 months of completing the review (Shea et al., 2017). Similarly, 13 (considering RoB when interpreting findings) showed limited adherence, with a slight majority of 40 meta-analytic systematic reviews rated *Yes* (60.6%), and 26 received a *No* rating (39.4%). Item 7 (providing rationale for excluded studies; *k* = 54; *No* = 81.8%) showed the lowest level of compliance.

Figure 2

Response Distribution for AMSTAR 2 Critical Domains, k=66



RQ 3: Relevant Transparency Checklist Items as Sources of Additional Nuance in Methodological Quality Within AMSTAR 2 Critical Domains

Research Question 3 examined the extent to which transparency checklist items derived from PRISMA (Page et al., 2021a) and its extensions (Rethlefsen et al., 2021) provide additional nuance regarding methodological quality within the critical domains of AMSTAR 2. A narrative comparison of AMSTAR 2 domain ratings with corresponding transparency-related reporting items demonstrated that the adapted transparency checklist offered meaningful additional granularity. The aim was not to re-evaluate methodological quality, but to assess whether transparency items could further qualify or contextualize AMSTAR 2 judgments. Across 66 meta-analytic systematic reviews, reporting quality varied substantially across items in the adapted PRISMA-based transparency checklist (see Appendix

E; Tables E4-E9). While several transparency items showed consistently high fulfillment, others revealed frequent reporting gaps that clarified potential sources of methodological limitations within AMSTAR 2 critical domains.

Compared to the AMSTAR 2 item 2 (preregistered protocol) rating (*Yes* = 51.5%; *k* = 34, *Partially Yes* = 13.6%; *k* = 9), the transparency checklist results revealed substantial heterogeneity in protocol-related reporting (Table E4). Most reviews reported working from predefined methods, with 65.1% (*k* = 43) being publicly available via a provided link and most commonly registered in PROSPERO or OSF, and 7.6% (*k* = 5) not being publicly accessible. Another 7.6% (*k* = 5) reported not working from a protocol. Nearly one-fifth of reviews (19.7%; *k* = 13) did not report protocol use at all. Reporting of protocol deviations was frequently incomplete. Most reviews did not report whether amendments were made (63.0%; *k* = 29). Among those that did (37.0%; *k* = 17), most provided reasons for amendments (88.2%; *k* = 15), but fewer specified the stage at which changes occurred (41.2%; *k* = 7). While AMSTAR 2 evaluates whether methods were established a priori and whether deviations were justified, transparency items allowed for finer distinctions regarding protocol availability, registration, and amendment reporting, revealing variability not fully captured by AMSTAR 2 alone.

AMSTAR 2 item 4 (comprehensive literature search) was partially fulfilled in more than half of the reviews (*k* = 35), although it was completely fulfilled in only 7.6% (k = 5) of the reviews. However, the transparency checklist items revealed notable gaps in how search strategies were reported. Although all reviews searched bibliographic databases, nearly half (43.9%; *k* = 29) did not report whether search limits or filters were used, and justifications for such limits were rarely provided (25.0%; *k* = 9). Searches of trial registries were mostly only partially reported (70%; *k* = 7), while involvement of experts was transparently documented (100%; *k* = 12). In addition, nearly half of the reviews (47.0%; *k* = 31) did not report whether or how searches were updated. Transparency gaps were particularly pronounced for advanced search practices. Peer review of search strategies was almost entirely unreported (1.5%; *k* = 1), and the derivation of search terms was unspecified in more than three-quarters of reviews (78.8%; *k* = 52; Table E5).

AMSTAR 2 item 7 (justified study exclusions) was infrequently fulfilled, with only 18.2% (k =12) of reviews rated *Yes*. The transparency checklist results clarified

this pattern: while most reviews reported general reasons for exclusion, only 13.6% (*k* = 9) provided study-level exclusions linked to specific reasons. The majority (78.8%; *k* = 52) relied on aggregated reasons without per-study justification. 5 reviews did not report on study exclusions (7.6%; Table E6). Thus, transparency items did not merely corroborate low AMSTAR 2 performance but highlighted a specific reporting practice, lack of per-study exclusion documentation, that likely contributed to unfavourable domain ratings.

AMSTAR 2 item 9 (RoB assessment) and 13 (consideration of RoB in interpretation of results) were rated as fulfilled in 75.8% (*k* = 50) and 60.6% (*k* = 40) of reviews, respectively, yet the transparency checklist items revealed substantial heterogeneity within these ratings. Most reviews reported conducting a RoB assessment (80.3%; *k* = 53) and specified the tool used, but fewer than half justified their RoB judgments. Reviewer procedures were inconsistently reported: while most reviews reported involvement of at least two authors (58.5%; *k* = 31), nearly one-quarter did not report the number of assessors involved (30.2%; *k* = 16, and a small number relied on single-author assessments. Among reviews with multiple assessors, discrepancy resolution procedures were reported in 63.9% of cases (*k* = 23). Among reviews that modified established instruments, 73.7% (*k* = 14) described these adaptations, while the remainder did not, limiting interpretability. Use of non-standardized or newly developed tools was rare (*k* = 4), and only half of such reviews reported decision rules for deriving overall judgments. As no review reported using automation tools for RoB assessment, this was left out of the analysis (Table E7).

AMSTAR 2 item 11 (appropriate statistical synthesis methods) demonstrated the highest fulfillment rate (97%; *k* = 64). Nevertheless, the transparency checklist items revealed notable omissions in reporting statistical details. While most reviews investigated heterogeneity (95.5%; *k* = 63), only about half (50.8%; *k* = 32) specified the moderator variables tested. Weighting methods were reported in fewer than one-quarter of reviews (24.6%; *k* = 16), and effect size interpretation thresholds were omitted in 41.7% (*k* = 25). These findings suggest that even when statistical methods are deemed appropriate within AMSTAR 2, transparency regarding analytic decisions and their interpretation varies considerably, limiting reproducibility and interpretability (Table E8).

AMSTAR 2 item 15 (publication bias assessment) was fulfilled in 74.2% of reviews. Transparency checklist items indicated that three-quarters of reviews presented assessments of RoB due to missing results (75.9%; *k* = 44; Table E9). Although statistical methods for detecting reporting bias were commonly used (*k* = 53; 80.3%), explicit decision thresholds were reported in only 11.3% of applicable reviews (*k* = 6). Additional transparency items revealed near-complete non-reporting of author contact for missing data, and reviewer procedures for graphical publication bias assessments. Therefore, item ID 14.09 (procedures for resolving discrepancies) as well as item ID 14.08 (reporting if reviewers worked independently) were not applicable and left out of the analysis. As no review used formal tools to assess reporting bias, a narrative comparison was not applicable.

Taken together, the relevant transparency checklist items demonstrated that substantive differences exist between reviews that score similarly within AMSTAR 2 critical domains. This pattern was most pronounced for domains related to the literature search, data extraction, and RoB assessment, where transparency items identified substantial variability even among reviews rated as high quality on AMSTAR 2. These additional nuances provide valuable context for interpreting methodological quality in meta-analytic systematic reviews in CCAP and may complement the AMSTAR 2 assessments in this review by offering a more granular understanding of reporting practices within the methodological quality of the assessed reviews.

RQ 4: Association Between Methodological Quality and Statistical Consistency

To explore whether methodological quality was associated with statistical consistency, the AMSTAR 2 score was operationalized as a continuous sum score ranging from 0 to 16, with higher values indicating greater methodological quality. Statistical consistency was assessed using *statcheck* at the publication level, which identifies inconsistencies between reported NHST results and recalculated *p*-values (Nuijten & Epskamp, 2024). Only reviews for which *statcheck* returned at least one extractable statistical test (*n* = 23; Appendix G, Figure 5) were included in this analysis. No gross inconsistencies were found.

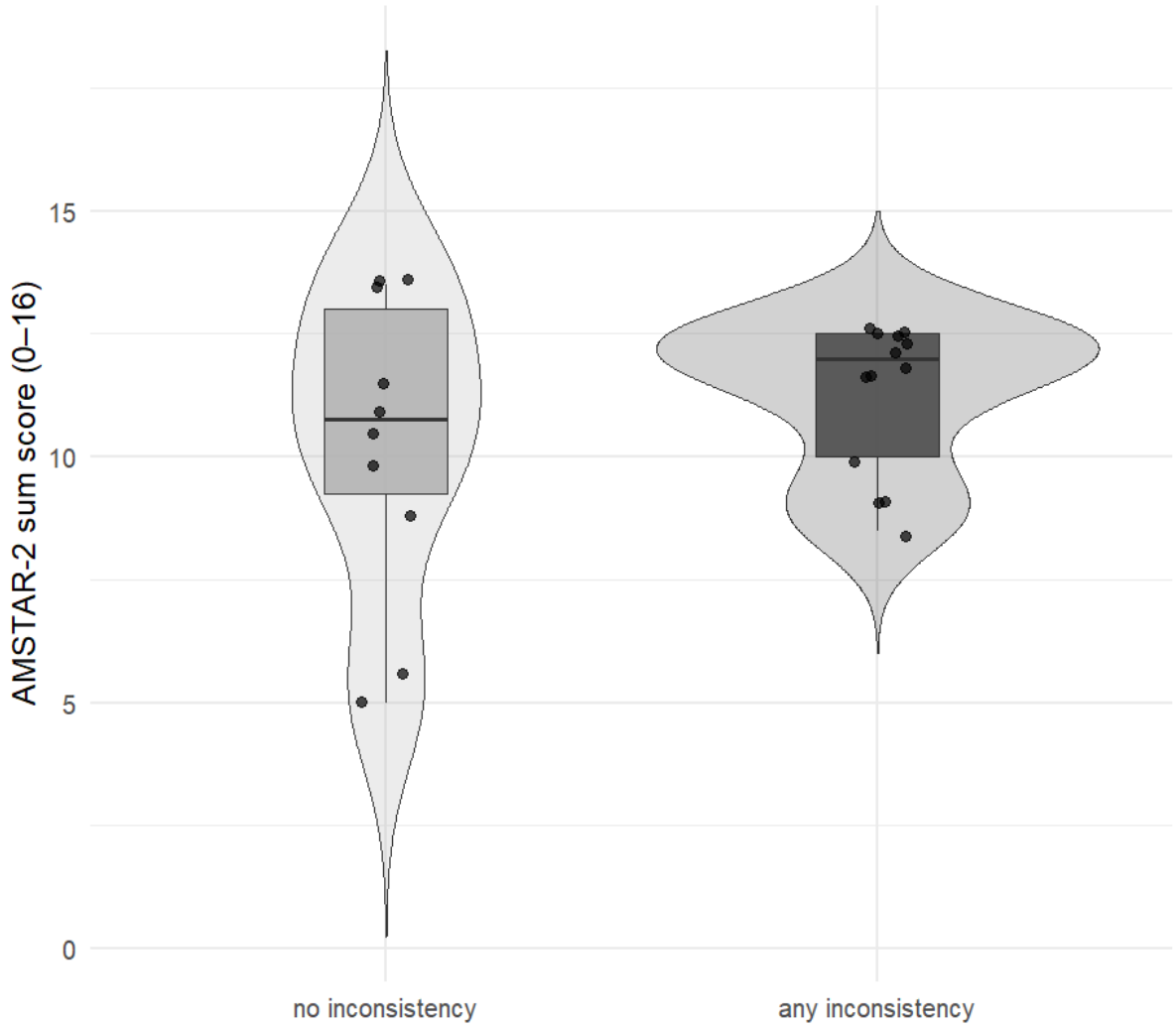
An independent-samples *t*-test compared AMSTAR-2 sum scores between reviews without inconsistencies (43.5%; *k* = 10) and those with at least one

inconsistency (56.5%; *k* = 13). Reviews without inconsistencies had a mean AMSTAR-2 score of 10.30 (*SD* = 3.08, 95% CI [8.09, 12.51]), whereas reviews with one or more inconsistencies had a mean score of 11.23 (*SD* = 1.54, 95% CI [10.30, 12.16]). This difference was not statistically significant, $t(12.42) = -0.87$, $p = .398$. The corresponding effect size was small, *Hedges' g* = 0.39, with a wide 95% CI [-0.45, 1.22], indicating a high degree of uncertainty around the estimate. Figure 3 presents the distribution of AMSTAR-2 scores by inconsistency status.

Taken together, the results suggest that methodological quality, as measured by AMSTAR-2, was not meaningfully associated with the presence of statistical inconsistencies among the subset of reviews containing extractable NHST results. Notably, taking into consideration the limited power, the effect size still represents a small to medium effect and was in the opposite direction from what was previously expected (i.e., reviews with at least one inconsistency had higher quality scores).

Figure 3

Continuous AMSTAR-2 Sum Score by Statistical Consistency, k=23



Note. Distribution of continuous AMSTAR 2 confidence scores for meta-analytic systematic reviews with and without statistical inconsistencies detected by *statcheck*. Violin shapes depict the kernel-smoothed distribution of scores within each group, overlaid with boxplots showing median and interquartile ranges, and jittered points representing individual reviews. Higher values indicate closer adherence to AMSTAR 2 methodological criteria. Because all included reviews received a "*Critically Low*" AMSTAR 2 overall confidence rating, only the continuous AMSTAR 2 confidence score was used to examine the association with statistical consistency.

Discussion

The present meta-review is the first to assess methodological quality of recent meta-analytic systematic reviews published in prestigious journals within the field of CCAP. Beyond summarizing methodological quality, this review uniquely explores nuances of AMSTAR 2 critical domains by incorporating relevant aspects of reporting transparency and examining the correlation between methodological quality and statistical consistency. The results provide an initial overview of methodological quality among the field's most prominent meta-analytic systematic reviews, highlighting both areas of encouraging uptake and substantial gaps in scientific credibility. The discussion follows the sequence of the research questions, beginning with overall methodological quality and its degree of implementation, and then examining performance across critical domains. It then reflects on the findings of the exploratory analyses and evaluates the strengths and limitations of the present meta-review. Finally, it addresses the practical implications for meta-analysts and meta-scientists, outlines directions for future research in CCAP, and concludes by advocating for stronger adoption of methodological quality assessment in future CCAP evidence syntheses.

Methodological Quality

RQ1: Overall Methodological Quality

This meta-review evaluated the methodological quality of meta-analytic systematic reviews published between 2022 and 2024 in leading CCAP journals. With respect to the first research question, the findings indicate that the overall methodological quality of recent meta-analytic systematic reviews is concerningly low. The established AMSTAR 2 overall confidence rating system heavily weights its seven critical domains, such that failure to meet one or more of these domains results in a severe downgrade to either *Low* or *Critically Low* (Shea et al., 2017). Two-thirds (k = 46) of the meta-analytic systematic reviews were rated as *Critically Low*, indicating multiple failures in key aspects such as protocol preregistration, adequacy of the search strategy, justification of excluded studies, appropriate synthesis methods, or handling of RoB and publication bias. An additional 27.3% (k = 18) were classified as *Low* due to violations of at least one critical domain. Only two reviews (3%) by Bellato et al. (2024) and Haza et al. (2024) achieved a *High* confidence rating, and none reached a *Moderate* rating. Although this is not a

recommended procedure (Shea et al., 2017), the mean AMSTAR 2 sum score of 11.2 out of 16 suggests that, on average, contemporary meta-analytic systematic reviews demonstrate moderate adherence to core methodological criteria. Together, these distributions indicate that most recent meta-analytic systematic reviews published in high-impact CCAP journals are characterized by critical methodological weaknesses that substantially limit confidence in their conclusions.

Similar patterns have been reported in prior meta-scientific assessments of the methodological quality of (meta-analytic) systematic reviews within CCAP (Bellato et al., 2023; Byrne et al., 2023; Cortese et al., 2019; Espada et al., 2023; Galán-Luque et al., 2023; García-Fernández et al., 2023; Kwan et al., 2020; O'Shea et al., 2025; Romero et al., 2023; Sala et al., 2024; Sánchez De Ribera et al., 2020; Scarlett et al., 2023), as well as in psychological science more broadly (Leclercq et al., 2020a; Lorenz et al., 2020). Alarmingly, even when authors mention their use of AMSTAR 2 in the review process, this still may not result in better methodological quality, as shown in a cross-sectional meta-scientific investigation by Bojcic et al. (2024). These studies collectively show that meta-analytic systematic reviews frequently lack key methodological safeguards.

The predominance of *Critically Low* ratings further suggests a persistent disconnect between the growing popularity of meta-analytic methods (Ioannidis, 2016) and the methodological rigor required to ensure trustworthy evidence synthesis (Shea et al., 2017). Although meta-analytic systematic reviews are often regarded as the highest level of evidence (Clarke & Chalmers, 2018; Murad et al., 2016; Strawbridge et al., 2025), the present findings, consistent with previous research, suggest that this assumption may not hold unless reviews adhere to established standards for methodological quality (Chappelle et al., 2023; Ioannidis, 2016; Møller et al., 2018). Within CCAP, where conclusions drawn from meta-analytic systematic reviews often inform interventions, policy decisions, and theoretical models (Bellato et al., 2023; Weisz et al., 2023), insufficient methodological quality poses a substantive risk to the validity of cumulative knowledge (Murad et al., 2016). This concern is heightened by the fact that these findings reflect a sample of high-impact meta-analytic systematic reviews within the field of CCAP.

At the same time, the relatively high mean sum score indicates that many reviews adequately implement several non-critical AMSTAR 2 components, suggesting that poor confidence ratings are often driven by a small number of crucial omissions, such as lack of justified study exclusion, inadequate literature search strategies, and failure to consider RoB in the reviews' findings rather than widespread failure across all dimensions. This pattern has previously been described as a "floor effect" in the AMSTAR 2 scoring system, which has limited discriminative ability and tends to overestimate *Low* or *Critically Low* confidence ratings (Leclercq et al., 2020b), which severely downgrades reviews if one or more critical domains are not fulfilled (Shea et al., 2017). Nonetheless, although AMSTAR 2 sets a high methodological threshold, *Low* and *Critically Low* ratings still reflect genuine deficiencies in methodological quality (Lorenz et al., 2020). This pattern also highlights concrete areas in which improvements may yield substantial gains, most notably preregistration, comprehensive search strategies, consistent justification of study exclusions, and rigorous assessment of RoB and publication bias.

Taken together, these results demonstrate that although meta-analytic systematic reviews in CCAP show partial adherence to methodological standards, the overall level of confidence that can be placed in their findings is very limited. Improving methodological rigor in future meta-analytic systematic reviews should therefore be a priority for authors, reviewers, and journals alike. Clearer methodological reporting, strengthened editorial policies requiring AMSTAR 2 self-evaluations and use during peer-review, and increased uptake of preregistration may help narrow the gap between current practices and the standards required for high-quality evidence synthesis (Bojcic et al., 2024). Academic institutions could further support meta-scientific goals beyond editorial guidelines by institutionalizing the evaluation of knowledge quality and scientific activity through forming meta-scientific teams within faculties, consisting of statisticians, methodologists, and research integrity specialists (Lozada-Martinez et al., 2025). In parallel, universities could develop a standardized research quality index to assess study design, reporting transparency, and reproducibility, generating quantifiable metrics of high-quality research and institutional performance tailored to the institution's research context (Lozada-Martinez et al., 2025). As many practitioners may not be fully

trained in evaluating meta-analytic systematic reviews (Garritty et al., 2019; OCEBM Levels of Evidence Working Group, 2011), universities could integrate structured meta-research training into curricula for postgraduate students and early-stage researchers, focusing on bias identification, evaluation of methodological quality, and detection of systematic errors (Lozada-Martinez et al., 2025). Finally, on a staff-member level, assessment should shift away from journal-based metrics, such as publication rate and impact factor, toward criteria emphasizing methodological quality with researchers engaged in high-quality peer review, transparent reporting, and data-sharing practices being appropriately incentivized (Lozada-Martinez et al., 2025).

RQ2: Methodological Quality Within AMSTAR 2 Critical Domains

Addressing RQ2, which focused on methodological quality of meta-analytic systematic reviews within the seven AMSTAR 2 critical domains, the present findings reveal substantial variability in adherence across domains. While some criteria were met by most reviews, other core methodological requirements were frequently neglected, resulting in mostly poor overall methodological quality across the field. Encouragingly, several critical domains demonstrated relatively high compliance. Most reviews adhered to appropriate statistical synthesis methods (AMSTAR 2 item 11), with 97% (k = 64) meeting this criterion, suggesting that once quantitative data extraction is undertaken, authors generally employ suitable meta-analytic techniques. Similarly, RoB assessments (item 9) and investigations of publication bias (item 15) were conducted in approximately three-quarters of meta-analytic systematic reviews. This represents a promising development compared to earlier CCAP-related meta-research, which predominantly reported inadequate RoB assessments (Byrne et al., 2023; Galán-Luque et al., 2023; Kwan et al., 2020; Romero et al., 2023; Sánchez De Ribera et al., 2020) and limited assessments of publication bias (Byrne et al., 2023; Sánchez De Ribera et al., 2020). These practices are central to evaluating the credibility of synthesized findings, and their relatively frequent use likely reflects growing awareness of bias-related threats in quantitative synthesis within psychological research (Leclercq et al., 2020a). This pattern may also be influenced by the study's focus on high-impact meta-analytic systematic reviews, as journals in this category often require such assessments. However, the present findings for item 9 may not be entirely robust, as a recent

meta-study by Kolaski et al. (2024a) found that only a small proportion of commonly used RoB tools fully meet AMSTAR 2 item 9 standards. This is particularly the case when tools are modified or applied across mixed study designs (Kolaski et al., 2024a). Moreover, AMSTAR 2 item 11 has received criticism for lacking sufficient detail on which meta-analytic method may be deemed appropriate, and no criterion explicitly addresses the use of one-way sensitivity analyses to test the robustness of results (De Santis et al., 2023; Leclercq et al., 2020a). This limitation may lead to an overestimation of the use of adequate statistical synthesis methods, and it is therefore advisable to discuss additional decision rules upon which the methodological appraisal will be based within the research team (Leclercq et al., 2020a). Due to time constraints, full double-coding and team discussion on the AMSTAR 2 ratings were not feasible for this thesis.

Moderate adherence was observed for AMSTAR 2 item 2, which concerns preregistered protocols. Although 51.5% (k =34) of reviews completely and 13.6% (*k* = 9) of reviews partially met this criterion, a substantial proportion of 34.8% (k = 23) of high-impact reviews were not preregistered. This finding in a high-impact sample is somewhat surprising in the sense that editorial policies of prestige journal outlets still allow the absence of preregistration in their review process. These omissions are particularly consequential because lack of preregistration can introduce selection bias, publication bias, and other forms of research distortion at early stages of the review process (Bakker et al., 2020; Munafò et al., 2017; Nosek et al., 2022). Though still a limited uptake of preregistration, it exceeds those reported in earlier CCAP meta-reviews (0%–40%; Byrne et al., 2023; Cortese et al., 2019; Espada et al., 2023; Galán-Luque et al., 2023; Kwan et al., 2020; Sánchez De Ribera et al., 2020; Scarlett et al., 2023) as well as rates reported in broader areas of psychology (e.g., 19% in clinical psychology; López-Nicolás et al., 2022; 36% in cognition training; Sandoval-Lentisco et al., 2024). This result is encouraging and may reflect increased attention to preregistration in recent years, particularly among reviews published between 2022 and 2024 (Moreau & Gamble, 2022).

Other critical domains, however, showed somewhat lower compliance. The integration of RoB considerations into the interpretation of results (item 13) and adherence to comprehensive search strategies (item 4) were mixed and often

insufficient. Although the majority of reviews (60.6%; *k* = 34) did incorporate RoB assessments into their narrative conclusions, 39.4% (*k* = 26) did not. This gap suggests that while many reviews conduct RoB assessments, they do not consistently use these assessments to contextualize their claims, thereby limiting the interpretive transparency recommended by AMSTAR 2 (Shea et al., 2017). Therefore, inconsistent implementation of RoB assessments remains a pervasive challenge (Kolaski et al., 2024b). Likewise, 39.4% (*k* = 26) failed to conduct or report a sufficiently comprehensive literature search. This reflects a serious issue, as inadequate search strategies can undermine reproducibility and contribute to research waste (López-Nicolás et al., 2022; Rethlefsen et al., 2024). AMSTAR 2 item 4 has been criticized for being overly subjective in defining what constitutes a "comprehensive" literature search, thereby making assessments highly dependent on reviewer expertise (Pieper et al., 2018a) and yielding lower interrater reliability than other AMSTAR 2 items (Lunny et al., 2025). Given that such expertise in evaluating meta-analytic systematic reviews is frequently lacking in practitioners (Garritty et al., 2019; OCEBM Levels of Evidence Working Group, 2011), this item warrants critical discussion among all raters (Lunny et al., 2025).

The most striking shortfall was observed for AMSTAR 2 item 7, which concerns justification for the exclusion of studies and was met by only 18.2% (*k* = 12) of reviews. Failure to provide clear rationales for study exclusion undermines transparency and makes it difficult to evaluate whether the included evidence base was assembled consistently and without bias. This pattern has also been reported in prior CCAP-related meta-research (Byrne et al., 2023; Espada et al., 2023; Galán-Luque et al., 2023; García-Fernández et al., 2023; Kwan et al., 2020; Romero et al., 2023; Sánchez De Ribera et al., 2020; Scarlett et al., 2023), with the magnitude observed here suggesting that documenting exclusions remains one of the most critically underdeveloped practices in CCAP-focused meta-analytic systematic reviews.

Taken together, these findings suggest that while current meta-analytic systematic reviews in CCAP often excel in domains related to statistical synthesis and bias assessment, they frequently fall short in early-stage procedural rigor and methodological reporting, particularly with respect to preregistration, comprehensive searching, and justification of study exclusions. These

inconsistencies likely help explain the predominantly *Low* and *Critically Low* overall confidence ratings observed in RQ1. Because AMSTAR 2 heavily weights failures in any critical domain (Shea et al., 2017), strong performance in later-stage synthesis practices cannot compensate for methodological weaknesses introduced during protocol development or study selection. They further highlight targeted opportunities for improving the methodological quality of future meta-analytic systematic reviews in CCAP. Strengthening adherence to preregistration, transparent reporting of search procedures, and explicit justification of study exclusions would likely yield the largest gains in AMSTAR 2 performance and, more importantly, in the credibility of evidence syntheses. Continued editorial emphasis on these elements, alongside clearer author guidelines and reviewer expectations (Bojcic et al., 2024), may be necessary to meaningfully elevate methodological standards within the field.

RQ 3: Relevant Transparency Checklist Items as Sources of Additional Nuance in Methodological Quality Within AMSTAR 2 Critical Domains

Building on RQ2, RQ3 examined whether transparency checklist items derived from PRISMA and its extensions provide additional nuance within AMSTAR 2 critical domains. The results indicate that transparency-related items consistently added meaningful contextual information across domains. Importantly, the transparency items did not contradict AMSTAR 2 ratings but refined their interpretation by revealing substantial heterogeneity in reporting completeness among reviews with similar methodological quality judgments. These results are descriptive and exploratory in nature. For several critical domains, most notably protocol registration (item 2), literature search strategy (item 4), and RoB assessment (item 9), AMSTAR 2 ratings encompassed reviews with markedly different levels of reporting transparency. This pattern supports the notion that methodological quality and reporting transparency, while conceptually distinct, are empirically intertwined and jointly informative for evaluating the trustworthiness of meta-analytic systematic reviews (Blümle et al., 2014; Lunny et al., 2025; Vazire, 2017).

For AMSTAR 2 item 2, approximately half of the reviews (k = 34) fulfilled the criteria, and 13.6% (k = 9) fulfilled it partially, yet transparency items revealed wide variation in protocol accessibility, amendment reporting, and documentation of

search updates. Approximately three-quarters of reviews reported preregistration or protocol use (72.7%; *k* = 48), with 65.1% (*k* = 43) being a publicly accessible protocol, mostly via PROSPERO or OSF. As mentioned above (see Section RQ2: Methodological Quality Within AMSTAR 2 Critical Domains) this preregistration rate reflects an encouraging result as it is higher than findings from previous CCAP meta-reviews (0%–40%; Byrne et al., 2023; Cortese et al., 2019; Espada et al., 2023; Galán-Luque et al., 2023; Kwan et al., 2020; Sánchez De Ribera et al., 2020; Scarlett et al., 2023) as well as broader trends in psychological research, where preregistration remains uncommon despite ongoing advocacy (López-Nicolás et al., 2022; Polanin et al., 2020; Sandoval-Lentisco et al., 2024; 2025).

However, transparency regarding protocol deviations remained limited. Only 37.0% (*k* = 17) of reviews reported protocol amendments, and although most of these justified the changes (88.2%; *k* = 15), fewer than half specified when the amendments occurred. This omission limits the ability to determine whether changes were made before or after critical review stages, such as data extraction or analysis, and prevents distinguishing between confirmatory and exploratory analyses (Nosek et al., 2019), thereby reducing the value of preregistration as a safeguard against undisclosed flexibility (Bakker et al., 2020). Given evidence that undisclosed deviations are common in psychological meta-analytic systematic reviews (Sandoval-Lentisco et al., 2025), it is likely that the true prevalence of protocol changes in the present sample was higher than reported. Transparency items, therefore, added important nuance to AMSTAR 2 ratings by clarifying the extent to which preregistered plans were verifiable and deviations traceable. This is particularly relevant given the lack of clear guidance in AMSTAR 2 item 2 on how to evaluate cases in which deviations from protocols are not explicitly reported (De Santis et al., 2023).

A similar pattern emerged for literature search strategies. More than half (*k* = 35) of the reviews partially met AMSTAR 2 item 4, and only 7.6% (*k* = 5) fully met the AMSTAR 2 criteria for this domain. Broad indicators of transparency, such as databases searched, were fully reported (100%; *k* = 66), consistent with previous research in psychology (López-Nicolás et al., 2022; Steil et al., 2022; Sandoval-Lentisco et al., 2024). However, essential details for reproducibility, such as the use and justification of search limits, the derivation of search terms, and peer review of

search strategies, were frequently unreported. This is in line with previous research, rendering the reporting of search strategies within (meta-analytic) systematic reviews insufficient and not reproducible (Koffel & Rethlefsen, 2016; Rethlefsen et al., 2024). These omissions do not necessarily indicate poor methodology but substantially constrain independent appraisal and replication of the search process. Transparency items, therefore, help contextualize AMSTAR 2 judgments by clarifying how “comprehensiveness” was operationalized and documented. This is particularly relevant given prior criticism that AMSTAR 2 item 4 is susceptible to subjective interpretation regarding what constitutes a comprehensive search (Pieper et al., 2018a). More transparent reporting may reduce this subjectivity by making methodological decisions explicit. Finally, it remains unclear whether trial registries need to be searched in all cases. While it is common practice to register randomized controlled trials (RCTs), this is not the case for non-randomized studies of interventions (NRSIs; De Santis et al., 2023), which may explain the small number of k = 10 applicable cases in the present sample.

In domains with low AMSTAR 2 compliance, transparency items identified specific reporting shortcomings underlying unfavorable confidence ratings. For justification of excluded studies (AMSTAR 2 item 7; *Yes* = 18.2%; k = 12), few reviews reported general reasons for exclusion within the transparency checklist rating (13.6%; k = 9). Notably, these two ratings should align, as both criteria require the same information. However, the absence of double-coding for AMSTAR 2 may have contributed to discrepancies between the two assessments (see Section Strengths and Limitations of the Present Meta-Review for further information). Most reviews in turn provided only aggregated exclusion reasons in a PRISMA flow diagram format (78.8%; k = 52; Page et al., 2021b). Though PRISMA does require referencing seemingly eligible studies, along with a rationale for their exclusion (Page et al., 2021a), most did not supply a full list of exclusions or study-level justifications as would also be required by AMSTAR 2 for a *Partially Yes* or *Yes* rating, respectively (Shea et al., 2017). This practice limits both interpretability and auditability and has been consistently reported in prior CCAP-related meta-research (Byrne et al., 2023; Espada et al., 2023; Galán-Luque et al., 2023; García-Fernández et al., 2023; Kwan et al., 2020; Romero et al., 2023; Sánchez De Ribera et al., 2020; Scarlett et al., 2023). Transparency items, therefore, not only confirmed

low AMSTAR 2 performance but also specified concrete reporting practices requiring improvement.

This gap may partly reflect limited awareness of AMSTAR 2 requirements among review authors. These findings align with a recent meta-scientific study by Puljak et al. (2025), who found that peer reviewers of systematic reviews predominantly used PRISMA, while rarely using AMSTAR 2. Their recommendations include requiring authors to submit key reporting materials at submission (PRISMA items, protocol or registration, and an AMSTAR 2 self-assessment), editors performing relevant pre-checks upon submission, providing structured reviewer prompts focused on validity, interpretation, and transparency, and involving both subject-matter and methodological reviewers, thereby highlighting the need for more unified evaluation frameworks (Puljak et al., 2025). However, another study by Bojcic et al. (2024) found that even when review authors explicitly stated the use of AMSTAR 2, only approximately one-third of their sample (k = 45) fulfilled item 7, and overall methodological quality remained very poor (Bojcic et al., 2024). This finding may appear surprising, as AMSTAR 2 item 7 could be considered one of the more straightforward criteria to fulfill. However, some criticism has been raised regarding item 7, particularly concerning whether the required exclusion list should be limited to the full text assessment stage or should include all studies excluded during the abstract screening stage (De Santis et al., 2023).

The added value of transparency items was most pronounced for AMSTAR 2 item 9. Although most reviews fulfilled item 9 (75.8%; k = 50) and were rated positively for conducting a RoB assessment (80.3%; k = 53), nearly one-fifth (18.2%; k = 12) did not report any RoB assessment. This pattern mirrors findings in cerebral palsy meta-research, where reporting of RoB often exceeds methodological rigor in conduct (Kolaski et al., 2024b). Transparency items revealed substantial heterogeneity in how RoB assessments were implemented and documented. Among the reviews that conducted a RoB assessment, 15.1% (k = 8) did not present the results. In addition, variability in reviewer involvement, discrepancy resolution procedures, justification of judgments, and reporting of bias detection thresholds suggests that reviews with similar AMSTAR 2 ratings may differ considerably in reproducibility and interpretability. Comparable patterns

emerged for transparency items relevant to AMSTAR 2 items 13 (RoB consideration in results) and 15 (publication bias), where bias assessments were often reported (80.3%; *k* = 53; for RoB; 75.9%; *k* = 50 for reporting bias), but rarely integrated into interpretation (20.8%; *k* = 11). This pattern supports concerns that bias assessments are frequently treated as procedural requirements rather than analytically integrated components of evidence synthesis (Kolaski et al., 2024a).

Transparent reporting, therefore, provides essential insight into whether RoB procedures were sufficiently documented to support independent evaluation (Viswanathan et al., 2017). Because no single RoB tool is suitable for all study designs, design-specific or multiple tools in cases of mixed study designs are often required to ensure valid RoB assessments (Kolaski et al., 2024a). These should be documented fully. Even in domains with high AMSTAR 2 compliance, such as statistical synthesis (Item 11), transparency items revealed incomplete reporting of weighting methods, moderator specifications, and effect size interpretation thresholds. This pattern is consistent with previous research highlighting that details essential for reproducibility, such as meta-analytic weighting methods, often remain undisclosed (López-Nicolás et al., 2022; Sandoval-Lentisco et al., 2024). These omissions may not invalidate conclusions but do limit reproducibility and secondary use of meta-analytic results (Nguyen et al., 2023; Rethlefsen et al., 2024).

Taken together, the findings from RQ3 demonstrate that transparency checklist items meaningfully complement AMSTAR 2 critical domains by revealing reporting heterogeneity within methodological quality ratings. Reviews that appear equivalent under AMSTAR 2 critical domains may nevertheless differ substantially in reporting completeness, reproducibility, and interpretability, dimensions that are essential for scientific credibility but only indirectly captured by methodological appraisal tools (Kolaski et al., 2023). These results reinforce the argument that assessments of methodological quality benefit from concurrent evaluation of transparency-related reporting practices (Matthias et al., 2020; Lorenz et al., 2020; Vazire, 2017). Transparency items can help researchers, reviewers, and clinicians better judge how much confidence to place in reviews that formally meet methodological standards but vary in reporting rigor. Incorporating transparency-focused assessments alongside established quality appraisal tools may therefore

enhance the evaluation and interpretation of meta-analytic systematic reviews in CCAP and beyond (Puljak et al., 2025).

RQ4: Association of Methodological Quality and Statistical Consistency

RQ4 examined whether the methodological quality of meta-analytic systematic reviews, as measured by the AMSTAR 2 sum score, was associated with the statistical consistency of reported NHST results. Because extractable and recalculable test statistics were available for fewer than one-third of the included meta-analytic systematic reviews (23 out of 66; see Appendix G, Figure 5), the analysis was necessarily limited and underpowered. Nonetheless, the analysis was conducted to provide an initial indication of whether methodological quality relates to statistical consistency in CCAP meta-analytic systematic reviews. Contrary to the expectation that higher methodological rigor would coincide with fewer reporting inconsistencies, the exploratory results provided no evidence for such an association.

Among the 23 reviews for which *statcheck* extracted at least one test result, approximately half contained at least one statistical inconsistency. Importantly, *statcheck* detected no gross inconsistencies, those that would change the inferential conclusions, suggesting that inconsistencies were minor. Reviews without inconsistencies had slightly lower average AMSTAR 2 scores than reviews with inconsistencies, but this difference was small and statistically nonsignificant. The effect size (*Hedges' g* = 0.39) was small and imprecise, with a wide CI [-0.45, 1.22]. Overall, these findings suggest that methodological quality, as captured by the AMSTAR 2 sum score, is not meaningfully related to the accuracy of reported NHST results within the available data. Notably, taking into consideration the limited power, the estimated effect size still falls in the small-to-medium range. Unexpectedly, the direction of the effect suggests higher quality scores among reviews with at least one inconsistency. Visual inspection of Figure G5 further indicates that inconsistencies were more often flagged in reviews reporting multiple test statistics, likely reflecting a higher probability of reporting errors in such reviews compared to reviews reporting merely one test statistic.

Several factors may further account for this null finding. First, AMSTAR 2 primarily assesses aspects of systematic review rigor, such as preregistration, comprehensiveness of search strategies, and assessment of bias (Shea et al.,

2017), rather than the accuracy of statistical reporting in the results section. A review may meet methodological best practices yet still contain occasional formatting or transcription errors in reporting statistical tests. Conversely, reviews with lower AMSTAR 2 scores may still report statistical tests accurately if authors are meticulous during manuscript preparation. In this sense, statistical conclusion validity as captured by *statcheck* and methodological quality as captured by AMSTAR 2 may reflect rather different aspects of research quality (Vazire et al., 2022).

Second, the small sample size (k = 23) and the limited number of extractable test statistics within each review severely constrained statistical power to detect associations. Many NHST results were not reported in a *statcheck*-compatible format (APA-style; Nuijten & Epskamp, 2024), were not extractable, or were not reported at all. Insufficient reporting of *p*-values and corresponding test-statistics has been identified as a common limitation of *statcheck*-based analyses (Polanin & Nuijten, 2018). While the absence of NHSTs is unproblematic if none were conducted, many meta-analyses do conduct such tests (Polanin & Pigott, 2015; Nuijten & Polanin, 2020), and problems arise when inferential claims are made without accompanying test statistics. These reporting issues restricted the analytic base and resulted in wide confidence intervals around all estimates.

Future research could benefit from more consistent reporting of NHST statistics in accordance with APA 7 guidelines (Nuijten & Epskamp, 2024). Such standardization would enhance the practicality of automated consistency checks using *statcheck* (Nuijten & Epskamp, 2024). Journals might also encourage the use of *statcheck* during peer review, as a simple and effective way to reduce statistical-reporting errors (Nuijten & Wicherts, 2024). Together, standardized reporting and automated screening can improve reporting quality in CCAP reviews. In parallel, improvements in *statcheck* and similar tools (Brown & Heathers, 2017), such as support for a wider range of statistical formatting, would also increase coverage.

Finally, the absence of gross inconsistencies suggests that, even when discrepancies occur, they rarely alter the substantive interpretation of results. This pattern aligns with broader findings indicating that minor inconsistencies are common across scientific fields, whereas consequential errors are relatively rare (Nuijten & Polanin, 2020; Olsen et al., 2019). Thus, although the methodological

quality of included reviews was generally very low (RQ1–RQ2), these limitations did not appear to correlate with major statistical inaccuracies in the subset examined.

In summary, exploratory findings from RQ4 indicate that methodological quality, as operationalized by AMSTAR 2 sum score, was not associated with the presence or absence of statistical inconsistencies detected by *statcheck*. These results highlight that methodological rigor and statistical reporting accuracy may represent distinct facets of research quality (Vazire et al., 2022). They also underscore the importance of considering multiple dimensions of quality when evaluating the reliability and trustworthiness of evidence synthesis in CCAP. Future work using larger samples or more granular measures of statistical reporting accuracy may help clarify whether relationships between methodological rigor and statistical consistency emerge under different analytic conditions.

Strengths and Limitations of the Present Meta-Review

One strength of this meta-review is its systematic approach to identifying and extracting data from recent high-impact CCAP meta-analytic systematic reviews. By focusing on the field's most frequently read syntheses, which are likely to shape research standards and practice, this meta-review provides an informative baseline of current methodological quality. A further strength lies in the study's novelty: to the author's knowledge, it is the first review to assess methodological quality across AMSTAR 2 critical domains in CCAP meta-analytic systematic reviews while also examining additional nuances related to reporting transparency, as well as its exploration with statistical consistency. Although some analyses deviated from the protocol, these deviations are transparently disclosed in the section "Deviations from Protocol."

As with any study, several limitations should be acknowledged. First, resource constraints precluded full independent double coding of methodological quality items and manual verification of *statcheck* output. The absence of double-coding may have contributed to discrepancies between the reporting transparency assessment (ID 16.02) and AMSTAR 2 item 7, although both capture the same information and should align. This misalignment affected study IDs 046, 173, and 181. For IDs 046 and 173, the review authors did provide a justified exclusion list, but not a full reference list, as would be required to fulfill AMSTAR 2 item 7 (Shea

et al., 2017), reporting only the publication year and first author (et al.) of their excluded studies. This detail was likely overlooked by both coders (SF and JR), leading to incorrect ratings for AMSTAR 2 item 7. For ID 181, AMSTAR 2 item 7 was also incorrectly coded as fulfilled by the thesis author (JR). These errors would have likely been detected through double-coding of methodological quality rating. Nonetheless, interrater reliability of methodological quality was acceptable: a random subsample of six meta-analytic systematic reviews (10%) was double-coded for methodological quality, yielding satisfactory agreement (Shea et al., 2017).

Second, some adaptations to the AMSTAR 2 checklist were made to accommodate non-intervention meta-analytic systematic reviews (see “Coding Procedures”), which may limit generalizability; however, similar adaptations have been proposed for items 3, 4, 9, and 13, as these may not apply equally to all systematic review types (Puljak et al., 2023). Third, AMSTAR 2 items requiring evaluation of primary studies or detailed protocol verification were not implemented, as these procedures exceeded the time-scope of this thesis and have been shown to require substantial resources (Lakens et al., 2017; Maassen et al., 2020; Tricco et al., 2016).

Fourth, the transparency-checklist used in this review is not a published or validated tool; however, it was developed based on previous literature (Lakens et al., 2017; López-Nicolás et al., 2022; Maassen et al., 2020) as well as current PRISMA (Page et al., 2021a) and PRISMA-S (Rethlefsen et al., 2021) guidelines. Because the transparency checklist does not encompass all AMSTAR 2 critical domain requirements, the scope of narrative comparison between the two tools was necessarily limited. Fifth, despite the use of HTML files, *statcheck* extracted test statistics from fewer than one-third of the included meta-analytic systematic reviews, which severely limited the power of the analysis.

Implications

The results of the present meta-review have several implications for research practice, editorial standards, and the evaluation of evidence within the field of CCAP. For authors of meta-analytic systematic reviews, the findings highlight the importance of strengthening methodological rigor, particularly through preregistration with accessible protocols, transparent documentation of

protocol deviations, comprehensive and reproducible search strategies, and complete reference lists with explicit justifications for study exclusions. Adherence to these practices, together with consulting the expertise of methodological reviewers (Pieper et al., 2018b) and improving reporting completeness (Lorenz et al., 2020), is likely to yield substantial gains in methodological quality and AMSTAR 2 confidence ratings. Moreover, statistical information should be reported in a *statcheck*-readable format to facilitate automated evaluation (Nuijten & Epskamp, 2024). For journals and reviewers, the predominance of *Low* and particularly *Critically Low* confidence ratings among reviews published in high-impact outlets underscores the need for more structured and consistent evaluation of methodological quality during peer review (Bojcic et al., 2024). Requiring preregistration protocols, PRISMA checklists, AMSTAR 2 self-assessments at submission, and use of *statcheck*, alongside editorial pre-checks and targeted reviewer prompts, may help align published evidence syntheses with established methodological standards (Bojcic et al., 2024).

Consistent with prior research, the findings demonstrate that adequate appraisal of methodological quality depends on reporting completeness (Matthias et al., 2020). Consequently, although methodological quality and reporting transparency represent complementary but distinct dimensions of review credibility, evaluations of meta-analytic systematic reviews should integrate both methodological appraisal tools and transparency-focused reporting checks (Blümle et al., 2014; Lunny et al., 2025). For the field CCAP and beyond, creation of an integrated evaluation framework encompassing both methodological quality and reporting quality may therefore be beneficial in the peer review process (Puljak et al., 2025). Moreover, at the field level, the applicability of AMSTAR 2 for assessing methodological quality should be improved, and further validation studies for non-intervention meta-analytic systematic reviews should be done, as well as further clarification of ambiguous items (De Santis et al., 2023; Kedzior-De Santis, 2019). Meta-analytic systematic reviews identified with the AMSTAR 2 ratings as having *Low* or *Critically Low* methodological quality could be flagged so that stakeholders and policy developers interpret their results with appropriate caution (De Santis & Matthias, 2023; Luchini et al., 2021; Lunny et al., 2022). Such a classification system could help address previous limitations and assist decision makers in

identifying high-quality (meta-analytic) systematic reviews (Luchini et al., 2021; Lunny et al., 2022). High scientific credibility is particularly crucial in CCAP, where meta-analytic conclusions frequently inform clinical practice, policy decisions, and theoretical development (Bellato et al., 2023; Weisz et al., 2023). Strengthening methodological rigor, reporting transparency, and statistical consistency in CCAP evidence synthesis is, therefore, not only a technical concern but a prerequisite for maintaining the credibility of research in a field that directly affects minors and their caregivers.

Conclusion

This meta-review provides a first comprehensive overview of methodological quality, key aspects of reporting transparency relevant to methodological rigor, and statistical consistency in recent meta-analytic systematic reviews published in leading CCAP journals. Despite the high level of trust placed in research syntheses and their central role in informing CCAP research, clinical practice, and policy, methodological quality was predominantly poor, with frequent failures across AMSTAR 2 critical domains such as item 7 (justified study exclusion), item 4 (comprehensive literature search), and item 13 (consideration of RoB in interpretation of results). The findings further show that reporting transparency adds essential nuance to methodological appraisal: reviews with similar AMSTAR 2 ratings often differed substantially in reporting completeness, reproducibility, and interpretability. Although no clear association emerged between methodological quality and statistical consistency, this likely reflects limited statistical power and underscores the importance of considering multiple, distinct dimensions of research quality. Overall, these results demonstrate that trust in meta-analytic systematic reviews cannot be inferred from publication venue or formal methodology alone but must be actively evaluated and rigorously implemented. Strengthening preregistration, transparent reporting, and consistent methodological appraisal, supported by clearer editorial standards and integrated evaluation frameworks, will be essential for enhancing the credibility of evidence synthesis in CCAP.

References

Aden-Buie, G. (2023). *writexl: Export data frames to Excel "xlsx" format (Version 1.5.1)* [R package]. Comprehensive R Archive Network. https://cran.r-project.org/package=writexl

American Psychological Association. (2020). *Publication manual of the American Psychological Association: The official guide to APA style* (7th ed.). American Psychological Association.

Bakker, M., Van Dijk, A., & Wicherts, J. M. (2012). The Rules of the Game Called Psychological Science. *Perspectives on Psychological Science*, *7*(6), 543–554. https://doi.org/10.1177/1745691612459060

Bakker, M., Veldkamp, C. L. S., Van Assen, M. A. L. M., Crompvoets, E. A. V., Ong, H. H., Nosek, B. A., Soderberg, C. K., Mellor, D., & Wicherts, J. M. (2020). Ensuring the quality and specificity of preregistrations. *PLOS Biology*, *18*(12), e3000937. https://doi.org/10.1371/journal.pbio.3000937

Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research Methods*, *43*(3), 666–678. https://doi.org/10.3758/s13428-011-0089-5

Balshem, H., Helfand, M., Schünemann, H. J., Oxman, A. D., Kunz, R., Brozek, J., Vist, G. E., Falck-Ytter, Y., Meerpohl, J., & Norris, S. (2011). GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology*, *64*(4), 401–406. https://doi.org/10.1016/j.jclinepi.2010.07.015

Bellato, A., Cristea, I. A., Giovane, C. D., Fazel, S., Polanczyk, G. V., Solmi, M., & Larsson, H. (2023). Evidence-based child and adolescent mental health care: The role of high-quality and transparently reported evidence synthesis studies. *JCPP Advances*, *3*(3), e12197. https://doi.org/10.1002/jcv2.12197

Biondi-Zoccai, G. (Ed.). (2016). Umbrella reviews: Evidence synthesis with overviews of reviews and meta-epidemiologic studies. *Springer International Publishing*. https://doi.org/10.1007/978-3-319-25655-9

Blümle, A., Von Elm, E., Antes, G., & Meerpohl, J. J. (2014). Messung und Bewertung der Studienqualität und Berichtsqualität [Measurement and evaluation of study quality and reporting quality]. *Zeitschrift für Evidenz Fortbildung und Qualität im Gesundheitswesen*, *108*(8–9), 495–503. https://doi.org/10.1016/j.zefq.2014.09.022

Bojcic, R., Todoric, M., & Puljak, L. (2024). Most systematic reviews reporting adherence to AMSTAR 2 had critically low methodological quality: a cross-sectional meta-research study. *Journal of Clinical Epidemiology*, *165*, 111210. https://doi.org/10.1016/j.jclinepi.2023.10.026

Boonyasai, R., Hahn, C., Bonnett, D., Fitzell, J., Goldstein, E., Hall, I., Hoadley, K., Lankford, A., Liang, L., Liu, S., Matosky, M., Matthews, T., Morgan, R., Moser, R., Muhuri, P., Pratt, R., Ryskulova, A., Smalley, L., Sripipatana, A., … & Valentine, M. (2022). 2022 National Healthcare Quality and Disparities Report.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2021). Introduction to meta-analysis (2nd ed.). Wiley

Botella, J., & Durán, J. I. (2019). A meta-analytical answer to the crisis of confidence of Psychology. *Anales de Psicología*, *35*(2), 350–356. https://doi.org/10.6018/analesps.35.2.345291

Brembs, B. (2018). Prestigious Science Journals Struggle to Reach Even Average Reliability. *Frontiers in Human Neuroscience*, *12*, 37. https://doi.org/10.3389/fnhum.2018.00037

Brewer, M. B., & Crano, W. D. (2014). Research design and issues of validity. In H. T. Reis & C. M. Judd (Eds.), *Handbook of Research Methods in Social and Personality Psychology* (2nd ed., pp. 11–26). Cambridge University Press.

Brown, N. J. L., & Heathers, J. A. J. (2017). The GRIM Test: A Simple Technique Detects Numerous Anomalies in the Reporting of Results in Psychology. *Social Psychological and Personality Science*, *8*(4), 363–369. https://doi.org/10.1177/1948550616673876

Brown, M., Matson, J. L., & Tevis, C. (2023). History of Clinical Child Psychology. In J. L. Matson (Ed.), *Handbook of Clinical Child Psychology* (pp. 3–17). Springer International Publishing. https://doi.org/10.1007/978-3-031-24926-6_1

Bundesinstitut für Arzneimittel und Medizinprodukte. - ICD-10-GM Version 2026. (o. D.). https://klassifikationen.bfarm.de/icd-10-gm/kode-suche/htmlgm2026/index.htm

Burns, P. B., Rohrich, R. J., & Chung, K. C. (2011). The Levels of Evidence and Their Role in Evidence-Based Medicine. *Plastic & Reconstructive Surgery*, *128*(1), 305–310. https://doi.org/10.1097/prs.0b013e318219c171

Byrne, S., Cobham, V., Richardson, M., & Imuta, K. (2023). Do parents enhance cognitive behavior therapy for youth anxiety? An overview of systematic reviews over time. *Clinical Child and Family Psychology Review*, *26*(3), 773–788. https://doi.org/10.1007/s10567-023-00436-5

Carlson, K. D., & Ji, F. X. (2011). Citing and building on meta-analytic findings: A review and recommendations. *Organizational Research Methods*, *14*(4), 696–717. https://doi.org/10.1177/1094428110384272

Carter, E. C., Schönbrodt, F. D., Gervais, W. M., & Hilgard, J. (2019). Correcting for Bias in Psychology: A Comparison of Meta-Analytic Methods. *Advances in Methods and Practices in Psychological Science*, *2*(2), 115–144. https://doi.org/10.1177/2515245919847196

Chapelle, C., Ollier, E., Bonjean, P., Locher, C., Zufferey, P. J., Cucherat, M., & Laporte, S. (2023). Replication of systematic reviews: is it to the benefit or detriment of methodological quality? *Journal of Clinical Epidemiology*, *162*, 98–106. https://doi.org/10.1016/j.jclinepi.2023.08.012

Chevret, S., Ferguson, N. D., & Bellomo, R. (2018). Are systematic reviews and meta-analyses still useful research? No. *Intensive Care Medicine*, *44*(4), 515–517. https://doi.org/10.1007/s00134-018-5066-3

Chin, J. M., Growns, B., Sebastian, J., Page, M. J., & Nakagawa, S. (2022). The transparency and reproducibility of systematic reviews in forensic science.

Forensic Science International, *340*, 111472. https://doi.org/10.1016/j.forsciint.2022.111472

Clarke, M., & Chalmers, I. (2018). Reflections on the history of systematic reviews. *BMJ Evidence-based Medicine*, *23*(4), 121–122. https://doi.org/10.1136/bmjebm-2018-110968

Cooper, H., Hedges, L. V., & Valentine, J. C. (2019). The handbook of research synthesis and meta-analysis (3rd ed.). Russell Sage Foundation.

Cortese, S., Tomlinson, A., & Cipriani, A. (2019). Meta-review: network meta-analyses in child and adolescent psychiatry. *Journal of the American Academy of Child & Adolescent Psychiatry*, *58*(2), 167–179. https://doi.org/10.1016/j.jaac.2018.07.891

Cristea, I. A., & Naudet, F. (2019). Increase value and reduce waste in research on psychological therapies. *Behaviour Research and Therapy*, *123*, 103479. https://doi.org/10.1016/j.brat.2019.103479

Dal Santo, T. D., Rice, D. B., Amiri, L. S., Tasleem, A., Li, K., Boruff, J. T., Geoffroy, M., Benedetti, A., & Thombs, B. D. (2023). Methods and results of studies on reporting guideline adherence are poorly reported: a meta-research study. *Journal of Clinical Epidemiology*, *159*, 225–234. https://doi.org/10.1016/j.jclinepi.2023.05.017

De Girolamo, G., Dagani, J., Purcell, R., Cocchi, A., & McGorry, P. D. (2012). Age of onset of mental disorders and use of mental health services: Needs, opportunities and obstacles. *Epidemiology and Psychiatric Sciences*, *21*(1), 47–57. https://doi.org/10.1017/S2045796011000746

De Santis, K. K., & Matthias, K. (2023). Different Approaches to Appraising Systematic Reviews of Digital Interventions for Physical Activity Promotion Using AMSTAR 2 Tool: Cross-Sectional Study. *International Journal of Environmental Research and Public Health*, *20*(6), 4689. https://doi.org/10.3390/ijerph20064689

De Santis, K. K., Pieper, D., Lorenz, R.C., Wegewitz, U., Siemens, W., & Matthias, K. (2023). User experience of applying AMSTAR 2 to appraise systematic

reviews of healthcare interventions: a commentary. *BMC Medical Research Methodology 23*(63). https://doi.org/10.1186/s12874-023-01879-8

Dekkers, T. J., Groenman, A. P., Cuijpers, P., Hoekstra, P. J., Luman, M., Orobio De Castro, B., Overbeek, G., Popma, A., Rommelse, N., Salemink, E., Stikkelbroek, Y. A. J., Van Den Hoofdakker, B. J., Van Der Oord, S., & Leijten, P. (2023). Commentary: Why treatment is the best choice for childhood mental disorders – a commentary on Roest et al. (2022). *Journal of Child Psychology and Psychiatry*, *64*(3), 470–473. https://doi.org/10.1111/jcpp.13715

Dilling, H., Mombour, W. & Schmidt, M. H (Eds.). (2015). *Internationale Klassifikation psychischer Störungen: ICD-10 Kapitel V (F) - Klinisch-diagnostische Leitlinien*. Hogrefe.

Draborg, E., Andreasen, J., Nørgaard, B., Juhl, C. B., Yost, J., Brunnhuber, K., Robinson, K. A., & Lund, H. (2022). Systematic reviews are rarely used to contextualise new results—a systematic review and meta-analysis of meta-research studies. *Systematic Reviews*, *11*(1), 189. https://doi.org/10.1186/s13643-022-02062-8

Dougherty, M. R., & Horne, Z. (2022). Citation counts and journal impact factors do not capture some indicators of research quality in the behavioural and brain sciences. *Royal Society Open Science*, *9*(8), 220334. https://doi.org/10.1098/rsos.220334

Ebrahim, S., Bance, S., Athale, A., Malachowski, C., & Ioannidis, J. P. (2015). Meta-analyses with industry involvement are massively published and report no caveats for antidepressants. *Journal of Clinical Epidemiology*, *70*, 155–163. https://doi.org/10.1016/j.jclinepi.2015.08.021

Epskamp, S. (2019). Reproducibility and replicability in a fast-paced methodological world. *Advances in Methods and Practices in Psychological Science*, *2*(2), 145–155. https://doi.org/10.1177/2515245919847421

Espada, J. P., Sánchez-López, A., & Morales, A. (2023). Effectiveness of psychological treatments for depression in childhood and adolescence: A

review of reviews. *Revista de Psicología Clínica con Niños y Adolescentes*, *10*(1). https://doi.org/10.21134/rpcna.2023.10.1.6

Fabiano, N., Gupta, A., Fiedorowicz, J. G., & Solmi, M. (2024). On the value of meta research for early career researchers: A commentary. *JCPP Advances*, *4*(2), e12235. https://doi.org/10.1002/jcv2.12235

Faggion, C. M. (2023). Methodological quality, risk of bias, and reporting quality: A confusion persists. *Journal of Evidence-Based Medicine*, *16*(3), 261–263. https://doi.org/10.1111/jebm.12550

Faggion, C. M., & Diaz, K. T. (2019). Overview authors rarely defined systematic reviews that are included in their overviews. *Journal of Clinical Epidemiology*, *109*, 70–79. https://doi.org/10.1016/j.jclinepi.2019.01.004

Falkai, P., & Wittchen, H. (Eds). (2018). Diagnostisches und Statistisches Manual Psychischer Störungen DSM-5®. Hogrefe. https://doi.org/10.1026/02803-000

Finch, A. J. (Ed.). (2012). *Specialty competencies in clinical child and adolescent psychology*. Oxford University Press.

Firke, S. (2024). *janitor: Simple tools for examining and cleaning dirty data (Version 2.2.1)* [R package]. Comprehensive R Archive Network. https://cran.r-project.org/package=janitor

Francis, G. (2013). Replication, statistical consistency, and publication bias. *Journal of Mathematical Psychology*, *57*(5), 153–169. https://doi.org/10.1016/j.jmp.2013.02.003

Furuya-Kanamori, L., Xu, C., Hasan, S. S., & Doi, S. A. (2021). Quality versus Risk-of-Bias assessment in clinical research. *Journal of Clinical Epidemiology*, *129*, 172–175. https://doi.org/10.1016/j.jclinepi.2020.09.044

Galán-Luque, T., Serrano-Ortiz, M., & Orgilés, M. (2023). Effectiveness of psychological interventions for child and adolescent specific anxiety disorders: A systematic review of systematic reviews and meta-analyses. *Revista de Psicología Clínica con Niños y Adolescentes*, *10*(1). https://doi.org/10.21134/rpcna.2023.10.1.3

García-Fernández, G., Krotter, A., & Udeanu, A. (2023). Effectiveness of psychological interventions for eating disorders in adolescence: An overview of systematic reviews. *Revista de Psicología Clínica con Niños y Adolescentes*, *10*(1). https://doi.org/10.21134/rpcna.2023.10.1.10

Garritty, C., Stevens, A., Hamel, C., Golfam, M., Hutton, B., & Wolfe, D. (2019). Knowledge synthesis in Evidence-Based Medicine. *Seminars in Nuclear Medicine*, *49*(2), 136–144. https://doi.org/10.1053/j.semnuclmed.2018.11.006

Gates, A., Gates, M., Duarte, G., Cary, M., Becker, M., Prediger, B., Vandermeer, B., Fernandes, R. M., Pieper, D., & Hartling, L. (2018). Evaluation of the reliability, usability, and applicability of AMSTAR, AMSTAR 2, and ROBIS: protocol for a descriptive analytic study. *Systematic Reviews*, *7*(1), 85. https://doi.org/10.1186/s13643-018-0746-1

Gates, M., Gates, A., Pieper, D., Fernandes, R. M., Tricco, A. C., Moher, D., Brennan, S. E., Li, T., Pollock, M., Lunny, C., Sepulveda, D., McKenzie, J. E., Scott, S. D., Robinson, K. A., Matthias, K., Bougioukas, K. I., Fusar-Poli, P., Whiting, P., Moss, S. J., & Hartling, L. (2022). Reporting guideline for overviews of reviews of healthcare interventions: Development of the PRIOR statement. *BMJ*, e070849. https://doi.org/10.1136/bmj-2022-070849

Ghai, S., Thériault, R., Forscher, P., Shoda, Y., Syed, M., Puthillam, A., Peng, H. C., Basnight-Brown, D., Majid, A., Azevedo, F., & Singh, L. (2025). A manifesto for a globally diverse, equitable, and inclusive open science. *Communications Psychology*, *3*(1), 16. https://doi.org/10.1038/s44271-024-00179-1

Goodman, S., & Dickersin, K. (2011). Metabias: A Challenge for Comparative Effectiveness Research. *Annals of Internal Medicine*, *155*(1), 61–62. https://doi.org/10.7326/0003-4819-155-1-201107050-00010

Guo, L., Miller, S., Zhou, W., Wei, Z., Ren, J., Huang, X., Xing, X., White, H., & Yang, K. (2025). Critical appraisal of methodological quality and completeness of reporting in Chinese social science systematic reviews

with meta-analysis: A systematic review. *Campbell Systematic Reviews*, *21*(1), e70014. https://doi.org/10.1002/cl2.70014

Guyatt, G. H., Oxman, A. D., Vist, G. E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., Schünemann, H. J., & GRADE Working Group (2008). GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ (Clinical research ed.)*, *336*(7650), 924–926. https://doi.org/10.1136/bmj.39489.470347.AD

Guyatt, G. H., Oxman, A. D., Kunz, R., Brozek, J., Alonso-Coello, P., Rind, D., Devereaux, P. J., Montori, V. M., Freyschuss, B., Vist, G., Jaeschke, R., Williams, J. W., Jr, Murad, M. H., Sinclair, D., Falck-Ytter, Y., Meerpohl, J., Whittington, C., Thorlund, K., Andrews, J., & Schünemann, H. J. (2011). GRADE guidelines 6. Rating the quality of evidence--imprecision. *Journal of Clinical Epidemiology*, *64*(12), 1283–1293. https://doi.org/10.1016/j.jclinepi.2011.01.012

Hardwicke, T. E., Jameel, L., Jones, M., Walczak, E. J., & Magis-Weinberg, L. (2014). Only Human: Scientists, Systems, and Suspect Statistics. *Opticon1826*, *0*(16). https://doi.org/10.5334/opt.ch

Hardwicke, T. E., Serghiou, S., Janiaud, P., Danchev, V., Crüwell, S., Goodman, S. N., & Ioannidis, J. P. A. (2020). Calibrating the scientific ecosystem through meta-research. *Annual Review of Statistics and its Application*, *7*, 11–37. https://doi.org/10.1146/annurev-statistics-031219-041104

He, L., Soh, K. L., Yu, J., Chen, A., & Dong, X. (2023). Effect of family-centered interventions for perinatal depression: an overview of systematic reviews. *Frontiers in Psychiatry*, *14*, 1094360. https://doi.org/10.3389/fpsyt.2023.1094360

Hendriks, F., Kienhues, D., & Bromme, R. (2020). Replication crisis = trust crisis? The effect of successful vs failed replications on laypeople's trust in researchers and research. *Public Understanding of Science*, *29*(3), 270–288. https://doi.org/10.1177/0963662520902383

Higgins, J. P. T., Thompson, S. G., & Spiegelhalter, D. J. (2009). A re-evaluation of random-effects meta-analysis. *Journal of the Royal Statistical Society*

Series A: Statistics in Society, *172*(1), 137–159. https://doi.org/10.1111/j.1467-985X.2008.00552.x

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2024). *Cochrane handbook for systematic reviews of interventions* (6.5). 69 Cochrane. https://www.cochrane.org/authors/handbooks-and-manuals/handbook/current

Hohn, R. E., Slaney, K. L., & Tafreshi, D. (2020). An empirical review of research and reporting practices in psychological meta-analyses. *Review of General Psychology*, *24*(3), 195–209. https://doi.org/10.1177/1089268020918844

Ioannidis, J. P. A. (2010). Meta-research: The art of getting it wrong. *Research Synthesis Methods*, *1*(3–4), 169–184. https://doi.org/10.1002/jrsm.19

Ioannidis, J. P. A. (2012). Why science is not necessarily Self-Correcting. *Perspectives on Psychological Science*, *7*(6), 645–654. https://doi.org/10.1177/1745691612464056

Ioannidis J. P. A. (2016). The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses. *The Milbank Quarterly*, *94*(3), 485–514. https://doi.org/10.1111/1468-0009.12210

Ioannidis, J. P. A. (2018). Meta-research: Why research on research matters. *PLOS Biology*, *16*(3), e2005468. https://doi.org/10.1371/journal.pbio.2005468

Ioannidis, J. P. A., Fanelli, D., Dunne, D. D., & Goodman, S. N. (2015). Meta-research: Evaluation and improvement of research methods and practices. *PLOS Biology*, *13*(10), e1002264. https://doi.org/10.1371/journal.pbio.1002264

Kedzior-De Santis, K. K. (2019). *Is AMSTAR2 an appropriate tool to assess the quality of systematic reviews in psychology?* ZPID (Leibniz Institute for Psychology Information). https://doi.org/10.23668/PSYCHARCHIVES.2479

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV

disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 593. https://doi.org/10.1001/archpsyc.62.6.593

Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Mohr, A. H., IJzerman, H., Nilsonne, G., Vanpaemel, W. & Frank, M. C. (2018). A Practical Guide for Transparency in Psychological Science. *Collabra Psychology*, *4*(1). https://doi.org/10.1525/collabra.158

Koffel, J. B., & Rethlefsen, M. L. (2016). Reproducibility of search strategies is poor in systematic reviews published in high-impact pediatrics, cardiology and surgery journals: A cross-sectional study. *PLOS ONE*, *11*(9), e0163309. https://doi.org/10.1371/journal.pone.0163309

Kolaski, K., Clarke, M. & Logan, L. R. (2024a). Analysis of risk of bias assessments in a sample of intervention systematic reviews, Part II: focus on risk of bias tools reveals few meet current appraisal standards. *Journal of Clinical Epidemiology*, *174*, 111460. https://doi.org/10.1016/j.jclinepi.2024.111460

Kolaski, K., Clarke, M., Rathnayake, D., & Romeiser Logan, L. (2024b). Analysis of risk of bias assessments in a sample of intervention systematic reviews, part I: many aspects of conduct and reporting need improvement. *Journal of Clinical Epidemiology*, *174*, 111480. https://doi.org/10.1016/j.jclinepi.2024.111480

Kolaski, K., Logan, L. R., & Ioannidis, J. P. A. (2023). Guidance to best tools and practices for systematic reviews. *Systematic Reviews*, *12*(1), 96. https://doi.org/10.1186/s13643-023-02255-9

Korbmacher, M., Azevedo, F., Pennington, C. R., Hartmann, H., Pownall, M., Schmidt, K., Elsherif, M., Breznau, N., Robertson, O., Kalandadze, T., Yu, S., Baker, B. J., O'Mahony, A., Olsnes, J. Ø.-., Shaw, J. J., Gjoneska, B., Yamada, Y., Röer, J. P., Murphy, J., … & Evans, T. (2023). The replication crisis has led to positive structural, procedural, and community changes. *Communications Psychology*, *1*(1), 3. https://doi.org/10.1038/s44271-023-00003-2

Kung, J., Chiappelli, F., Cajulis, O. O., Avezova, R., Kossan, G., Chew, L., & Maida, C. A. (2010). From Systematic Reviews to Clinical Recommendations for

Evidence-Based Health Care: Validation of Revised Assessment of Multiple Systematic Reviews (R-AMSTAR) for Grading of Clinical Relevance. *The Open Dentistry Journal*, *4*, 84–91. https://doi.org/10.2174/1874210601004020084

Kwan, I., Dickson, K., Richardson, M., MacDowall, W., Burchett, H., Stansfield, C., Brunton, G., Sutcliffe, K., & Thomas, J. (2020). Cyberbullying and children and young people's mental health: A systematic map of systematic reviews. *Cyberpsychology, Behavior, and Social Networking*, *23*(2), 72–82. https://doi.org/10.1089/cyber.2019.0370

Lakens, D., Page-Gould, E., Van Assen, M. A. L. M., Spellman, B., Schönbrodt, F. D., Hasselman, F., Corker, K. S., Grange, J. A., Sharples, A., Cavender, C., Augusteijn, H. E. M., Augusteijn, H., Gerger, H., Locher, C., Miller, I. D., Anvari, F., & Scheel, A. M. (2017). *Examining the reproducibility of meta-analyses in Psychology: A preliminary report*. https://doi.org/10.31222/osf.io/xfbjf

Leclercq, V., Beaudart, C., Ajamieh, S., Tirelli, E., & Bruyère, O. (2020a). Methodological quality of meta-analyses indexed in PsycINFO: leads for enhancements: a meta-epidemiological study. *BMJ open*, *10*(8), e036349. https://doi.org/10.1136/bmjopen-2019-036349

Leclercq, V., Beaudart, C., Tirelli, E., & Bruyère, O. (2020b). Psychometric measurements of AMSTAR 2 in a sample of meta-analyses indexed in PsycINFO. *Journal of Clinical Epidemiology*, *119*, 144–145. https://doi.org/10.1016/j.jclinepi.2019.10.005

Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS medicine*, *6*(7), e1000100. https://doi.org/10.1371/journal.pmed.1000100

López-Nicolás, R., López-López, J. A., Rubio-Aparicio, M., & Sánchez-Meca, J. (2022). A meta-review of transparency and reproducibility-related reporting practices in published meta-analyses on clinical psychological interventions

(2000–2020). *Behavior Research Methods*, *54*(1), 334–349. https://doi.org/10.3758/s13428-021-01644-z

Lorenz, R. C., Matthias, K., Pieper, D., Wegewitz, U., Morche, J., Nocon, M., ... & Jacobs, A. (2019). A psychometric study found AMSTAR 2 to be a valid and moderately reliable appraisal tool. *Journal of Clinical Epidemiology*, *114*, 133-140.

Lorenz, R. C., Matthias, K., Pieper, D., Wegewitz, U., Morche, J., Nocon, M., ... & Jacobs, A. (2020). AMSTAR 2 overall confidence rating: lacking discriminating capacity or requirement of high methodological quality?. *Journal of Clinical Epidemiology*, *119*, 142 – 144

Lozada-Martinez, I. D., Neira-Rodado, D., Martinez-Guevara, D., Cruz-Soto, H. S., Sanchez-Echeverry, M. P., & Liscano, Y. (2025). Why is it important to implement meta-research in universities and institutes with medical research activities? *Frontiers in Research Metrics and Analytics*, *10*, 1497280. https://doi.org/10.3389/frma.2025.1497280

Luchini, C., Veronese, N., Nottegar, A., Shin, J. I., Gentile, G., Granziol, U., Soysal, P., Alexinschi, O., Smith, L., & Solmi, M. (2021). Assessing the quality of studies in meta-research: Review/guidelines on the most important quality assessment tools. *Pharmaceutical Statistics*, *20*(1), 185–195. https://doi.org/10.1002/pst.2068

Lund, H., Brunnhuber, K., Juhl, C., Robinson, K., Leenaars, M., Dorch, B. F., Jamtvedt, G., Nortvedt, M. W., Christensen, R., & Chalmers, I. (2016). Towards evidence based research. *BMJ (Clinical research ed.)*, *355*, i5440. https://doi.org/10.1136/bmj.i5440

Lunny, C., Jain, N., Nazari, T., Kosaner-Kließ, M., Santos, L., Goodman, I., Osman, A. A. M., Berrone, S., Dadam, M. N., Brenna, C. T. A., Hussein, H., Dahdal, G., A, D. C., Ferri, N., Kanji, S., Chi, Y., Pieper, D., Shea, B., Parker, A., … & Tricco, A. C. (2025). Exploring the methodological quality and risk of bias in 200 systematic reviews: A comparative study of ROBIS and AMSTAR-2 tools. *Research Synthesis Methods*, 1–30. https://doi.org/10.1017/rsm.2025.10032

Lunny, C., Kanji, S., Thabet, P., Haidich, A., Bougioukas, K. I., & Pieper, D. (2024). Assessing the methodological quality and risk of bias of systematic reviews: primer for authors of overviews of systematic reviews. *BMJ Medicine*, *3*(1), e000604. https://doi.org/10.1136/bmjmed-2023-000604

Lunny, C., Whitelaw, S., Chi, Y., Zhang, J. H., Ferri, N., Kanji, S., Pieper, D., Shea, B., Dourka, J., Veroniki, A., Arden, C., Pham, B., Reid, E., Bagheri, E., & Tricco, A. (2022). Decision makers find it difficult to compare and select similar systematic reviews based on quality, methods and results: a cross-sectional survey. 10.21203/rs.3.rs-2416773/v1.

Maassen, E., Van Assen, M. A. L. M., Nuijten, M. B., Olsson-Collentine, A., & Wicherts, J. M. (2020). Reproducibility of individual effect sizes in meta-analyses in psychology. *PLOS ONE*, *15*(5), e0233107. https://doi.org/10.1371/journal.pone.0233107

Matson, J. L. (Ed.). (2023). *Handbook of clinical child psychology: Integrating theory and research into practice*. Springer International Publishing. https://doi.org/10.1007/978-3-031-24926-6

Matthias, K., Rissling, O., Pieper, D., Morche, J., Nocon, M., Jacobs, A., ... & Lorenz, R. C. (2020). The methodological quality of systematic reviews on the treatment of adult major depression needs improvement according to AMSTAR 2: a cross-sectional study. *Heliyon*, *6*(9).

McGee, R. G., & Dawson, A. C. (2020). Fake news and fake research: Why meta-research matters more than ever. *Journal of Paediatrics and Child Health*, *56*(12), 1868–1871. https://doi.org/10.1111/jpc.15237

McKiernan, E. C., Schimanski, L. A., Muñoz Nieves, C., Matthias, L., Niles, M. T., & Alperin, J. P. (2019). Use of the Journal Impact Factor in academic review, promotion, and tenure evaluations. *eLife*, *8*, e47338. https://doi.org/10.7554/eLife.47338

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015

statement. *Systematic Reviews*, *4*(1). https://doi.org/10.1186/2046-4053-4-1

Møller, M. H., Ioannidis, J. P. A. & Darmon, M. (2018). Are systematic reviews and meta-analyses still useful research? We are not sure. *Intensive Care Medicine*, *44*(4), 518–520. https://doi.org/10.1007/s00134-017-5039-y

Moreau, D., & Corballis, M. C. (2019). When averaging goes wrong: The case for mixture model estimation in psychological science. *Journal of Experimental Psychology: General*, *148*(9), 1615–1627. https://doi.org/10.1037/xge0000504

Moreau, D., & Gamble, B. (2022). Conducting a meta-analysis in the age of open science: Tools, tips, and practical recommendations. *Psychological Methods*, *27*(3), 426–432. https://doi.org/10.1037/met0000351

Muka, T., Glisic, M., Milic, J., Verhoog, S., Bohlius, J., Bramer, W., Chowdhury, R., & Franco, O. H. (2020). A 24-step guide on how to design, conduct, and successfully publish a systematic review and meta-analysis in medical research. *European Journal of Epidemiology*, *35*(1), 49–60. https://doi.org/10.1007/s10654-019-00576-5

Müller, K. (2020). *here: A simpler way to find your files (Version 1.0.1)* [R package]. Comprehensive R Archive Network. https://cran.r-project.org/package=here

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie Du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*(1), 0021. https://doi.org/10.1038/s41562-016-0021

Murad, M. H., Asi, N., Alsawas, M., & Alahdab, F. (2016). New evidence pyramid. *Evidence Based Medicine*, *21*(4), 125–127. https://doi.org/10.1136/ebmed-2016-110401

Murad, M. H., Montori, V. M., Ioannidis, J. P. A., Jaeschke, R., Devereaux, P. J., Prasad, K., Neumann, I., Carrasco-Labra, A., Agoritsas, T., Hatala, R., Meade, M. O., Wyer, P., Cook, D. J., & Guyatt, G. (2014). How to read a

systematic review and meta-analysis and apply the results to patient care. *JAMA*, *312*(2), 171. https://doi.org/10.1001/jama.2014.5559

Nakagawa, S., & Lagisz, M. (2024). *The ABC of academic writing: Non-native speakers' perspective*.

Nguyen, P., McKenzie, J. E., Alqaidoom, Z., Hamilton, D. G., Moher, D., & Page, M. J. (2025). Reproducibility of meta-analytic results in systematic reviews of interventions: meta-research study. *BMJ Medicine*, *4*(1), e002024. https://doi.org/10.1136/bmjmed-2025-002024

Nguyen, P., McKenzie, J. E., Hamilton, D. G., Moher, D., Tugwell, P., Fidler, F. M., Haddaway, N. R., Higgins, J. P. T., Kanukula, R., Karunananthan, S., Maxwell, L. J., McDonald, S., Nakagawa, S., Nunan, D., Welch, V. A., & Page, M. J. (2023). Systematic reviewers' perspectives on sharing review data, analytic code, and other materials: A survey. *Cochrane Evidence Synthesis and Methods*, *1*(2), e12008. https://doi.org/10.1002/cesm.12008

Nosek, B. A., Beck, E. D., Campbell, L., Flake, J. K., Hardwicke, T. E., Mellor, D. T., van 't Veer, A. E., & Vazire, S. (2019). Preregistration Is Hard, And Worthwhile. *Trends in Cognitive Sciences*, *23*(10), 815–818. https://doi.org/10.1016/j.tics.2019.07.009

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Struhl, M. K., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology*, *73*, 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

Nuijten, M. B., & Epskamp, S. (2024). statcheck: Extract statistics from articles and recompute p-values [Computer software]. https://CRAN.R-project.org/package=statcheck

Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior Research Methods*, *48*(4), 1205–1226. https://doi.org/10.3758/s13428-015-0664-2

Nuijten, M. B., & Polanin, J. R. (2020). “statcheck”: Automatically detect statistical reporting inconsistencies to increase reproducibility of meta-analyses. *Research Synthesis Methods*, *11*(5), 574–579. https://doi.org/10.1002/jrsm.1408

Nuijten, M. B., & Wicherts, J. M. (2024). Implementing *statcheck* during peer review is related to a steep decline in statistical-reporting inconsistencies. *Advances in Methods and Practices in Psychological Science*, *7*(2), 25152459241258945. https://doi.org/10.1177/25152459241258945

OCEBM Levels of Evidence Working Group. (2011). *The Oxford 2011 levels of evidence* [Webpage]. Centre for Evidence-Based Medicine, University of Oxford. https://www.cebm.ox.ac.uk/resources/levels-of-evidence/ocebm-levels-of-evidence

O’Kelly, F., DeCotiis, K., Aditya, I., Braga, L. H., & Koyle, M. A. (2020). Assessing the methodological and reporting quality of clinical systematic reviews and meta-analyses in paediatric urology: can practices on contemporary highest levels of evidence be built?. *Journal of Pediatric Urology*, *16*(2), 207-217.

Olsen, J., Mosen, J., Voracek, M., & Kirchler, E. (2019). Research practices and statistical reporting quality in 250 economic psychology master’s theses: a meta-research investigation. *Royal Society Open Science*, *6*(12), 190738. https://doi.org/10.1098/rsos.190738

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

OpenAI. (2024). *ChatGPT (Version 4)* [Large language model]. OpenAI. https://chatgpt.com

OpenAI. (2025). *ChatGPT (Version 5)* [Large language model]. OpenAI. https://chatgpt.com

O’Shea, J., Jenkins, R., Nicholls, D., Downs, J., & Hudson, L. D. (2025). Prevalence, severity and risk factors for mental disorders among sexual and gender minority young people: A systematic review of systematic

reviews and meta-analyses. *European Child & Adolescent Psychiatry*, *34*(3), 959–982. https://doi.org/10.1007/s00787-024-02552-1

Oxman, A. D., & Guyatt, G. H. (1991). Validation of an index of the quality of review articles. *Journal of Clinical Epidemiology*, *44*(11), 1271–1278. https://doi.org/10.1016/0895-4356(91)90160-b

Page, M. J., Altman, D. G., Shamseer, L., McKenzie, J. E., Ahmadzai, N., Wolfe, D., Yazdi, F., Catalá-López, F., Tricco, A. C., & Moher, D. (2018). Reproducible research practices are underused in systematic reviews of biomedical interventions. *Journal of Clinical Epidemiology*, *94*, 8–18. https://doi.org/10.1016/j.jclinepi.2017.10.017

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … & Moher, D. (2021a). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *Systematic Reviews*, *10*(1). https://doi.org/10.1186/s13643-021-01626-4

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., McGuinness, L. A., … & McKenzie, J. E. (2021b). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ* (Clinical research ed.), *372*, n160. https://doi.org/10.1136/bmj.n160

Peels, R., & Bouter, L. (2021). Replication and trustworthiness. *Accountability in Research*, *30*(2), 77–87. https://doi.org/10.1080/08989621.2021.1963708

Peña, A. M., Ladd, C. C., Anderson, J. M., Torgerson, T., Hartwell, M., Johnson, B. S., McMurray, M., & Vassar, M. (2022). An analysis of the evidence underpinning the American Urologic Association Clinical Practice Guidelines. *Urology*, *161*, 42–49. https://doi.org/10.1016/j.urology.2021.12.019

Pennington, C. R. (2023). *A Student's Guide to Open Science: Using the Replication Crisis to Reform Psychology*. Open University Press.

Perry, R., Whitmarsh, A., Leach, V., & Davies, P. (2021). A comparison of two assessment tools used in overviews of systematic reviews: ROBIS versus AMSTAR-2. *Systematic Reviews*, *10*(1), 273. https://doi.org/10.1186/s13643-021-01819-x

Pieper, D., Koensgen, N., Breuing, J., Ge, L. & Wegewitz, U. (2018b). How is AMSTAR applied by authors – a call for better reporting. *BMC Medical Research Methodology*, *18*(1), 56. https://doi.org/10.1186/s12874-018-0520-z

Pieper, D., Puljak, L., González-Lorenzo, M. & Minozzi, S. (2018a). Minor differences were found between AMSTAR 2 and ROBIS in the assessment of systematic reviews including both randomized and nonrandomized studies. *Journal of Clinical Epidemiology*, *108*, 26–33. https://doi.org/10.1016/j.jclinepi.2018.12.004

Polanczyk, G. V., Salum, G. A., Sugaya, L. S., Caye, A., & Rohde, L. A. (2015). Annual Research Review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry*, *56*(3), 345–365. https://doi.org/10.1111/jcpp.12381

Polanin, J. R., Hennessy, E. A., & Tsuji, S. (2020). Transparency and reproducibility of meta-analyses in Psychology: A meta-review. *Perspectives on Psychological Science*, *15*(4), 1026–1041. https://doi.org/10.1177/1745691620906416

Polanin, J. R., & Nuijten, M. B. (2018). Verifying the accuracy of statistical significance testing in Campbell Collaboration systematic reviews through the use of the R package statcheck. *Campbell Systematic Reviews*, *14*(1), 1–36. https://doi.org/10.4073/csrm.2018.1

Polanin, J. R., & Pigott, T. D. (2015). The use of meta-analytic statistical significance testing. *Research Synthesis Methods*, *6*(1), 63–73. https://doi.org/10.1002/jrsm.1124

Pollock M., Fernandes R. M., Becker L. A., Pieper D., & Hartling L. (2020). Chapter V: Overviews of Reviews [last updated August 2023]. In: Higgins J. P., Thomas J., Chandler J., Cumpston M., Li T., Page M.J., et al., editor(s). *Cochrane Handbook for Systematic Reviews of Interventions* version 6.5. Cochrane, 2024. Available from *cochrane.org/handbook*.

Puljak, L., Bala, M. M., Mathes, T., Pericic, T. P., Wegewitz, U., Faggion, C. M., Matthias, K., Storman, D., Zajac, J., Rombey, T., Bruschettini, M., & Pieper, D. (2023). AMSTAR 2 is only partially applicable to systematic reviews of non-intervention studies: a meta-research study. *Journal of Clinical Epidemiology*, *163*, 11–20. https://doi.org/10.1016/j.jclinepi.2023.08.021

Puljak L., Pintur S., Rombey T., Lockwood C., & Pieper D. (2025). Use of structured tools by peer reviewers of systematic reviews: a cross-sectional study reveals high familiarity with PRISMA but limited use of other tools. *Journal of Clinical Epidemiology*. https://doi.org/10.1016/j.jclinepi.2025.112084.

Qi, W., Huang, S., Zhao, J., Cui, X., Wei, Z., Cui, G., Guo, Q., & Hu, J. (2024). The preventive effect of psychological and psychosocial interventions on postpartum depression: An overview of systematic reviews. *Journal of Psychiatric Research*, *182*, 21–33. https://doi.org/10.1016/j.jpsychires.2024.11.050

R Core Team. (2021). *R: A language and environment for statistical computing (Version 4.1.0)* [Computer software]. R Foundation for Statistical Computing. https://www.r-project.org

Reitsma J. B., Rutjes A. W. S., Whiting P., Vlassov V. V., Leeflang M. M. G., & Deeks J. J. (2009). Chapter 9: Assessing methodological quality. In: Deeks J. J., Bossuyt P. M., & Gatsonis, C. (editors), Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 1.0.0. The Cochrane Collaboration, 2009. Available from: http://srdta.cochrane.org/.

Rethlefsen, M. L., Brigham, T. J., Price, C., Moher, D., Bouter, L. M., Kirkham, J. J., Schroter, S., & Zeegers, M. P. (2024). Systematic review search strategies are poorly reported and not reproducible: a cross-sectional metaresearch study. *Journal of Clinical Epidemiology*, *166*, 111229. https://doi.org/10.1016/j.jclinepi.2023.111229

Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., & Koffel, J. B. (2021). PRISMA-S: an extension to the PRISMA Statement for reporting literature searches in systematic reviews. *Systematic Reviews*, *10*(1), 39. https://doi.org/10.1186/s13643-020-01542-z

Romero, E., Álvarez-Voces, M., Díaz-Vázquez, B., & López-Romero, L. (2023). What works in the psychological treatment of child conduct problems? An umbrella review of meta-analytic studies. *Revista de Psicología Clínica con Niños y Adolescentes*, *10*(1). https://doi.org/10.21134/rpcna.2023.10.1.1

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. John Wiley & Sons.

Sacco, R., Camilleri, N., Eberhardt, J., Umla-Runge, K., & Newbury-Birch, D. (2024). A systematic review and meta-analysis on the prevalence of mental disorders among children and adolescents in Europe. *European Child & Adolescent Psychiatry*, *33*(9), 2877–2894. https://doi.org/10.1007/s00787-022-02131-2

Sala, A., Porcaro, L., & Gómez, E. (2024). Social media use and adolescents' mental health and well-being: An umbrella review. *Computers in Human Behavior Reports*, *14*, 100404. https://doi.org/10.1016/j.chbr.2024.100404

Sánchez De Ribera, O., Trajtenberg, N., & Christensen, L. S. (2020). Evaluating the quality of meta-analytical reviews using the AMSTAR-2: A systematic review of meta-analytical reviews regarding child sexual abuse interventions. *Child Abuse & Neglect*, *104*, 104463. https://doi.org/10.1016/j.chiabu.2020.104463

Sandoval-Lentisco, A., López-Nicolás, R., Tortajada, M., López-López, J. A., & Sánchez-Meca, J. (2024). Transparency in cognitive training meta-

analyses: A meta-review. *Neuropsychology Review*. https://doi.org/10.1007/s11065-024-09638-2

Sandoval-Lentisco, A., Tortajada, M., López-Nicolás, R., López-López, J. A., Wagenmakers, E. J., Sánchez-Meca, J., & Hardwicke, T. E. (2025). Preregistration of psychology meta-analyses: A cross-sectional study of prevalence and practice. *Advances in Methods and Practices in Psychological Science*, *8*(1), 25152459241300113. https://doi.org/10.1177/25152459241300113

Scarlett, H., Moirangthem, S., & Van Der Waerden, J. (2023). The impact of paternal mental illness on child development: An umbrella review of systematic reviews and meta-analyses. *European Child & Adolescent Psychiatry*. https://doi.org/10.1007/s00787-023-02261-1

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D. A. (2017). AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, j4008. https://doi.org/10.1136/bmj.j4008

Siegel, M., Eder, J. S. N., Wicherts, J. M., & Pietschnig, J. (2022). Times are changing, bias isn't: A meta-meta-analysis on publication bias detection practices, prevalence rates, and predictors in industrial/organizational psychology. *Journal of Applied Psychology*, *107*(11), 2013–2039. https://doi.org/10.1037/apl0000991

Siegel, M., Fanninger, S., Riedel, J., & Zemp, M. (2025). Transparency, quality, and statistical consistency of meta-analytic systematic reviews in clinical child and adolescent psychology (2022–2024): Study protocol for a meta-review. *Frontiers in Psychology, 16*. https://doi.org/10.3389/fpsyg.2025.1535606

Siddaway, A. P., Wood, A. M., & Hedges, L. V. (2019). How to do a systematic review: A best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annual Review of Psychology*, *70*(1), 747–770.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Steil, A. V., Dias, N. M., Lopes, F. M., Silva, M. L. B. D., Bousfield, A. B. D. S., & De Luca Canto, G. (2022). Reporting characteristics of systematic reviews in Psychology: A scoping review. *Journal of Health Psychology*, *27*(13), 2964–2981. https://doi.org/10.1177/13591053221074592

Stevens, J. R. (2017). Replicability and Reproducibility in Comparative Psychology. *Frontiers in Psychology*, *8*, 862. https://doi.org/10.3389/fpsyg.2017.00862

Stone, J. C., Leonardi-Bee, J., Barker, T. H., Sears, K., Klugar, M., Munn, Z., & Aromataris, E. (2024). Common tool structures and approaches to risk of bias assessment: implications for systematic reviewers. *JBI Evidence Synthesis*, *22*(3), 389–393. https://doi.org/10.11124/jbies-23-00463

Strawbridge, R., Sharma, D., Kisely, S., Cristea, I. A., Young, A. H., & Kaufman, K. R. (2025). Enhancing the quality of systematic reviews and meta-analyses. *BJPsych Open*, *11*(6), e266. https://doi.org/10.1192/bjo.2025.10876

Solmi, M., Radua, J., Olivola, M., Croce, E., Soardo, L., Salazar De Pablo, G., Il Shin, J., Kirkbride, J. B., Jones, P., Kim, J. H., Kim, J. Y., Carvalho, A. F., Seeman, M. V., Correll, C. U., & Fusar-Poli, P. (2022). Age at onset of mental disorders worldwide: Large-scale meta-analysis of 192 epidemiological studies. *Molecular Psychiatry*, *27*(1), 281–295. https://doi.org/10.1038/s41380-021-01161-7

Sotola, L. K. (2022). Garbage In, Garbage Out? Evaluating the Evidentiary Value of Published Meta-analyses Using Z-Curve Analysis. *Collabra Psychology*, *8*(1). https://doi.org/10.1525/collabra.32571

Świątkowski, W., & Dompnier, B. (2017). Replicability Crisis in Social Psychology: Looking at the Past to Find New Pathways for the Future. *International Review of Social Psychology*, *30*(1), 111. https://doi.org/10.5334/irsp.66

Vazire, S. (2017). Quality uncertainty erodes trust in science. *Collabra Psychology*, *3*(1). https://doi.org/10.1525/collabra.74

Vazire, S. (2018). Implications of the credibility revolution for productivity, creativity, and progress. *Perspectives on Psychological Science*, *13*(4), 411–417. https://doi.org/10.1177/1745691617751884

Vazire, S., Schiavone, S. R., & Bottesini, J. G. (2022). Credibility Beyond Replicability: Improving the Four Validities in Psychological Science. *Current Directions in Psychological Science*, *31*(2), 162–168. https://doi.org/10.1177/09637214211067779

Viswanathan, M., Patnode, C. D., Berkman, N. D., Bass, E. B., Chang, S., Hartling, L., Murad, M. H., Treadwell, J. R., & Kane, R. L. (2017). Recommendations for assessing the risk of bias in systematic reviews of health-care interventions. *Journal of Clinical Epidemiology*, *97*, 26–34. https://doi.org/10.1016/j.jclinepi.2017.12.004

Vosgerau J., Simonsohn U., Nelson L. D., & Simmons J. P. (2019). 99% impossible: a valid, or falsifiable, internal meta-analysis. *Journal of Experimental Psychology Gen, 148*(9):1628–39

Weissgerber, T. L. (2021). The science of science: Clinical Science launches a new translational meta-research collection. *Clinical Science*, *135*(16), 2031–2034. https://doi.org/10.1042/cs20210777

Weisz, J. R., Venturo-Conerly, K. E., Fitzpatrick, O. M., Frederick, J. A., & Ng, M. Y. (2023). What four decades of meta-analysis have taught us about youth psychotherapy and the science of research synthesis. *Annual Review of Clinical Psychology*, *19*(1), 79–105. https://doi.org/10.1146/annurev-clinpsy-080921-082920

Whiting, P., Savović, J., Higgins, J. P., Caldwell, D. M., Reeves, B. C., Shea, B., Davies, P., Kleijnen, J., & Churchill, R. (2015). ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology*, *69*, 225–234. https://doi.org/10.1016/j.jclinepi.2015.06.005

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis* (2nd ed. 2016). Springer International Publishing: Imprint: Springer. https://doi.org/10.1007/978-3-319-24277-4

Wickham, H. (2025). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.6.0, https://stringr.tidyverse.org.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., & François, R. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, *4*(43). https://doi.org/10.21105/joss.01686

Wickham, H., & Bryan, J. (2023). *readxl: Read Excel files (Version 1.4.3)* [R package]. Comprehensive R Archive Network. https://cran.r-project.org/package=readxl

Wickham, H., François, R., Henry, L., Müller, K., & Vaughan, D. (2025a). *dplyr: A Grammar of Data Manipulation*. R package version 1.1.4, https://dplyr.tidyverse.org.

Wickham, H., & Henry, L. (2025). *purrr: Functional Programming Tools*. R package version 1.2.0, https://purrr.tidyverse.org/.

Wickham, H., Hester, J., & Bryan, J. (2025b). *readr: Read Rectangular Text Data*. R package version 2.1.6, https://readr.tidyverse.org.

Wickham, H., Pedersen, T., & Seidel, D. (2025c). *scales: Scale Functions for Visualization*. R package version 1.4.0, https://scales.r-lib.org.

Wickham, H., Vaughan, D., & Girlich, M. (2025d). *tidyr: Tidy Messy Data*. R package version 1.3.1, https://tidyr.tidyverse.org.

World Health Organization (2003). Caring for children and adolescents with mental disorders: setting WHO directions. World Health Organization. World Health Organization. https://iris.who.int/handle/10665/42679

World Health Organization. (2024a). *Mental health of adolescents*. https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health

World Health Organization. (2024b). *The adolescent health indicators recommended by the Global Action for Measurement of Adolescent health: Guidance for monitoring adolescent health at country, regional and global levels.*

https://iris.who.int/bitstream/handle/10665/376852/9789240092198-eng.pdf?sequence=1

Zbukvic, I., McKay, S., Cooke, S., Anderson, R., Pilkington, V., McGillivray, L., Bailey, A., Purcell, R., & Tye, M. (2024). Evidence for targeted and universal secondary school-based programs for anxiety and depression: An overview of systematic reviews. *Adolescent Research Review*. https://doi.org/10.1007/s40894-023-00211-1

Appendix

Appendix A: Abstracts

German Abstract

Meta-analytische Systematic Reviews sind zentral für evidenzbasierte Entscheidungen in der klinischen Kinder- und Jugendpsychologie (KKJP). Ihre Vertrauenswürdigkeit hängt jedoch maßgeblich von methodischer Qualität, transparenter Berichterstattung und statistischer Konsistenz ab. Vor dem Hintergrund der Replikationskrise und wachsender Zweifel an der Verlässlichkeit von Forschungssynthesen bietet dieses Meta-Review eine systematische Bewertung der methodischen Qualität aktueller einflussreicher meta-analytischer Systematic Reviews im Bereich der KKJP. Zudem wurden zentrale Aspekte der transparenten Berichterstattung sowie der Zusammenhang von methodischer Qualität mit statistischer Konsistenz untersucht.

Eine systematische Literaturrecherche identifizierte 66 meta-analytische Systematic Reviews, die zwischen 2022 und 2024 in sieben führenden KKJP-Zeitschriften veröffentlicht wurden. Die methodische Qualität wurde mithilfe einer adaptierten Version von AMSTAR 2 bewertet und durch eine narrative Analyse zentraler Transparenzaspekte ergänzt, basierend auf PRISMA 2020 und dessen Erweiterungen. Die statistische Konsistenz berichteter Nullhypothesensignifikanztests wurde mit *statcheck* geprüft.

Die methodische Qualität war überwiegend niedrig: Rund 70% der Reviews wurden als *Kritisch Niedrig* eingestuft, lediglich 3 % erreichten die Qualitätsstufe *Hoch*. Die Einhaltung der kritischen AMSTAR 2 Domänen variierte stark. Während statistische Synthesemethoden häufig adäquat berichtet wurden, zeigten sich deutliche Defizite bei der Präregistrierung von Protokollen, der Vollständigkeit der Literaturrecherche und der Begründung ausgeschlossener Studien. Die narrative Analyse der Transparenzkriterien offenbarte zudem erhebliche Heterogenität innerhalb der kritischen Domänen von AMSTAR 2 und identifizierte Berichterstattungslücken, insbesondere bei Protokollnutzung, Suchstrategien und Verfahren zur Bewertung des Risk of Bias. In den Reviews mit extrahierbaren statistischen Tests wurden keine groben statistischen Inkonsistenzen festgestellt

und die methodische Qualität stand in keinem signifikanten Zusammenhang mit der statistischen Konsistenz. Zusammenfassend legen die Ergebnisse nahe, dass aktuelle einflussreiche meta-analytische Systematic Reviews in der KKJP trotz Veröffentlichung in renommierten Fachzeitschriften etablierte methodische Standards nicht erfüllen. Eine konsequente Stärkung methodischer Qualität sowie transparenter und korrekter statistischer Berichterstattung ist daher unerlässlich, um die Glaubwürdigkeit und den klinischen Nutzen von Evidenzsynthesen in der KKJP zu sichern.

English Abstract

Meta-analytic systematic reviews are central to evidence-based decision-making in Clinical Child and Adolescent Psychology (CCAP), yet their credibility depends on methodological rigor, transparent reporting, and statistical consistency. In the context of the replication crisis and concerns about the trustworthiness of research syntheses, this meta-review evaluates the methodological quality of recent high-impact CCAP meta-analytic systematic reviews, examines how key aspects of reporting transparency complement methodological quality indicators, and explores the association of methodological quality and statistical consistency.

A systematic literature search identified 66 meta-analytic systematic reviews published between 2022 and 2024 in seven leading CCAP journals. Methodological quality was assessed using an adapted version of AMSTAR 2, supplemented by a narrative exploration of relevant reporting transparency items derived from PRISMA 2020 and its extensions. Statistical consistency of null hypothesis significance testing results was examined using *statcheck*.

Overall methodological quality was predominantly low. Nearly 70% of reviews were rated as *Critically Low* in confidence, with only 3% rated as *High*. Adherence to AMSTAR 2 critical domains varied substantially, with high compliance for statistical synthesis methods but pronounced weaknesses in protocol preregistration, comprehensive literature searches, and justification of excluded studies. The narrative integration of transparency checklist items revealed considerable heterogeneity within AMSTAR 2 critical domains, highlighting reporting gaps that provided nuance to methodological assessments, particularly regarding protocol use, search strategies, and risk-of-bias procedures. Among reviews with extractable statistical tests, no gross statistical inconsistencies were detected, and methodological quality was not meaningfully associated with statistical consistency. Together, these findings provide an upper-bound baseline of methodological quality and indicate that many influential meta-analytic systematic reviews in CCAP fall short of established methodological standards, despite being published in high-impact journals. Strengthening methodological rigor, reporting transparency, and statistical reporting accuracy is therefore essential to ensure the credibility and clinical utility of evidence syntheses in CCAP.

Appendix B: PRIOR Checklist

Table 3

PRIOR Checklist

Section Topic	#	Item	Location reported
		Title	
Title	1	Identify the report as an overview of reviews.	Title
		Abstract	
Abstract	2	Provide a comprehensive and accurate summary of the purpose, methods, and results of the overview of reviews.	Sections “The Present Meta-Review: Methodological Quality in CCAP” and “Methods”
		Introduction	
Rationale	3	Describe the rationale for conducting the overview of reviews in the context of existing knowledge.	Section “Introduction”
Objectives	4	Provide an explicit statement of the objective(s) or question(s) addressed by the overview of reviews.	Section “Research Questions”
		Methods	

Eligibility criteria	5a	Specify the inclusion and exclusion criteria for the overview of reviews. If supplemental primary studies were included, this should be stated, with a rationale.	Section “Inclusion and Exclusion Criteria”
	5b	Specify the definition of ‘systematic review’ as used in the inclusion criteria for the overview of reviews.	Section “Inclusion and Exclusion Criteria”
Information sources	6	Specify all databases, registers, websites, organizations, reference lists, and other sources searched or consulted to identify systematic reviews and supplemental primary studies (if included).	Sections “Information Sources and Search Strategy” and “Identification and Selection of Studies”
Search strategy	7	Present the full search strategies for all databases, registers and websites, such that they could be reproduced. Describe any search filters and limits applied.	Table 1
Selection process	8a	Describe the methods used to decide whether a systematic review or supplemental primary study (if included) met the inclusion criteria of the overview of reviews.	Section “Inclusion and Exclusion Criteria” and “Study Selection”
	8b	Describe how overlap in the populations, interventions, comparators, and/or outcomes of systematic reviews was identified and managed during study selection.	N/A: Not relevant to the scope of the present thesis.
Data collection process	9a	Describe the methods used to collect data from reports.	Section “Data Collection Process”

	9b	If applicable, describe the methods used to identify and manage primary study overlap at the level of the comparison and outcome during data collection. For each outcome, specify the method used to illustrate and/or quantify the degree of primary study overlap across systematic reviews.	N/A: Not relevant to the scope of the present thesis.
	9c	If applicable, specify the methods used to manage discrepant data across systematic reviews during data collection.	N/A: Not relevant to the scope of the present thesis.
Data items	10	List and define all variables and outcomes for which data were sought. Describe any assumptions made and/or measures taken to identify and clarify missing or unclear information.	Section “Data Items”
Risk of bias assessment	11a	Describe the methods used to assess risk of bias or methodological quality of the included systematic reviews.	Section “RQ 1 and 2: Methodological Quality”
	11b	Describe the methods used to collect data on (from the systematic reviews) and/or assess the risk of bias of the primary studies included in the systematic reviews. Provide a justification for instances where flawed, incomplete, or missing assessments are identified but not re-assessed.	Section “RQ 1 and 2: Methodological Quality”
	11c	Describe the methods used to assess the risk of bias of supplemental primary studies (if included).	N/A: Not relevant to the scope of the present thesis.
Synthesis methods	12a	Describe the methods used to summarize or synthesize results and provide a rationale for the choice(s).	Section “Data Analysis”

	12b	Describe any methods used to explore possible causes of heterogeneity among results.	N/A: Not relevant to the scope of the present thesis.
	12c	Describe any sensitivity analyses conducted to assess the robustness of the synthesized results.	N/A: Not relevant to the scope of the present thesis.
Reporting bias assessment	13	Describe the methods used to collect data on (from the systematic reviews) and/or assess the risk of bias due to missing results in a summary or synthesis (arising from reporting biases at the levels of the systematic reviews, primary studies, and supplemental primary studies, if included).	N/A: Not relevant to the scope of the present thesis.
Certainty Assessment	14	Describe the methods used to collect data on (from the systematic reviews) and/or assess certainty (or confidence) in the body of evidence for an outcome.	N/A: Not relevant to the scope of the present thesis.
		Results	
Systematic review and supplemental primary study selection	15a	Describe the results of the search and selection process, including the number of records screened, assessed for eligibility, and included in the overview of reviews, ideally with a flow diagram.	Appendix F: PRISMA flow diagram
	15b	Provide a list of studies that might appear to meet the inclusion criteria, but were excluded, with the main reason for exclusion.	Appendix D
Characteristics of systematic reviews and	16	Cite each included systematic review and supplemental primary study (if included) and present its characteristics.	Appendix C, Appendix H

supplemental primary studies			
Primary study overlap	17	Describe the extent of primary study overlap across the included systematic reviews.	N/A: Not relevant to the scope of the present thesis.
Risk of bias in systematic reviews, primary studies, and supplemental primary studies	18a	Present assessments of risk of bias or methodological quality for each included systematic review.	Section “Methodological Quality Assessment”, risk of bias: N/A
	18b	Present assessments (collected from systematic reviews or assessed anew) of the risk of bias of the primary studies included in the systematic reviews.	N/A: Not relevant to the scope of the present thesis.
	18c	Present assessments of the risk of bias of supplemental primary studies (if included).	N/A: Not relevant to the scope of the present thesis.
Summary or synthesis of results	19a	For all outcomes, summarize the evidence from the systematic reviews and supplemental primary studies (if included). If meta-analyses were done, present for each the summary estimate and its precision and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	N/A: Not relevant to the scope of the present thesis.
	19b	If meta-analyses were done, present results of all investigations of possible causes of heterogeneity.	N/A: Not relevant to the scope of the present thesis.

	19c	If meta-analyses were done, present results of all sensitivity analyses conducted to assess the robustness of synthesized results.	N/A: Not relevant to the scope of the present thesis.
Reporting biases	20	Present assessments (collected from systematic reviews and/or assessed anew) of the risk of bias due to missing primary studies, analyses, or results in a summary or synthesis (arising from reporting biases at the levels of the systematic reviews, primary studies, and supplemental primary studies, if included) for each summary or synthesis assessed.	N/A: Not relevant to the scope of the present thesis.
Certainty of evidence	21	Present assessments (collected or assessed anew) of certainty (or confidence) in the body of evidence for each outcome.	N/A: Not relevant to the scope of the present thesis.
		Discussion	
Discussion	22a	Summarize the main findings, including any discrepancies in findings across the included systematic reviews and supplemental primary studies (if included).	Section “Discussion”
	22b	Provide a general interpretation of the results in the context of other evidence.	Section “Discussion”
	22c	Discuss any limitations of the evidence from systematic reviews, their primary studies, and supplemental primary studies (if included) included in the overview of reviews. Discuss any limitations of the overview of reviews methods used.	Section “Limitations”

	22d	Discuss implications for practice, policy, and future research (both systematic reviews and primary research). Consider the relevance of the findings to the end users of the overview of reviews, e.g., healthcare providers, policymakers, patients, among others.	Section “Conclusion”
		Other Information	
Registration and protocol	23a	Provide registration information for the overview of reviews, including register name and registration number, or state that the overview of reviews was not registered.	Section “Statements and Declarations”
	23b	Indicate where the overview of reviews protocol can be accessed, or state that a protocol was not prepared.	Section “Statements and Declarations”
	23c	Describe and explain any amendments to information provided at registration or in the protocol. Indicate the stage of the overview of reviews at which amendments were made.	Section “Deviations from Protocol”
Support	24	Describe sources of financial or non-financial support for the overview of reviews, and the role of the funders or sponsors in the overview of reviews.	Section “Statements and Declarations”
Competing interests	25	Declare any competing interests of the overview of reviews' authors.	Section “Statements and Declarations”
Author information	26a	Provide contact information for the corresponding author.	Section “Statements and Declarations”

	26b	Describe the contributions of individual authors and identify the guarantor of the overview of reviews.	N/A: Not relevant to the scope of the present thesis.
Availability of data and other materials	27	Report which of the following are available, where they can be found, and under which conditions they may be accessed: template data collection forms; data collected from included systematic reviews and supplemental primary studies; analytic code; any other materials used in the overview of reviews.	Section “Statements and Declarations”

Appendix C: Included Meta-Analytic Systematic Reviews

Alozkan-Sever, C., Uppendahl, J. R., Cuijpers, P., De Vries, R., Rahman, A., Mittendorfer- Rutz, E., Akhtar, A., Zheng, Z., & Sijbrandij, M. (2023). Research review: Psychological and psychosocial interventions for children and adolescents with depression, anxiety, and post-traumatic stress disorder in low- and middle-income countries – a systematic review and meta-analysis. *Journal of Child Psychology and Psychiatry*, *64*(12), 1776–1788. https://doi.org/10.1111/jcpp.13891

Armoon, B., Ghadipasha, A., Mohammadi, R., Lesage, A., Harooni, J., & Griffiths, M. D. (2025). The global prevalence of mental health disorders among runaway and homeless youth: A meta-analysis. *European Child & Adolescent Psychiatry*, *34*(2), 535–564. https://doi.org/10.1007/s00787-024-02519-2

Bapuji, S. B., Hansen, A., Marembo, M. H., Olivier, P., & Yap, M. B. H. (2024). Modifiable parental factors associated with the mental health of youth from immigrant families in high-income countries: A systematic review and meta-analysis. *Clinical Psychology Review*, *110*, 102429. https://doi.org/10.1016/j.cpr.2024.102429

Bellato, A., Sesso, G., Milone, A., Masi, G., & Cortese, S. (2024). Systematic review and meta-analysis: Altered autonomic functioning in youths with emotional dysregulation. *Journal of the American Academy of Child & Adolescent Psychiatry*, *63*(2), 216–230. https://doi.org/10.1016/j.jaac.2023.01.017

Betancourt, J. L., Alderson, R. M., Roberts, D. K., & Bullard, C. C. (2024). Self-esteem in children and adolescents with and without attention-deficit/hyperactivity disorder: A meta-analytic review. *Clinical Psychology Review*, *108*, 102394. https://doi.org/10.1016/j.cpr.2024.102394

Bloss, C., Brown, S., & Sawrikar, V. (2024). Does behavioural parent training reduce internalising symptoms (or not) among children with externalising problems? Systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, *33*(8), 2485–2501. https://doi.org/10.1007/s00787-022-02122-3

Bo, A., Goings, T. C., Evans, C. B. R., Sharma, A., Jennings, Z., Durand, B., Bardeen, A., & Murray-Lichtman, A. (2023). Culturally sensitive prevention programs for substance use among adolescents of color: A systematic review and meta-analysis of randomized controlled trials. *Clinical Psychology Review*, *99*, 102233. https://doi.org/10.1016/j.cpr.2022.102233

Boldrini, T., Ghiandoni, V., Mancinelli, E., Salcuni, S., & Solmi, M. (2023). Systematic review and meta-analysis: Psychosocial treatments for disruptive behavior symptoms and disorders in adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry*, *62*(2), 169–189. https://doi.org/10.1016/j.jaac.2022.05.002

Boumpa, V., Papatoukaki, A., Kourti, A., Mintzia, S., Panagouli, E., Bacopoulou, F., Psaltopoulou, T., Spiliopoulou, C., Tsolia, M., Sergentanis, T. N., & Tsitsika, A. (2024). Sexual abuse and post-traumatic stress disorder in childhood, adolescence and young adulthood: A systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, *33*(6), 1653–1673. https://doi.org/10.1007/s00787-022-02015- 5

Breaux, R., Baweja, R., Eadeh, H.-M., Shroff, D. M., Cash, A. R., Swanson, C. S., Knehans, A., & Waxmonsky, J. G. (2023). Systematic review and meta-analysis: Pharmacological and nonpharmacological interventions for persistent nonepisodic irritability. *Journal of the American Academy of Child & Adolescent Psychiatry*, *62*(3), 318–334. https://doi.org/10.1016/j.jaac.2022.05.012

Cooke, J. E., Deneault, A., Devereux, C., Eirich, R., Fearon, R. M. P., & Madigan, S. (2022). Parental sensitivity and child behavioral problems: A meta-analytic review. *Child Development*, *93*(5), 1231–1248. https://doi.org/10.1111/cdev.13764

Cushing, T., Robertson, S., Mannes, J., Marshall, N., Carey, M. J., Duschinsky, R., & Meiser- Stedman, R. (2024). The relationship between attachment and posttraumatic stress in children and adolescents: A meta-analytic review. *Development and Psychopathology*, *36*(3), 1055–1069. https://doi.org/10.1017/S0954579423000299

Davis, R. S., Meiser-Stedman, R., Afzal, N., Devaney, J., Halligan, S. L., Lofthouse, K., Smith, P., Stallard, P., Ye, S., & Hiller, R. M. (2023). Systematic review and meta-analysis: Group-based interventions for

treating posttraumatic stress symptoms in children and adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry*, *62*(11), 1217–1232. https://doi.org/10.1016/j.jaac.2023.02.013

Dean, R. L., Lester, K. J., Grant, E., Field, A. P., Orchard, F., & Pile, V. (2025). The impact of interventions for depression on self-perceptions in young people: A systematic review & meta-analysis. *Clinical Psychology Review*, *115*, 102521. https://doi.org/10.1016/j.cpr.2024.102521

Defoe, I. N., Treffers, S., & Stams, G. J. (2023). Research review: Siblings matter. A multi level meta-analysis on the association between cannabis use among adolescent siblings. *Journal of Child Psychology and Psychiatry*, *64*(11), 1532–1544. https://doi.org/10.1111/jcpp.13836

Dülsen, P., & Baumeister, H. (2025). Internet- and mobile-based anxiety and depression interventions for children and adolescents: Efficacy and negative effects - a systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, *34*(1), 101–121. https://doi.org/10.1007/s00787-024-02404-y

England-Mason, G., Andrews, K., Atkinson, L., & Gonzalez, A. (2023). Emotion socialization parenting interventions targeting emotional competence in young children: A systematic review and meta-analysis of randomized controlled trials. *Clinical Psychology Review*, *100*, 102252. https://doi.org/10.1016/j.cpr.2023.102252

Ezquerra, B., Alacreu-Crespo, A., Peñuelas-Calvo, I., Abascal-Peiró, S., Jiménez-Muñoz, L., Nicholls, D., Baca-García, E., & Porras-Segovia, A. (2024). Characteristics of single vs. multiple suicide attempters among adolescents: A systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, *33*(10), 3405–3418. https://doi.org/10.1007/s00787-023-02260-2

Farkas, B. F., Takacs, Z. K., Kollárovics, N., & Balázs, J. (2024). The prevalence of self-injury in adolescence: A systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, *33*(10), 3439–3458. https://doi.org/10.1007/s00787-023- 02264-y

Finlay-Jones, A. L., Ang, J. E., Brook, J., Lucas, J. D., MacNeill, L. A., Mancini, V. O., Kottampally, K., Elliott, C., Smith, J. D., & Wakschlag, L. S. (2024). Systematic review and meta-analysis: Early irritability as a transdiagnostic

neurodevelopmental vulnerability to later mental health problems. *Journal of the American Academy of Child & Adolescent Psychiatry*, *63*(2), 184–215. https://doi.org/10.1016/j.jaac.2023.01.018

Francis, E. R., Tsaligopoulou, A., Stock, S. E., Pingault, J., & Baldwin, J. R. (2023). Subjective and objective experiences of childhood adversity: A meta-analysis of their agreement and relationships with psychopathology. *Journal of Child Psychology and Psychiatry*, *64*(8), 1185–1199. https://doi.org/10.1111/jcpp.13803

Frisira, E., Holland, J., & Sayal, K. (2025). Systematic review and meta-analysis: Relative age in attention-deficit/hyperactivity disorder and autism spectrum disorder. *European Child & Adolescent Psychiatry*, *34*(2), 381–401. https://doi.org/10.1007/s00787-024- 02459-x

Gu, Q., Liu, J., Zhang, X., Huang, A., Yu, X., Wu, K., & Huang, Y. (2025). Association between heavy metals exposure and risk of attention deficit hyperactivity disorder (ADHD) in children: A systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, *34*(3), 921–941. https://doi.org/10.1007/s00787-024-02546-z

Guzman-Holst, C., Zaneva, M., Chessell, C., Creswell, C., & Bowes, L. (2022). Research review: Do antibullying interventions reduce internalizing symptoms? A systematic review, meta-analysis, and meta-regression exploring intervention components, moderators, and mechanisms. *Journal of Child Psychology and Psychiatry*, *63*(12), 1454–1465. https://doi.org/10.1111/jcpp.13620

Haza, B., Gosling, C. J., Ciminaghi, F., Conty, L., & Pinabiaux, C. (2024). Research review: Social cognition and everyday social skills in children and adolescents with attention deficit/hyperactivity disorder: A meta-analysis of case–control studies. *Journal of Child Psychology and Psychiatry*, *65*(9), 1245–1254. https://doi.org/10.1111/jcpp.14006

Holzman, J. B. W., Kennedy, S. M., Grassie, H. L., & Ehrenreich-May, J. (2022). Associations between dispositional parental emotion regulation and youth mental health symptoms: A systematic review and meta-analysis. *Clinical Psychology Review*, *95*, 102174. https://doi.org/10.1016/j.cpr.2022.102174

Jensen, A. C., & Thomsen, A. E. (2024). Parental differential treatment of siblings linked with internalizing and externalizing behavior: A meta-analysis. *Child Development*, *95*(4), 1384–1405. https://doi.org/10.1111/cdev.14091

Joseph, H. M., Lorenzo, N. E., Fisher, N., Novick, D. R., Gibson, C., Rothenberger, S. D., Foust, J. E., & Chronis-Tuscano, A. (2023). Research review: A systematic review and meta-analysis of infant and toddler temperament as predictors of childhood attention deficit/hyperactivity disorder. *Journal of Child Psychology and Psychiatry*, *64*(5), 715–735. https://doi.org/10.1111/jcpp.13753

Khazdouz, M., Safarzadeh, R., Hejrani, B., Hasani, M., Mahdavi, F. S., Ejtahed, H.-S., & Qorbani, M. (2025). The association between junk foods consumption and attention deficit hyperactivity disorder in children and adolescents: A systematic review and meta-analysis of observational studies. *European Child & Adolescent Psychiatry*, *34*(3), 825–834. https://doi.org/10.1007/s00787-024-02521-8

Kim, H., Kim, J. H., Yi, J. H., Kim, J. Y., Solmi, M., Cortese, S., Smith, L., Koyanagi, A., Shin, J. I., Cheon, K.-A., & Fusar-Poli, P. (2024). Correlations between sleep problems, core symptoms, and behavioral problems in children and adolescents with autism spectrum disorder: A systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, *33*(5), 1539–1549. https://doi.org/10.1007/s00787-023-02253- 1

Kopp, M., Lindauer, M., & Garthus-Niegel, S. (2024). Association between maternal employment and the child's mental health: A systematic review with meta-analysis. *European Child & Adolescent Psychiatry*, *33*(9), 3025–3042. https://doi.org/10.1007/s00787-023-02164-1

Lång, U., Yates, K., Leacy, F. P., Clarke, M. C., McNicholas, F., Cannon, M., & Kelleher, I. (2022). Systematic review and meta-analysis: Psychosis risk in children and adolescents with an at-risk mental state. *Journal of the American Academy of Child & Adolescent Psychiatry*, *61*(5), 615–625. https://doi.org/10.1016/j.jaac.2021.07.593

Lee, A. H., Kitagawa, Y., Mirhashem, R., Rodriguez, M., Hilerio, R., & Bernard, K. (2025). Do dimensions of childhood adversity differ in their direct associations with youth psychopathology? A meta-analysis. *Development*

and Psychopathology, *37*(2), 871– 901. https://doi.org/10.1017/S0954579424000737

Li, C., Wang, P., Martin-Moratinos, M., Bella-Fernández, M., & Blasco-Fontecilla, H. (2024). Traditional bullying and cyberbullying in the digital age and its associated mental health problems in children and adolescents: A meta-analysis. *European Child & Adolescent Psychiatry*, *33*(9), 2895–2909. https://doi.org/10.1007/s00787-022-02128- x

Lin, S. C., Kehoe, C., Pozzi, E., Liontos, D., & Whittle, S. (2024). Research review: Child emotion regulation mediates the association between family factors and internalizing symptoms in children and adolescents – a meta-analysis. *Journal of Child Psychology and Psychiatry*, *65*(3), 260–274. https://doi.org/10.1111/jcpp.13894

Loheide-Niesmann, L., Riem, M. M. E., & Cima, M. (2024). The impact of maternal childhood maltreatment on child externalizing behaviour and the mediating factors underlying this association: A three-level meta-analysis and systematic review. *European Child & Adolescent Psychiatry*, *33*(8), 2445–2470. https://doi.org/10.1007/s00787-022-02117-0

Maciel, L., Gomis-Pomares, A., Day, C., & Basto-Pereira, M. (2023). Cross-cultural adaptability of parenting interventions designed for childhood behavior problems: A meta-analysis. *Clinical Psychology Review*, *102*, 102274. https://doi.org/10.1016/j.cpr.2023.102274

Manuele, S. J., Yap, M. B. H., Lin, S. C., Pozzi, E., & Whittle, S. (2023). Associations between paternal versus maternal parenting behaviors and child and adolescent internalizing problems: A systematic review and meta-analysis. *Clinical Psychology Review*, *105*, 102339. https://doi.org/10.1016/j.cpr.2023.102339

McKern, N., Dargue, N., & Sweller, N. (2023). Comparing gesture frequency between autistic and neurotypical individuals: A systematic review and meta-analysis. *Psychological Bulletin*, *149*(11–12), 724–745. https://doi.org/10.1037/bul0000408

Moltrecht, B., Lange, A. M. C., Merrick, H., & Radley, J. (2024). Whole-family programmes for families living with parental mental illness: A systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, *33*(9), 3203–3246. https://doi.org/10.1007/s00787-024-02380-3

Newlove-Delgado, T., Russell, A. E., Mathews, F., Cross, L., Bryant, E., Gudka, R., Ukoumunne, O. C., & Ford, T. J. (2023). Annual research review: The impact of COVID-19 on psychopathology in children and young people worldwide: systematic review of studies with pre- and within-pandemic data. *Journal of Child Psychology and Psychiatry*, *64*(4), 611–640. https://doi.org/10.1111/jcpp.13716

Nieto-Retuerto, M., Torres-Gomez, B., & Alonso-Arbiol, I. (2024). Parental mentalization and children's externalizing problems: A systematic review and meta-analysis. *Development and Psychopathology, Advance online publication.* https://doi.org/10.1017/S0954579424001391

O'Halloran, L., Coey, P., & Wilson, C. (2022). Suicidality in autistic youth: A systematic review and meta-analysis. *Clinical Psychology Review*, *93*, 102144. https://doi.org/10.1016/j.cpr.2022.102144

Perlstein, S., Fair, M., Hong, E., & Waller, R. (2023). Treatment of childhood disruptive behavior disorders and callous-unemotional traits: A systematic review and two multilevel meta-analyses. *Journal of Child Psychology and Psychiatry*, *64*(9), 1372– 1387. https://doi.org/10.1111/jcpp.13774

Restoy, D., Oriol-Escudé, M., Alonzo-Castillo, T., Magán-Maganto, M., Canal-Bedia, R., Díez-Villoria, E., Gisbert-Gustemps, L., Setién-Ramos, I., Martínez-Ramírez, M., Ramos-Quiroga, J. A., & Lugo-Marín, J. (2024). Emotion regulation and emotion dysregulation in children and adolescents with autism spectrum disorder: A meta-analysis of evaluation and intervention studies. *Clinical Psychology Review*, *109*, 102410. https://doi.org/10.1016/j.cpr.2024.102410

Robberegt, S. J., Brouwer, M. E., Kooiman, B. E. A. M., Stikkelbroek, Y. A. J., Nauta, M. H., & Bockting, C. L. H. (2023). Meta-analysis: Relapse prevention strategies for depression and anxiety in remitted adolescents and young adults. *Journal of the American Academy of Child & Adolescent Psychiatry*, *62*(3), 306–317. https://doi.org/10.1016/j.jaac.2022.04.014

Sacco, R., Camilleri, N., Eberhardt, J., Umla-Runge, K., & Newbury-Birch, D. (2024). A systematic review and meta-analysis on the prevalence of mental disorders among children and adolescents in Europe. *European Child & Adolescent Psychiatry*, *33*(9), 2877–2894. https://doi.org/10.1007/s00787-022-02131-2

Shen, J., Zhao, S., Horn, T., Benkart, R., Busch, T., Vrabec, A., & Taylor, H. G. (2023). Family matters: A systematic review and meta-analysis on the efficacy of family-oriented interventions for children with acquired brain injuries. *Clinical Psychology Review*, *99*, 102218. https://doi.org/10.1016/j.cpr.2022.102218

Shephard, E., Zuccolo, P. F., Idrees, I., Godoy, P. B. G., Salomone, E., Ferrante, C., Sorgato, P., Catão, L. F. C. C., Goodwin, A., Bolton, P. F., Tye, C., Groom, M. J., & Polanczyk, G. V. (2022). Systematic review and meta-analysis: The science of early-life precursors and interventions for attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *61*(2), 187–226. https://doi.org/10.1016/j.jaac.2021.03.016

Siciliano, R. E., Anderson, A. S., & Compas, B. E. (2022). Autonomic nervous system correlates of posttraumatic stress symptoms in youth: Meta-analysis and qualitative review. *Clinical Psychology Review*, *92*, 102125. https://doi.org/10.1016/j.cpr.2022.102125

Somers, K., Spruit, A., Stams, G. J., Vandevelde, S., Lindauer, R., & Assink, M. (2024). Identifying effective moderators of cognitive behavioural trauma treatment with caregiver involvement for youth with PTSD: A meta-analysis. *European Child & Adolescent Psychiatry*, *33*(7), 2067–2081. https://doi.org/10.1007/s00787-022-02088- 2

Tang, P., Kostyrka-Allchorne, K., Butura, A.-M., Phillips-Owen, J., & Sonuga-Barke, E. (2024). Reciprocal developmental pathways between future-related thinking and symptoms of adolescent depression and anxiety: A systematic review and meta-analysis of longitudinal studies. *Clinical Psychology Review*, *112*, 102465. https://doi.org/10.1016/j.cpr.2024.102465

Tindall, L., Kerrigan, P., Li, J., Hayward, E., & Gega, L. (2024). Is behavioural activation an effective treatment for depression in children and adolescents? An updated systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, *33*(12), 4133– 4156. https://doi.org/10.1007/s00787-024-02429-3

Trompeter, N., Dârvariu, Ș., Brieva-Toloza, A. V., Opitz, M.-C., Rabelo-da-Ponte, F. D., Sharpe, H., Desrivieres, S., Schmidt, U., & Micali, N. (2024). The

prospective relationship between anxiety symptoms and eating disorder symptoms among adolescents: A systematic review and meta-analysis of a bi-directional relationship. *European Child & Adolescent Psychiatry, Advance online publication.* https://doi.org/10.1007/s00787-024-02601-9

Tung, I., Hipwell, A. E., Grosse, P., Battaglia, L., Cannova, E., English, G., Quick, A. D., Llamas, B., Taylor, M., & Foust, J. E. (2024). Prenatal stress and externalizing behaviors in childhood and adolescence: A systematic review and meta-analysis. *Psychological Bulletin*, *150*(2), 107–131. https://doi.org/10.1037/bul0000407

Van IJzendoorn, M. H., Schuengel, C., Wang, Q., & Bakermans-Kranenburg, M. J. (2023). Improving parenting, child attachment, and externalizing behaviors: Meta-analysis of the first 25 randomized controlled trials on the effects of Video-feedback Intervention to promote positive parenting and sensitive discipline. *Development and Psychopathology*, *35*(1), 241–256. https://doi.org/10.1017/S0954579421001462

Van Meter, A. R., Knowles, E. A., & Mintz, E. H. (2023). Systematic review and meta-analysis: International prevalence of suicidal ideation and attempt in youth. *Journal of the American Academy of Child & Adolescent Psychiatry*, *62*(9), 973–986. https://doi.org/10.1016/j.jaac.2022.07.867

Venturo-Conerly, K. E., Eisenman, D., Wasil, A. R., Singla, D. R., & Weisz, J. R. (2023). Meta-analysis: The effectiveness of youth psychotherapy interventions in low- and middle-income countries. *Journal of the American Academy of Child & Adolescent Psychiatry*, *62*(8), 859–873. https://doi.org/10.1016/j.jaac.2022.12.005

Walsh, E. H., McMahon, J., & Herring, M. P. (2022). Research review: The effect of school based suicide prevention on suicidal ideation and suicide attempts and the role of intervention and contextual factors among adolescents: a meta-analysis and meta regression. *Journal of Child Psychology and Psychiatry*, *63*(8), 836–845. https://doi.org/10.1111/jcpp.13598

Wang, L. A. L., Petrulla, V., Zampella, C. J., Waller, R., & Schultz, R. T. (2022). Gross motor impairment and its relation to social skills in autism spectrum disorder: A systematic review and two meta-analyses. *Psychological Bulletin*, *148*(3–4), 273–300. https://doi.org/10.1037/bul0000358

Wang, Y., Chen, X., Wang, A., Jordan, L. P., & Lu, S. (2024). Research review: Grandparental care and child mental health – a systematic review and meta-analysis. *Journal of Child Psychology and Psychiatry*, *65*(4), 568–586. https://doi.org/10.1111/jcpp.13943

Wendel, F., Bender, S., Breitinger, E., Coenen, M., Hummel, J., Immich, G., Kirschneck, M., Klünder, V., Kunzler, A. M., Lieb, K., Movsisyan, A., Li, L. Y., Ravens-Sieberer, U., Rehfuess, E., Voss, S., & Jung-Sievers, C. (2024). Interventions to build resilience and to ameliorate negative psychosocial effects of the COVID-19 pandemic on children and adolescents: A systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, *33*(11), 3707–3726. https://doi.org/10.1007/s00787-023- 02280-y

Wergeland, G. J., Ghaderi, A., Fjermestad, K., Enebrink, P., Halsaa, L., Njardvik, U., Riise, E. N., Vorren, G., & Öst, L.-G. (2025). Family therapy and cognitive behavior therapy for eating disorders in children and adolescents in routine clinical care: A systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, *34*(3), 883–902. https://doi.org/10.1007/s00787-024-02544-1

Woolgar, F., Garfield, H., Dalgleish, T., & Meiser-Stedman, R. (2022). Systematic review and meta-analysis: Prevalence of posttraumatic stress disorder in trauma-exposed preschool-aged children. *Journal of the American Academy of Child & Adolescent Psychiatry*, *61*(3), 366–377. https://doi.org/10.1016/j.jaac.2021.05.026

Yang, Y., Shields, G. S., Zhang, Y., Wu, H., Chen, H., & Romer, A. L. (2022). Child executive function and future externalizing and internalizing problems: A meta-analysis of prospective longitudinal studies. *Clinical Psychology Review*, *97*, 102194. https://doi.org/10.1016/j.cpr.2022.102194

Zhang, Q., & Ji, W. (2024). Overparenting and offspring depression, anxiety, and internalizing symptoms: A meta-analysis. *Development and Psychopathology*, *36*(3), 1307–1322. https://doi.org/10.1017/S095457942300055X

Appendix D: Reports Excluded at Full-Text Screening and Reasons for Exclusion

Seventy-nine reports were excluded during full-text screening; for the flow diagram, see Appendix F.

Wrong Age Group

k = 23

Adams, D., Dargue, N., & Paynter, J. (2023). Longitudinal studies of challenging behaviours in autistic children and adults: A systematic review and meta-analysis. *Clinical Psychology Review*, *104*, 102320. https://doi.org/10.1016/j.cpr.2023.102320

Brumariu, L. E., Waslin, S. M., Gastelle, M., Kochendorfer, L. B., & Kerns, K. A. (2023). Anxiety, academic achievement, and academic self-concept: Meta-analytic syntheses of their relations across developmental periods. *Development and Psychopathology*, *35*(4), 1597–1613. https://doi.org/10.1017/S0954579422000323

Chen, Y., Xi, Z., Saunders, R., Simmons, D., Totsika, V., & Mandy, W. (2024). A systematic review and meta-analysis of the relationship between sensory processing differences and internalising/externalising problems in autism. *Clinical Psychology Review*, *114*, 102516. https://doi.org/10.1016/j.cpr.2024.102516

Clauss, K., Gorday, J. Y., & Bardeen, J. R. (2022). Eye tracking evidence of threat-related attentional bias in anxiety- and fear-related disorders: A systematic review and meta-analysis. *Clinical Psychology Review*, *93*, 102142. https://doi.org/10.1016/j.cpr.2022.102142

Curran, T., & Hill, A. P. (2022). Young people's perceptions of their parents' expectations and criticism are increasing over time: Implications for perfectionism. *Psychological Bulletin*, *148*(1–2), 107–128. https://doi.org/10.1037/bul0000347 Deneault, A., Cabrera, N. J., & Bureau, J. (2022). A meta-analysis on observed paternal and maternal sensitivity. *Child Development*, *93*(6), 1631–1648. https://doi.org/10.1111/cdev.13832

Ditzer, J., Wong, E. Y., Modi, R. N., Behnke, M., Gross, J. J., & Talmon, A. (2023). Child maltreatment and alexithymia: A meta-analytic review. *Psychological Bulletin*, *149*(5–6), 311–329. https://doi.org/10.1037/bul0000391

Edwards, H., Wright, S., Sargeant, C., Cortese, S., & Wood-Downie, H. (2024). Research review: A systematic review and meta-analysis of sex differences in narrow constructs of restricted and repetitive behaviours and interests in autistic children, adolescents, and adults. *Journal of Child Psychology and Psychiatry*, *65*(1), 4–17. https://doi.org/10.1111/jcpp.13855

Emmer, C., Dorn, J., & Mata, J. (2024). The immediate effect of discrimination on mental health: A meta-analytic review of the causal evidence. *Psychological Bulletin*, *150*(3), 215–252. https://doi.org/10.1037/bul0000419

Geer, E. A., Barroso, C., Conlon, R. A., Dasher, J. M., & Ganley, C. M. (2024). A meta-analytic review of the relation between spatial anxiety and spatial skills. *Psychological Bulletin*, *150*(4), 464–486. https://doi.org/10.1037/bul0000420

Khan, A. N., & Jaffee, S. R. (2022). Alexithymia in individuals maltreated as children and adolescents: A meta-analysis. *Journal of Child Psychology and Psychiatry*, *63*(9), 963– 972. https://doi.org/10.1111/jcpp.13616

Li, W., Zhao, Z., Chen, D., Peng, Y., & Lu, Z. (2022). Prevalence and associated factors of depression and anxiety symptoms among college students: A systematic review and meta-analysis. *Journal of Child Psychology and Psychiatry*, *63*(11), 1222–1230. https://doi.org/10.1111/jcpp.13606

Lins-Silva, D. H., Bandeira, I. D., Faria-Guimarães, D., Dorea-Bandeira, I., Tolentino, A., Miranda-Scippa, Â., Hermens, D. F., & Lucena, R. (2025). Prevalence of mental disorders among family members of individuals on the autism spectrum: Systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, *34*(4), 1267–1279. https://doi.org/10.1007/s00787-024-02575-8

Marx, I., Cortese, S., Koelch, M. G., & Hacker, T. (2022). Meta-analysis: Altered perceptual timing abilities in attention-deficit/hyperactivity disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, *61*(7), 866–880. https://doi.org/10.1016/j.jaac.2021.12.004

Nicholls-Clow, R., Simmonds-Buckley, M., & Waller, G. (2024). Avoidant/restrictive food intake disorder: Systematic review and meta-analysis demonstrating the impact of study quality on prevalence rates. *Clinical Psychology Review*, *114*, 102502. https://doi.org/10.1016/j.cpr.2024.102502

O'Shea, J., Jenkins, R., Nicholls, D., Downs, J., & Hudson, L. D. (2025). Prevalence, severity and risk factors for mental disorders among sexual and gender minority young people: A systematic review of systematic reviews and meta-analyses. *European Child & Adolescent Psychiatry*, *34*(3), 959–982. https://doi.org/10.1007/s00787-024-02552-1

Pawar, A. K. S., Firmin, E. S., Wilens, T. E., & Hammond, C. J. (2024). Systematic review and meta-analysis: Medical and recreational cannabis legalization and cannabis use among youth in the United States. *Journal of the American Academy of Child & Adolescent Psychiatry*, *63*(11), 1084–1113. https://doi.org/10.1016/j.jaac.2024.02.016

Potrebny, T., Nilsen, S. A., Bakken, A., Von Soest, T., Kvaløy, K., Samdal, O., Sivertsen, B., Aase, H., & Bang, L. (2025). Secular trends in mental health problems among young people in Norway: A review and meta-analysis. *European Child & Adolescent Psychiatry*, *34*(1), 69–81. https://doi.org/10.1007/s00787-024-02371-4

Pozuelo, J. R., Desborough, L., Stein, A., & Cipriani, A. (2022). Systematic review and meta-analysis: Depressive symptoms and risky behaviors among adolescents in low- and middle-income countries. *Journal of the American Academy of Child & Adolescent Psychiatry*, *61*(2), 255–276. https://doi.org/10.1016/j.jaac.2021.05.005

Tu, E.-N., Manley, H., Saunders, K. E. A., & Creswell, C. (2024). Systematic review and meta-analysis: Risks of anxiety disorders in offspring of parents

with mood disorders. *Journal of the American Academy of Child & Adolescent Psychiatry*, *63*(4), 407–421. https://doi.org/10.1016/j.jaac.2023.06.022

Uljarević, M., Spackman, E. K., Whitehouse, A. J. O., Frazier, T. W., Billingham, W., Condron, P., Hardan, A., & Leekam, S. R. (2023). Arriving at the empirically based conceptualization of restricted and repetitive behaviors: A systematic review and meta-analytic examination of factor analyses. *Clinical Psychology Review*, *103*, 102286. https://doi.org/10.1016/j.cpr.2023.102286

Wilkop, M., Wade, T. D., Keegan, E., & Cohen-Woods, S. (2023). Impairments among DSM- 5 eating disorders: A systematic review and multilevel meta-analysis. *Clinical Psychology Review*, *101*, 102267. https://doi.org/10.1016/j.cpr.2023.102267

Zhu, J., Exner-Cortens, D., Dobson, K., Wells, L., Noel, M., & Madigan, S. (2024). Adverse childhood experiences and intimate partner violence: A meta-analysis. *Development and Psychopathology*, *36*(2), 929–943. https://doi.org/10.1017/S0954579423000196

Wrong Outcome

k = 14

Bernabe-Zuñiga, J. E., Rodriguez-Lucenilla, M. I., Alias-Castillo, A. J., Rueda-Ruzafa, L., Roman, P., & Del Mar Sanchez-Joya, M. (2025). Early interventions with parental participation and their implications on the neurodevelopment of premature children: A systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, *34*(3), 853–865. https://doi.org/10.1007/s00787-024-02528-1

Bryant, A., Schlesinger, H., Sideri, A., Holmes, J., Buitelaar, J., & Meiser-Stedman, R. (2023). A meta-analytic review of the impact of ADHD medications on anxiety and depression in children and adolescents. *European Child & Adolescent Psychiatry*, *32*(10), 1885–1898. https://doi.org/10.1007/s00787-022-02004-8

Delagneau, G., Twilhaar, E. S., Testa, R., Van Veen, S., & Anderson, P. (2023). Association between prenatal maternal anxiety and/or stress and offspring's cognitive functioning: A meta-analysis. *Child Development*, *94*(3), 779–801. https://doi.org/10.1111/cdev.13885

Duong, K. L., Yang, B. R., Yun, H., & Chae, J. (2024). Effect of methylphenidate on height in pediatric attention-deficit hyperactivity disorder patients: A systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, *33*(6), 1755–1770. https://doi.org/10.1007/s00787-023-02273-x

Guo, B.-Q., Li, H.-B., Zhai, D.-S., & Yang, L.-Q. (2024). Prevalence of autism spectrum disorder diagnosis by birth weight, gestational age, and size for gestational age: A systematic review, meta-analysis, and meta-regression. *European Child & Adolescent Psychiatry*, *33*(7), 2035–2049. https://doi.org/10.1007/s00787-022-02078-4

Movahed, F., Heidari, E., Sadeghi, D., Rezaei Nejad, A., Abyaneh, R., Zarei, M., Beigi, F., Abdollahi, A., & Shafiee, A. (2025). Incident diabetes in adolescents using antidepressant: A systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, *34*(2), 599–610. https://doi.org/10.1007/s00787-024-02502-x

Nigg, J. T., Bruton, A., Kozlowski, M. B., Johnstone, J. M., & Karalunas, S. L. (2024). Systematic review and meta-analysis: Do white noise or pink noise help with task performance in youth with attention-deficit/hyperactivity disorder or with elevated attention problems? *Journal of the American Academy of Child & Adolescent Psychiatry*, *63*(8), 778–788. https://doi.org/10.1016/j.jaac.2023.12.014

Orlowski, E. W., Bender, A. M., & Karver, M. S. (2023). A systematic review and meta-analysis of clinician behaviors and characteristics related to alliance building with youth clients. *Clinical Psychology Review*, *102*, 102273. https://doi.org/10.1016/j.cpr.2023.102273

Perrone, L., Thorpe, D., Shariat Panahi, G., Kitagawa, Y., Lindhiem, O., & Bernard, K. (2024). Meta-analysis of associations between childhood adversity and diurnal cortisol regulation. *Development and Psychopathology*, *36*(3), 1323–1355. https://doi.org/10.1017/S0954579423000561

Rapp, H., Wang Xu, J., & Enright, R. D. (2022). A meta-analysis of forgiveness education interventions' effects on forgiveness and anger in children and adolescents. *Child Development*, *93*(5), 1249–1269. https://doi.org/10.1111/cdev.13771

Salazar De Pablo, G., Pastor Jordá, C., Vaquerizo-Serrano, J., Moreno, C., Cabras, A., Arango, C., Hernández, P., Veenstra-VanderWeele, J., Simonoff, E., Fusar-Poli, P., Santosh, P., Cortese, S., & Parellada, M. (2023). Systematic review and meta-analysis: Efficacy of pharmacological interventions for irritability and emotional dysregulation in autism spectrum disorder and predictors of response. *Journal of the American Academy of Child & Adolescent Psychiatry*, *62*(2), 151–168. https://doi.org/10.1016/j.jaac.2022.03.033

Schulz, S., Nelemans, S., Hadiwijaya, H., Klimstra, T., Crocetti, E., Branje, S., & Meeus, W. (2023). The future is present in the past: A meta-analysis on the longitudinal associations of parent–adolescent relationships with peer and romantic relationships. *Child Development*, *94*(1), 7–27. https://doi.org/10.1111/cdev.13849

Serravalle, L., Trespalacios, F., & Ellenbogen, M. A. (2025). Hypothalamic–pituitary–adrenal axis functioning in offspring of parents with a major affective disorder: A meta-analytic review. *European Child & Adolescent Psychiatry*, *34*(4), 1249–1265. https://doi.org/10.1007/s00787-024-02553-0

Vermillet, A., Tølbøll, K., Litsis Mizan, S., C. Skewes, J., & Parsons, C. E. (2022). Crying in the first 12 months of life: A systematic review and meta-analysis of cross-country parent-reported data and modeling of the “cry curve.” *Child Development*, *93*(4), 1201– 1222. https://doi.org/10.1111/cdev.13760

Wrong Study Type

k = 40

Becker, T. D., Castañeda Ramirez, S., Bruges Boude, A., Leong, A., Ivanov, I., & Rice, T. R. (2024). Interventions for prevention and treatment of substance use in youth with traumatic childhood experiences: A systematic review and synthesis of the literature. *European Child & Adolescent Psychiatry*, *33*(10), 3419–3438. https://doi.org/10.1007/s00787-023-02265-x

Bhatia, D. (2024). Editorial: Cannabis legalization and youth cannabis use: Findings from an updated systematic review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, *63*(11), 1075–1077. https://doi.org/10.1016/j.jaac.2024.05.005

Carneiro, F. A. T., Leong, V., Nóbrega, S., Salinas-Quiroz, F., Costa, P. A., & Leal, I. (2024). Are the children alright? A systematic review of psychological adjustment of children conceived by assisted reproductive technologies. *European Child & Adolescent Psychiatry*, *33*(8), 2527–2546. https://doi.org/10.1007/s00787-022-02129-w

Carroll, S. L., Mikhail, M. E., & Burt, S. A. (2023). The development of youth antisocial behavior across time and context: A systematic review and integration of person-centered and variable-centered research. *Clinical Psychology Review*, *101*, 102253. https://doi.org/10.1016/j.cpr.2023.102253

Chelouche-Dwek, G., & Fonagy, P. (2025). Mentalization-based interventions in schools for enhancing socio-emotional competencies and positive behaviour: A systematic review. *European Child & Adolescent Psychiatry*, *34*(4), 1295–1315. https://doi.org/10.1007/s00787-024-02578-5

Cook, B., Van Bockstaele, S., Crow, S. B., Sainsbury, D., Butterworth, S., & Filson, S. (2024). Neurodevelopmental disorders in children with cleft lip and palate: A systematic review. *European Child & Adolescent Psychiatry*. https://doi.org/10.1007/s00787-024-02636-y

Coutts-Bain, D., Sharpe, L., Techakesari, P., Forrester, M. A., & Hunt, C. (2023). A mixed-methods review and meta-synthesis of fears of recurrence and

progression in people with mental health conditions. *Clinical Psychology Review*, *105*, 102342. https://doi.org/10.1016/j.cpr.2023.102342

Cruciani, G., Quintigliano, M., Mezzalira, S., Scandurra, C., & Carone, N. (2024). Attitudes and knowledge of mental health practitioners towards LGBTQ+ patients: A mixed-method systematic review. *Clinical Psychology Review*, *113*, 102488. https://doi.org/10.1016/j.cpr.2024.102488

Demetriou, E. A., Boulton, K. A., Bowden, M. R., & Guastella, A. J. (2024). Adjustment and homesickness in hospitalised children: A systematic review. *Clinical Psychology Review*, *110*, 102431. https://doi.org/10.1016/j.cpr.2024.102431

Fang, Y., Luo, J., Boele, M., Windhorst, D., Van Grieken, A., & Raat, H. (2024). Parent, child, and situational factors associated with parenting stress: A systematic review. *European Child & Adolescent Psychiatry*, *33*(6), 1687–1705. https://doi.org/10.1007/s00787-022-02027-1

Fornefeld, D., Zellin, U., Schmidt, P., & Fricke, O. (2025). The supporting role of dogs in the inpatient setting: A systematic review of the therapeutic effects of animal-assisted therapy with dogs for children and adolescents in an inpatient setting. *European Child & Adolescent Psychiatry*, *34*(1), 3–17. https://doi.org/10.1007/s00787-023-02326-1

Fusar-Poli, P., & Salazar De Pablo, G. (2022). No meta-analytic effect of age on probability of developing psychosis in individuals at clinical high risk. *Journal of the American Academy of Child & Adolescent Psychiatry*, *61*(6), 731–732. https://doi.org/10.1016/j.jaac.2021.11.027

Gallop, L., Westwood, S. J., Hemmings, A., Lewis, Y., Campbell, I. C., & Schmidt, U. (2025). Effects of repetitive transcranial magnetic stimulation in children and young people with psychiatric disorders: A systematic review. *European Child & Adolescent Psychiatry*, *34*(2), 403–422. https://doi.org/10.1007/s00787-024-02475-x

Gallop, L., Westwood, S. J., Lewis, Y., Campbell, I. C., & Schmidt, U. (2024). Effects of transcranial direct current stimulation in children and young people with psychiatric disorders: A systematic review. *European Child &*

Adolescent Psychiatry, *33*(9), 3003– 3023. https://doi.org/10.1007/s00787-023-02157-0

Girela-Serrano, B. M., Spiers, A. D. V., Ruotong, L., Gangadia, S., Toledano, M. B., & Di Simplicio, M. (2024). Impact of mobile phones and wireless devices use on children and adolescents' mental health: A systematic review. *European Child & Adolescent Psychiatry*, *33*(6), 1621–1651. https://doi.org/10.1007/s00787-022-02012-8

Hayes, D., Mansfield, R., Mason, C., Santos, J., Moore, A., Boehnke, J., Ashworth, E., Moltrecht, B., Humphrey, N., Stallard, P., Patalay, P., & Deighton, J. (2024). The impact of universal, school based, interventions on help seeking in children and young people: A systematic literature review. *European Child & Adolescent Psychiatry*, *33*(9), 2911– 2928. https://doi.org/10.1007/s00787-022-02135-y

Kong, L., Nivins, S., Chen, X., Liang, Y., Gissler, M., & Lavebratt, C. (2025). Association of preterm birth and birth size status with neurodevelopmental and psychiatric disorders in spontaneous births. *European Child & Adolescent Psychiatry*, *34*(1), 261–273. https://doi.org/10.1007/s00787-024-02489-5

Koumoula, A., Marchionatti, L. E., Caye, A., Karagiorga, V. E., Balikou, P., Lontou, K., Arkoulaki, V., Simioni, A., Serdari, A., Kotsis, K., Basta, M., Kapsimali, E., Mitropoulou, A., Klavdianou, N., Zeleni, D., Mitroulaki, S., Botzaki, A., Gerostergios, G., Samiotakis, G., … Salum, G. A. (2024). The science of child and adolescent mental health in Greece: A nationwide systematic review. *European Child & Adolescent Psychiatry*, *33*(10), 3359–3375. https://doi.org/10.1007/s00787-023-02213-9

Lång, U., & Kelleher, I. (2022). Dr. Lång et al. reply. *Journal of the American Academy of Child & Adolescent Psychiatry*, *61*(6), 733–734. https://doi.org/10.1016/j.jaac.2021.12.005

Lee, K. S., Hagan, C. N., Hughes, M., Cotter, G., McAdam Freud, E., Kircanski, K., Leibenluft, E., Brotman, M. A., & Tseng, W.-L. (2023). Systematic review and meta-analysis: Task-based fMRI studies in youths with

irritability. *Journal of the American Academy of Child & Adolescent Psychiatry*, *62*(2), 208–229. https://doi.org/10.1016/j.jaac.2022.05.014

Merín, L., Toledano-González, A., Fernández-Aguilar, L., Nieto, M., Del Olmo, N., & Latorre, J. M. (2024). Evaluation of the association between excessive screen use, sleep patterns and behavioral and cognitive aspects in preschool population. A systematic review. *European Child & Adolescent Psychiatry*, *33*(12), 4097–4114. https://doi.org/10.1007/s00787-024-02430-w

Miu, A. C., Szentágotai-Tătar, A., Balázsi, R., Nechita, D., Bunea, I., & Pollak, S. D. (2022). Emotion regulation as mediator between childhood adversity and psychopathology: A meta-analysis. *Clinical Psychology Review*, *93*, 102141. https://doi.org/10.1016/j.cpr.2022.102141

Nivison, M. D., Filetti, C. R., Carlson, E. A., Jacobvitz, D. B., & Roisman, G. I. (2024). Agreement between retrospective and prospective assessments of childhood abuse revisited. *Development and Psychopathology, Advance online publication.* https://doi.org/10.1017/S0954579424001032

Oh, S., Choi, J., Han, D. H., & Kim, E. (2024). Effects of game-based digital therapeutics on attention deficit hyperactivity disorder in children and adolescents as assessed by parents or teachers: A systematic review and meta-analysis. *European Child & Adolescent Psychiatry*, *33*(2), 481–493. https://doi.org/10.1007/s00787-023-02174-z

Ollerup, D. R., Rasmussen, L., Elkrog, S. L., Stoltz-Andersen, M. H., Stubmark, H., & Wesselhoeft, R. (2025). Psychotropic drug use among children and adolescents in the Nordic countries: A systematic review. *European Child & Adolescent Psychiatry*, *34*(3), 903–919. https://doi.org/10.1007/s00787-024-02545-0

Pedruzo, B., Aymerich, C., Pacho, M., Herrero, J., Laborda, M., Bordenave, M., Giuliano, A. J., McCutcheon, R. A., Gutiérrez-Rojas, L., McGuire, P., Stone, W. S., Fusar-Poli, P., González-Torres, M. Á., & Catalan, A. (2024). Longitudinal change in neurocognitive functioning in children and adolescents at clinical high risk for psychosis: A systematic review.

European Child & Adolescent Psychiatry, *33*(10), 3377–3387. https://doi.org/10.1007/s00787-023-02221-9

Price, M. A., Weisz, J. R., McKetta, S., Hollinsaid, N. L., Lattanner, M. R., Reid, A. E., & Hatzenbuehler, M. L. (2022). Meta-analysis: Are psychotherapies less effective for black youth in communities with higher levels of anti-black racism? *Journal of the American Academy of Child & Adolescent Psychiatry*, *61*(6), 754–763. https://doi.org/10.1016/j.jaac.2021.07.808

Richard, J., & King, S. M. (2023). Annual research review: Emergence of problem gambling from childhood to emerging adulthood: a systematic review. *Journal of Child Psychology and Psychiatry*, *64*(4), 645–688. https://doi.org/10.1111/jcpp.13713

Rienecke, R. D., Trotter, X., & Jenkins, P. E. (2024). A systematic review of eating disorders and family functioning. *Clinical Psychology Review*, *112*, 102462. https://doi.org/10.1016/j.cpr.2024.102462

Roach, E. L., Haft, S. L., Huang, J., & Zhou, Q. (2023). Systematic review: The association between race-related stress and trauma and emotion dysregulation in youth of color. *Journal of the American Academy of Child & Adolescent Psychiatry*, *62*(2), 190–207. https://doi.org/10.1016/j.jaac.2022.04.013

Seker, A., Qirko-Gurakuqi, A., Tabaku, M., Javate, K. R. P., & Rathwell, I. (2024). Maternal atopic conditions and autism spectrum disorder: A systematic review. *European Child & Adolescent Psychiatry*, *33*(11), 3727–3737. https://doi.org/10.1007/s00787-023-02285-7

Shaffer, R. C., Reisinger, D. L., Schmitt, L. M., Lamy, M., Dominick, K. C., Smith, E. G., Coffman, M. C., & Esbensen, A. J. (2023). Systematic review: Emotion dysregulation in syndromic causes of intellectual and developmental disabilities. *Journal of the American Academy of Child & Adolescent Psychiatry*, *62*(5), 518–557. https://doi.org/10.1016/j.jaac.2022.06.020

Sibley, M. H., Flores, S., Murphy, M., Basu, H., Stein, M. A., Evans, S. W., Zhao, X., Manzano, M., & Van Dreel, S. (2025). Research review:

Pharmacological and non pharmacological treatments for adolescents with attention deficit/hyperactivity disorder – a systematic review of the literature. *Journal of Child Psychology and Psychiatry*, *66*(1), 132–149. https://doi.org/10.1111/jcpp.14056

Siegel, P., & Peterson, B. S. (2025). Advancing the treatment of anxiety disorders in transition-age youth: A review of the therapeutic effects of unconscious exposure. *Journal of Child Psychology and Psychiatry*, *66*(1), 98–121. https://doi.org/10.1111/jcpp.14037

Sigrist, C., Vöckel, J., MacMaster, F. P., Farzan, F., Croarkin, P. E., Galletly, C., Kaess, M., Bender, S., & Koenig, J. (2022). Transcranial magnetic stimulation in the treatment of adolescent depression: A systematic review and meta-analysis of aggregated and individual-patient data from uncontrolled studies. *European Child & Adolescent Psychiatry*, *31*(10), 1501–1525. https://doi.org/10.1007/s00787-022-02021-7

Speroni, J., Fanniff, A. M., Edgemon, J. M., Martini, V., & Haas, A. L. (2023). Alcohol mixed with energy drinks and aggressive behaviors in adolescents and young adults: A systematic review. *Clinical Psychology Review*, *104*, 102319. https://doi.org/10.1016/j.cpr.2023.102319

Taubner, S., Ioannou, Y., Saliba, A., Sales, C. M. D., Volkert, J., Protić, S., Adler, A., Barkauskiene, R., Conejo-Cerón, S., Di Giacomo, D., Mestre, J. M., Moreno-Peral, P., Vieira, F. M., Mota, C. P., Henriques, M. I. R. S., Røssberg, J. I., Perdih, T. S., Schmidt, S. J., Zettl, M., … Heinonen, E. (2024). Mediators of outcome in adolescent psychotherapy and their implications for theories and mechanisms of change: A systematic review. *European Child & Adolescent Psychiatry*, *33*(10), 3321–3358. https://doi.org/10.1007/s00787-023-02186-9

Vijayakumar, N., & Whittle, S. (2023). A systematic review into the role of pubertal timing and the social environment in adolescent mental health problems. *Clinical Psychology Review*, *102*, 102282. https://doi.org/10.1016/j.cpr.2023.102282

Wols, A., Pingel, M., Lichtwarck-Aschoff, A., & Granic, I. (2024). Effectiveness of applied and casual games for young people's mental health: A systematic review of randomised controlled studies. *Clinical Psychology Review*, *108*, 102396. https://doi.org/10.1016/j.cpr.2024.102396

Wu, B., Zhang, X., Xie, H., Zhang, B., Ling, Y., Gan, R., Qiu, L., Roberts, N., Jia, Z., & Gong, Q. (2024). Research review: Shared and distinct structural and functional brain alterations in adolescents with major depressive disorder' – a multimodal meta-analysis. *Journal of Child Psychology and Psychiatry*, jcpp.14104. https://doi.org/10.1111/jcpp.14104

Submission Date Before March 2021

k = 2

Burnette, J. L., Billingsley, J., Banks, G. C., Knouse, L. E., Hoyt, C. L., Pollack, J. M., & Simon, S. (2023). A systematic review and meta-analysis of growth mindset interventions: For whom, how, and why might such interventions work? *Psychological Bulletin*, *149*(3–4), 174–205. https://doi.org/10.1037/bul0000368

Ritchie, M. B., Neufeld, R. W. J., Yoon, M., Li, A., & Mitchell, D. G. V. (2022). Predicting youth aggression with empathy and callous unemotional traits: A meta-analytic review. *Clinical Psychology Review*, *98*, 102186. https://doi.org/10.1016/j.cpr.2022.102186

Appendix E: Transparency Checklist

Table 4

Transparency Checklist Items Corresponding to AMSTAR 2 Question 2: Did the Report of the Review Contain an Explicit Statement that the Review Methods were Established Prior to the Conduct of the Review and Did the Report Justify Any Significant Deviations From the Protocol?, k = 66

ID	PRISMA Item	Text	Frequency	Proportion
07.04	Search Strategy	Did the author(s) report updating their search? *If author(s) state that search was run once, code "Yes – no update".*		
		a) Yes – full search	22	33.3%
		b) Yes – only selected information sources	1	1.5%
		c) Yes – no update	12	18.2%
		d) No – not reported	31	47.0%
24.01	Registration and protocol	Did the author(s) report working from a protocol or a priori established methods? *There is no need to do an internet search for a protocol to answer this question.*		
		a) Yes, and protocol is publicly available (author(s) cite the bibliographic reference for the published protocol in the main text)	0	
		b) Yes, and protocol is publicly available (author(s) cite a website where the protocol can be accessed)	43	65.1%
		c) Yes, but protocol is not publicly available (i.e. author(s) state that they worked from a protocol but do not report the location of the protocol, or only report that it is available on request)	5	7.6%
		d) Unclear - the use of a protocol was only implied (e.g., by the term 'pre-specified'), please describe why you think a protocol was used	0	
		e) No - they reported not working from a protocol	5	7.6%
		f) Not reported	13	19.7%

ID	PRISMA Item	Text	Frequency	Proportion
24.02	Registration and protocol	If Yes – protocol was used: Where was it registered? (select all that apply)		
		a) PROSPERO	39	83.0%
		b) OSF	8	17.0%
		c) As a journal article	0	
		d) ResearchGate	0	
		e) INPLASY	0	
		f) Research Registry-Registry of Systematic Reviews/Meta-Analyses	0	
		g) protocols.io	0	
		h) Other, please specify: Freetext	0	
24.03	Registration and protocol	Did the author(s) report any amendments from the original protocol?		
		a) Yes – they reported amendments	17	37.0%
		b) Yes – they reported that no amendments were made	0	
		c) Not reported	29	63.0%
		d) Not applicable	20	
24.04	Registration and protocol	If an amendment was made: Did the author(s) note the reason for the amendment?		
		a) Yes	15	88.2%
		b) No	2	11.8%
		c) Not applicable	49	
24.05	Registration and protocol	If an amendment was made: Did the author(s) note the stage of the review when the amendment was made?		
		a) Yes	7	41.2%
		b) No	10	58.8%
		c) Not applicable	49	

Note. Frequencies indicate the number of reviews meeting each criterion. Percentages are based on the total sample (k = 66) unless otherwise specified. For items with responses coded as *Not applicable*, these cases were excluded from percentage calculations, resulting in reduced denominators for some items.

Table 5

Transparency Checklist Items Corresponding to AMSTAR 2 Question 4: Did the Review Authors Use a Comprehensive Literature Search Strategy?, k = 66

ID	PRISMA Item	Text	Frequency	Proportion
06.01	Information Sources and Methods	Did the author(s) search bibliographic databases?		
		a) Yes	66	100%
		b) No	0	
06.03	Information Sources and Methods	If the author(s) reported searching one or several trial registers (e.g., ClinicalTrials.gov, WHO International Clinical Trials Registry Platform), did they report: registry name, citation/link, date restrictions, date searched?		
		a) Yes	1	10.0%
		b) Partially	7	70.0%
		c) No	2	20.0%
		d) Not applicable	56	
06.06	Information Sources and Methods	If the author(s) contacted individuals to identify studies, was the type of individuals specified? (e.g., researchers in the field)		
		a) Yes	12	100%
		b) Partially	0	
		c) No	0	
		d) Not applicable	56	
07.01	Search Strategies	Did the author(s) report whether search limits or filters were used?		
		a) Yes – search limits/filters used	26	36.4%
		b) Yes – no search limits/filters used	11	16.7%
		c) No – not reported	29	43.9%

ID	PRISMA Item	Text	Frequency	Proportion
07.02	Search Strategies	If Yes: Did the author(s) provide justification for the use of the limits applied to the search(es)? Note: See PRISMA-S p.11 for details if needed.		
		a) Yes	9	25.0%
		b) No	27	75.0%
		c) Not applicable	30	
07.03	Search Strategies	How did the author(s) arrive at their search terms? (select all that apply) *Code "expert opinion" when they indicated they thought of the search terms themselves.* *Code "key literature" when they surveyed the literature in the field for possible terms. This is not the same as adapting search terms from existing previous systematic reviews. Code "not cited" for search terms from previous systematic reviews, when they did not specify from which systematic reviews these terms were taken.*		
		a) Search terms from previous systematic reviews (cited)	5	7.6%
		b) Search terms from previous systematic reviews (not cited)	1	1.5%
		c) Expert opinion	8	12.1%
		d) Key literature	0	
		e) Other: Freetext	0	
		f) Not reported	52	78.8%
07.04	Search Strategy	Did the author(s) report updating their search? *If author(s) state that search was run once, code "Yes – no update".*		
		a) Yes – full search	22	33.3%
		b) Yes – only selected information sources	1	1.5%
		c) Yes – no update	12	18.2%
		d) No – not reported	31	47.0%
07.06	Search Strategies	Was the search strategy peer-reviewed? *In PRISMA-S, this refers to the peer review of search strategies prior to executing the search. Peer review is used to identify errors, missing keywords or subject headings, and other issues within a search strategy. One commonly used tool for search strategy peer review is the Peer Review of Electronic*		

ID	PRISMA Item	Text	Frequency	Proportion
		Search Strategies (PRESS) Guideline. Code "yes – it was peer-reviewed by another researcher" (both options), when another researcher not involved in the initial search string development validated/checked the search string.		
		a) Yes – it was peer-reviewed by another researcher using a formal tool (e.g., PRESS)	0	
		b) Yes – it was peer-reviewed by another researcher (without formal tool)	1	1.5%
		c) Yes – it was not peer-reviewed	0	
		d) No – not reported	65	98.5%
07.09		If the search strategy was peer-reviewed: Did the author(s) report any details on the peer-review process and/or the tool used?		
		a) Yes	0	
		b) No	1	100%
		c) Not applicable	65	
07.07	Search Strategies	If the search strategy was validated (e.g., identification of a set of clearly eligible studies): Did the author(s) report details on this validation procedure and the validation set of studies?		
		a) Yes	0	
		b) No	0	
		c) Not applicable	66	100%

Note. Frequencies indicate the number of reviews meeting each criterion. Percentages are based on the total sample (k = 66) unless otherwise specified. For items with responses coded as *Not applicable*, these cases were excluded from percentage calculations, resulting in reduced denominators for some items.

Table 6

Transparency Checklist Items Corresponding to AMSTAR 2 Question 7: Did the Review Authors Provide a List of Excluded Studies and Justify the Exclusions?, k = 66

ID	PRISMA Item	Text	Frequency	Proportion
16.02	Study Selection	Did the author(s) cite studies that might appear to meet inclusion criteria, but were excluded at fulltext stage?		
		a) Yes, a reference and reason for every excluded article was provided	9	13.6%
		b) No, only a list of reasons (with or without number of studies) was provided, but not on a per-study basis	52	78.8%
		c) No – not reported	5	7.6%

Note. Frequencies indicate the number of reviews meeting each criterion. Percentages are based on the total sample (*k* = 66) unless otherwise specified. For items with responses coded as *Not applicable*, these cases were excluded from percentage calculations, resulting in reduced denominators for some items.

Table 7

Transparency Checklist Items Corresponding to AMSTAR 2 Question 9: Did the Review Authors Use a Satisfactory Technique for Assessing the Risk of Bias (RoB) in Individual Studies that Were Included in the Review?, and AMSTAR 2 Question 13: Did the review authors account for RoB in individual studies when interpreting/ discussing the results of the review?, k = 66

ID	PRISMA Item	Text	Frequency	Proportion
11.01	Study risk of bias assessment	Did the author(s) report having undertaken a risk of bias (or quality) assessment?		
		a) Yes – undertaken	53	80.3%
		b) No – argued why not undertaken	1	1.5%
		c) No – not undertaken or not reported	12	18.2%

11.02	Study risk of bias assessment	If Yes – undertaken: Did the author(s) specify the tool (validated or self-developed) to assess risk of bias or study quality?		
		a) Yes (name and citation provided for validated tools, content provided for self-developed tools)	51	96.2%
		b) No	2	3.8%
		c) Not applicable	13	
11.03	Study risk of bias assessment	If a non-standardized and/or new tool (e.g., one specially developed for the review) was used to assess risk of bias: Did the author(s) report the rules they applied to reach an overall judgement?		
		a) Yes	2	50%
		b) No	2	50%
		c) Not applicable	62	
11.08	Study risk of bias assessment	If adaptations to an existing tool were made: Did the author(s) describe these adaptations?		
		a) Yes	14	73.7%
		b) No	5	26.3%
		c) Not applicable	47	
11.09	Study risk of bias assessment	If a new tool was developed: Did the authors make it publicly accessible?		
		a) Yes	2	100%
		b) No	0	
		c) Not applicable	64	
11.04	Study risk of bias assessment	If undertaken: how many author(s) were involved in the risk of bias (or quality) assessment?		
		a) All included studies assessed by at least two author(s)	31	58.5%
		b) All included studies assessed by one author, with verification by another	4	7.5%
		c) All included studies assessed by only one author	1	1.9%
		d) Other (please specify): Freetext	1	1.9%
		e) Not reported	16	30.2%
		f) Not applicable	13	

11.10	Study risk of bias assessment	If more than one author: Did the author(s) report how discrepancies were resolved?		
		a) Yes	23	63.9%
		b) No	13	36.1%
		c) Not applicable	30	
11.05	Study risk of bias assessment	If undertaken: Did the author(s) report using any automation tool(s) for assessing risk of bias?		
		a) Yes – they used (an) automation tool(s)	0	
		b) Yes – they reported they did not use (an) automation tool(s)	0	
		c) No – not reported	52	100%
		d) Not applicable	14	
11.07	Study risk of bias assessment	Did the author(s) report whether they contacted (or attempted to contact) corresponding author(s) of included studies for any unpublished data (e.g., unreported outcomes, unreported information required for the risk of bias assessment)?		
		a) Yes - Review author(s) reported that they contacted (or attempted to contact) author(s) of included studies	1	1.9%
		b) Yes - Review author(s) reported that they DID NOT contact author(s) of included studies	1	1.9%
		c) No such statement was made	51	96.2%
		d) Not applicable	13	
18.01	Risk of bias in studies	Did the author(s) present the risk of bias assessments for each included study?		
		a) Yes	45	84.9%
		b) No – even though risk of bias assessment was reported	8	15.1%
		c) Not applicable	13	
18.02	Risk of bias in studies	Did the author(s) provide justification for their risk of bias assessment?		
		a) Yes	11	20.8%
		b) No	42	79.2%
		c) Not applicable	13	

21.02	Certainty of evidence	Did the author(s) present their assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis?		
		a) Yes	44	75.9%
		b) No	14	24.1%
		c) Not applicable	8	

Note. Frequencies indicate the number of reviews meeting each criterion. Percentages are based on the total sample (k = 66) unless otherwise specified. For items with responses coded as *Not applicable*, these cases were excluded from percentage calculations, resulting in reduced denominators for some items.

Table 8

Transparency Checklist Items Corresponding to AMSTAR 2 Question 11: If Meta-Analysis Was Performed Did the Review Authors Use Appropriate Methods for Statistical Combination of Results?, k = 66

ID	PRISMA Item	Text	Frequency	Proportion
12.03	Effect measures	Did the author(s) state thresholds/ranges used to interpret the size of the effect(s) and the rationale? A citation sufficiently covers the threshold/range and rationale. *Code NA for prevalence studies*		
		a) Yes	35	58.3%
		b) No	25	41.7%
		c) Not applicable	5	
13.05	Synthesis methods	Did the author(s) specify the meta-analysis weighting method used (e.g., Mantel-Haenszel, inverse-variance)? (Note: such information may appear in the Methods section or on the forest plot)		
		a) Yes	16	24.6%
		b) No	49	75.4%
		c) Not applicable	1	
13.06	Synthesis methods	If Yes: Which weighting method was used?		
		a) Inverse variance	8	50.0%
		b) Mantel-Haenszel	1	6.25%
		c) Peto	0	
		d) Sample size	1	6.25%
		e) Other, please specify: Freetext	6	37.5%
		f) Not applicable	50	
13.14	Synthesis methods	Did the author(s) report using any method to formally investigate heterogeneity among effect sizes (e.g., subgroup analysis, meta-regression)?		
		a) Yes – conducted	63	95.5%

ID	PRISMA Item	Text	Frequency	Proportion
		b) Yes – not conducted (also includes methods used to explore heterogeneity in studies/effect sizes not amenable for meta-analysis, e.g., structured tables.)	1	1.5%
		c) No – not reported	2	3.0%
13.15	Synthesis methods	If Yes - conducted: Did the author(s) report which moderator variables, including their type (continuous; categorical) and levels (if categorical), were tested? This needs to be specified within the methods section.		
		a) Yes	32	50.8%
		b) No	31	49.2%
		c) Not applicable	3	

Note. Frequencies indicate the number of reviews meeting each criterion. Percentages are based on the total sample (k = 66) unless otherwise specified. For items with responses coded as *Not applicable*, these cases were excluded from percentage calculations, resulting in reduced denominators for some items.

Table 9

Transparency Checklist Items Corresponding to AMSTAR 2 Question 15: If They Performed Quantitative Synthesis Did the Review Authors Carry Out an Adequate Investigation of Publication Bias (Small Study Bias) and Discuss its Likely Impact on the Results of the Review?, k = 66

ID	PRISMA Item	Text	Frequency	Proportion
14.01	Reporting bias assessment	Which methods did the author(s) use to assess reporting bias?		
		a) Statistical (i.e., publication bias detection methods)	53	80.3%
		b) Graphical (e.g., funnel plot)	41	62.1%
		c) Tools (e.g., checklists, scales, domain-based tools)	0	
		d) Other, please specify: Freetext	0	
		e) Not reported	8	12.1%
14.11	Reporting bias assessment	IF A STATISTICAL METHOD WAS USED: Did the author(s) specify the thresholds used to decide whether publication bias (as indicated by this method) was present?		
		a) Yes	6	11.3%
		b) No	47	88.7%
		c) Not applicable	13	
14.05	Reporting bias assessment	IF TOOLS WERE USED: Did the author(s) specify the methodological components/domains/items of the tool?		
		a) Yes	0	
		b) No	0	
		c) Not applicable	66	100%
14.07	Reporting bias assessment	IF GRAPHICAL METHODS OR TOOLS WERE USED: Did the author(s) report how many reviewers were involved in the reporting bias assessment?		
		a) Yes	0	
		b) No	39	100%
		c) Not applicable	27	

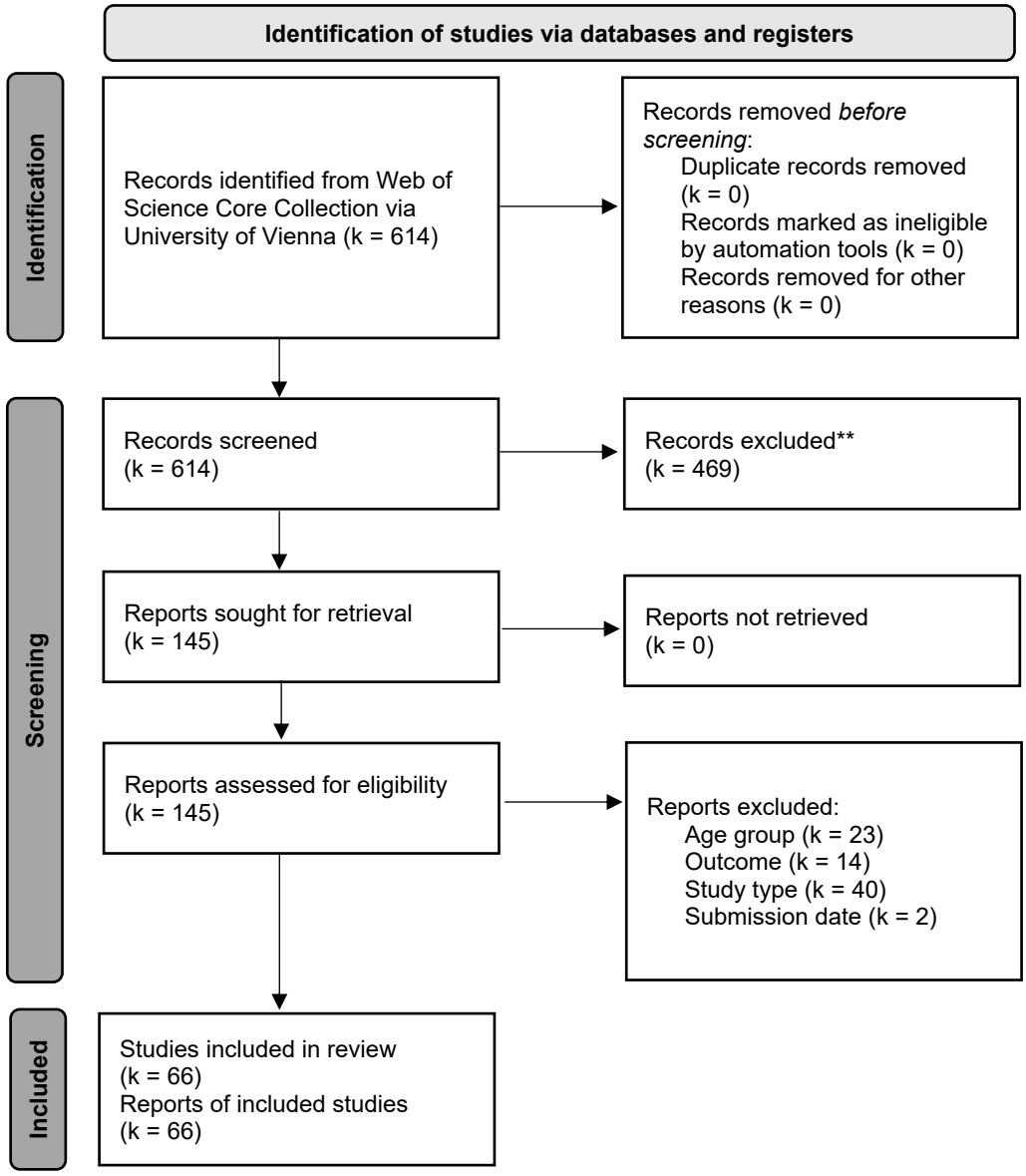
14.08	Reporting bias assessment	If more than one: Did the author(s) report whether they worked independently?		
		a) Yes	0	
		b) No	0	
		c) Not applicable	66	100%
14.09	Reporting bias assessment	If more than one: Did the author(s) report how discrepancies were resolved?		
		a) Yes	0	
		b) No	0	
		c) Not applicable	66	100%
21.01	Certainty of evidence	If a tool was used to assess reporting bias, did the author(s) report the results for each item included in the tool?		
		a) Yes	0	
		b) No	0	
		c) Not applicable	66	100%

Note. Frequencies indicate the number of reviews meeting each criterion. Percentages are based on the total sample (k = 66) unless otherwise specified. For items with responses coded as *Not applicable*, these cases were excluded from percentage calculations, resulting in reduced denominators for some items.

Appendix F: PRISMA 2020 Flow Diagram

Figure 4

Flow Diagram of the Study Selection Process in Accordance with PRISMA Guidelines (Page et al., 2021b)



Note. Adapted from Page et al. (2021). Counts in the diagram follow the original convention using *n*. Elsewhere in this thesis, *k* denotes meta-analytic systematic reviews, and *n* is used to refer to participant.

Appendix G: Association of Methodological Quality and Statistical Consistency

Figure 5

Extractable Test-Statistics by Statcheck

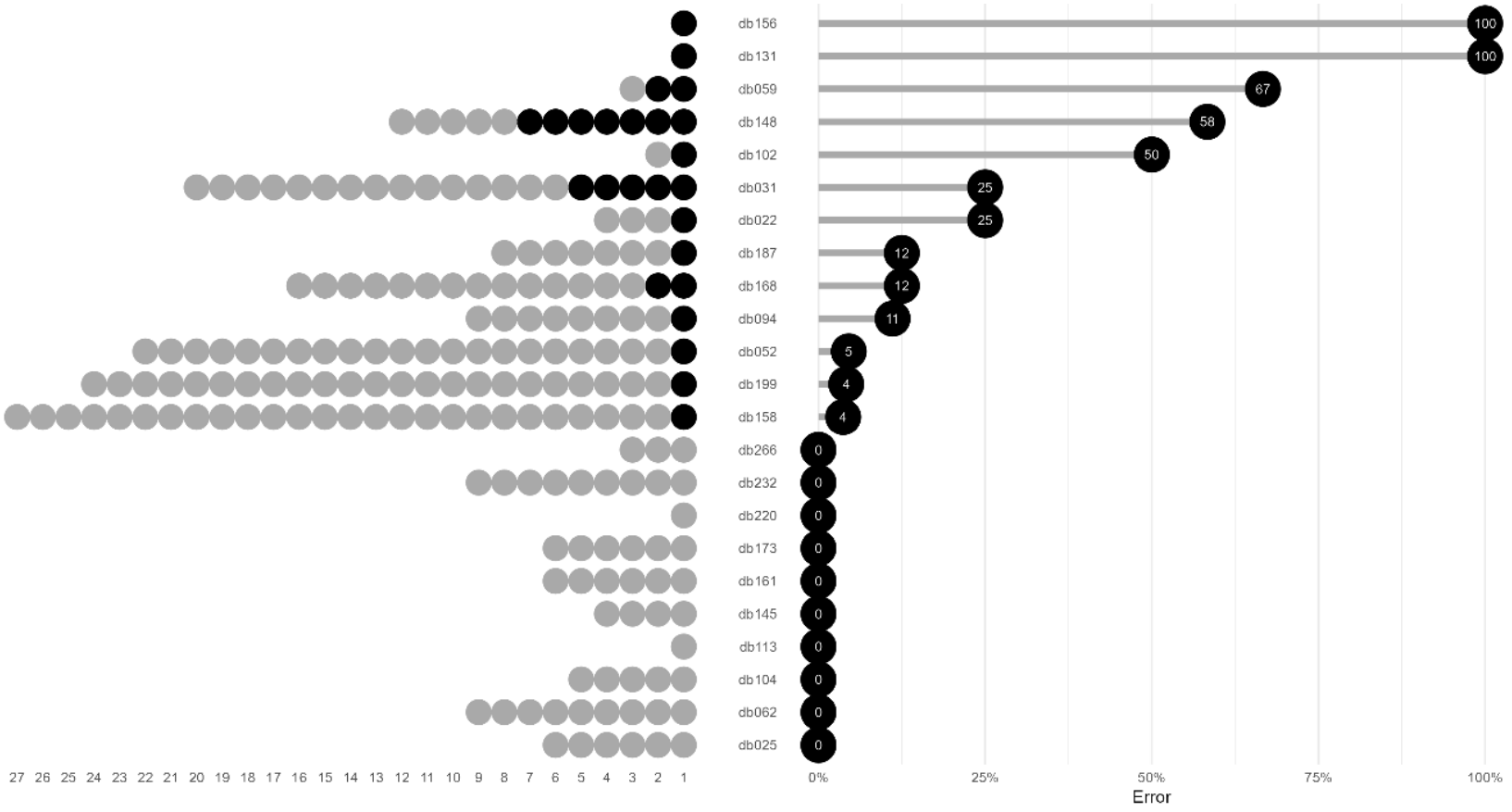


Table 10

Study Level Variables of Included Meta-Analytic Systematic Reviews

Citation	AMSTAR sum	≥1 statistical inconsistency
Alozkan Sever et al. (2023)	13.5	NA
Armoon et al. (2025)	12.0	NA
Bapuji et al. (2024)	9.0	NA
Bellato et al. (2024)	15.5	NA
Betancourt et al. (2024)	7.0	NA
Bloss et al. (2024)	12.5	Yes
Bo et al. (2023)	13.0	NA
Boldrini et al. (2023)	14.0	NA
Boumpa et al. (2024)	13.5	NA
Breaux et al. (2023)	8.5	NA
Cooke et al. (2022)	9.5	No
Cushing et al. (2024)	11.5	No
Davis et al. (2023)	13.5	NA
Dean et al. (2025)	12.5	NA
Defoe et al. (2023)	5.5	No
Dülsen and Baumeister, (2025)	10.0	NA

Citation	AMSTAR sum	≥1 statistical inconsistency
England- Mason et al. (2023)	10.0	No
Ezquerra et al. (2024)	8.5	NA
Farkas et al. (2024)	5.0	No
Finlay-Jones et al. (2024)	12.0	NA
Francis et al. (2023)	13.5	NA
Frisira et al. (2025)	11.0	NA
Gu et al. (2025)	13.0	NA
Guzman-Holst et al. (2022)	13.5	NA
Haza et al. (2024)	15.0	NA
Holzman et al. (2022)	9.0	NA
Jensen and Thomsen (2024)	6.5	NA
Joseph et al. (2023)	9.5	NA
Khazdouz et al. (2025)	12.0	NA
Kim et al. (2024)	11.5	NA
Kopp et al. (2024)	9.5	NA
Lång et al. (2022)	10.5	NA
Lee et al. (2025)	9.0	NA
Li et al. (2024)	8.5	NA

Citation	AMSTAR sum	≥1 statistical inconsistency
Lin et al. (2024)	13.0	NA
Loheide-Niesmann et al. (2024)	12.5	Yes
Maciel et al. (2023)	11.5	NA
Manuele et al. (2023)	11.5	NA
McKern et al. (2023)	12.0	Yes
Moltrecht et al. (2024)	10.0	NA
Newlove Delgado et al. (2023)	11.0	NA
Nieto- Retuerto et al. (2024)	12.0	Yes
O'Halloran et al. (2022)	12.5	NA
Perlstein et al. (2023)	12.5	NA
Restoy et al. (2024)	13.5	NA
Robberegt et al. (2023)	13.0	NA
Sacco et al. (2024)	10.5	NA
Shen et al. (2023)	13.5	No
Shephard et al. (2022)	12.5	Yes
Siciliano et al. (2022)	10.0	NA
Somers et al. (2024)	7.5	NA
Tang et al. (2024)	11.0	No

Citation	AMSTAR sum	≥1 statistical inconsistency
Tindall et al. (2024)	13.5	Yes
Trompeter et al. (2024)	13.5	No
Tung et al. (2024)	10.0	Yes
Van IJzendoorn et al. (2023)	12.5	Yes
Van Meter et al. (2023)	8.5	NA
Venturo-Conerly et al. (2023)	13.0	NA
Walsh et al. (2022)	14.0	NA
Wang et al. (2022)	9.0	No
Wang et al. (2024)	13.5	NA
Wendel et al. (2024)	14.5	NA
Wergeland et al. (2025)	11.5	No
Woolgar et al. (2022)	11.0	NA
Yang et al. (2022)	9.0	NA
Zhang and Ji (2024)	6.0	NA

Note. NA=Not available (*statcheck* could not extract any test statistics).

Appendix H: Study Characteristics

Table 11

Study Characteristics of Included Meta-Analytic Systematic Reviews

Citation	Journal	*k* Studies	*n* Children	*n* Parents	Age Group(s)	Child outcome(s)	Intervention Focus
Alozkan-Sever et al. (2023)	Journal of Child Psychology and Psychiatry	31	6,123	—	Children (2-10 years); Adolescents (11-20 years)	Depression; anxiety; PTSD	Yes
Armoon et al. (2025)	European Child & Adolescent Psychiatry	101	124,478	—	Adolescents (11-20 years)	General mental health and psychopathology	No
Bapuji et al. (2024)	Clinical Psychology Review	56	40,551 (including youth and parents)		Infants (0-24 months); Children (2-10 years); Adolescents (11-20 years)	General mental health and psychopathology	No
Bellato et al. (2024)	Journal of the American Academy of Child and Adolescent Psychiatry	12	1,016	—	Children (2-10 years); Adolescents (11-20 years)	Emotional dysregulation	No
Betancourt et al. (2024)	Clinical Psychology Review	26	13,403	—	Children (2-10 years); Adolescents (11-20 years)	ADHD	Yes

Citation	Journal	*k* Studies	*n* Children	*n* Parents	Age Group(s)	Child outcome(s)	Intervention Focus
Bloss et al. (2024)	European Child & Adolescent Psychiatry	24	NR	—	Children (2-10 years); Adolescents (11-20 years)	Internalizing; externalizing	Yes
Bo et al. (2023)	Clinical Psychology Review	30	NR	—	Adolescents (11-20 years)	Substance use	Yes
Boldrini et al. (2023)	Journal of the American Academy of Child and Adolescent Psychiatry	16	1,954	—	Adolescents (11-20 years)	Disruptive behavior disorder	Yes
Boumpa et al. (2024)	European Child & Adolescent Psychiatry	28	28,693	—	Children (2-10 years); Adolescents (11-20 years)	PTSD	No
Breaux et al. (2023)	Journal of the American Academy of Child and Adolescent Psychiatry	101	6,953	—	Children (2-10 years); Adolescents (11-20 years)	ASD; ADHD; disruptive behavior; emotional dysregulation	Yes
Cooke et al. (2022)	Child Development	108	40,147	—	Infants (0-24 months); Children (2-10 years);	Internalizing; externalizing	No

Citation	Journal	*k* Studies	*n* Children	*n* Parents	Age Group(s)	Child outcome(s)	Intervention Focus
					Adolescents (11-20 years)		
Cushing et al. (2024)	Development and Psychopathology	30	10,431	—	Children (2-10 years); Adolescents (11-20 years)	PTSD	No
Davis et al. (2023)	Journal of the American Academy of Child and Adolescent Psychiatry	42	5,998	—	Infants (0-24 months); Children (2-10 years); Adolescents (11-20 years)	PTSD	Yes
Dean et al. (2025)	Clinical Psychology Review	20	2,041	—	Adolescents (11-20 years)	Depression	Yes
Defoe et al. (2023)	Journal of Child Psychology and Psychiatry	20	144,815	4,199	Adolescents (11-20 years)	Substance use	No
Dülsen and Baumeister, (2025)	European Child & Adolescent Psychiatry	17	1,465	—	Children (2-10 years); Adolescents (11-20 years)	Anxiety; depression	Yes

Citation	Journal	*k* Studies	*n* Children	*n* Parents	Age Group(s)	Child outcome(s)	Intervention Focus
England-Mason et al. (2023)	Clinical Psychology Review	26	2,342	—	Infants (0-24 months); Children (2-10 years)	Internalizing; externalizing	Yes
Ezquerra et al. (2024)	European Child & Adolescent Psychiatry	14	4,286	—	Adolescents (11-20 years)	Suicidality	No
Farkas et al. (2024)	European Child & Adolescent Psychiatry	97	439,818	—	Adolescents (11-20 years)	Non-suicidal self-injury	No
Finlay-Jones et al. (2024)	Journal of the American Academy of Child and Adolescent Psychiatry	98	932,229	—	Infants (0-24 months); Children (2-10 years)	Internalizing; externalizing	No
Francis et al. (2023)	Journal of Child Psychology and Psychiatry	39	32,952	—	Adolescents (11-20 years)	Adverse childhood experiences	No
Frisira et al. (2025)	European Child & Adolescent Psychiatry	32	27,017,971	—	Children (2-10 years); Adolescents (11-20 years)	ADHD; ASD	No

Citation	Journal	*k* Studies	*n* Children	*n* Parents	Age Group(s)	Child outcome(s)	Intervention Focus
Gu et al. (2025)	European Child & Adolescent Psychiatry	31	25,258	—	Children (2-10 years); Adolescents (11-20 years)	ADHD	Yes
Guzman-Holst et al. (2022)	Journal of Child Psychology and Psychiatry	22	58,091	—	Children (2-10 years); Adolescents (11-20 years)	Internalizing	Yes
Haza et al. (2024)	Journal of Child Psychology and Psychiatry	142	16,283	—	Children (2-10 years); Adolescents (11-20 years)	ADHD	Yes
Holzman et al. (2022)	Clinical Psychology Review	38	3,488	—	Children (2-10 years); Adolescents (11-20 years)	PTSD	No
Jensen and Thomsen (2024)	Child Development	88	NR	—	Children (2-10 years); Adolescents (11-20 years)	Internalizing; externalizing	No
Joseph et al. (2023)	Journal of Child Psychology and Psychiatry	48	112,716	—	Infants (0-24 months); Children (2-10 years);	ADHD	No

Citation	Journal	*k* Studies	*n* Children	*n* Parents	Age Group(s)	Child outcome(s)	Intervention Focus
					Adolescents (11-20 years)		
Khazdouz et al. (2025)	European Child & Adolescent Psychiatry	10	58,296	—	Children (2-10 years); Adolescents (11-20 years)	ADHD	No
Kim et al. (2024)	European Child & Adolescent Psychiatry	22	2,655	—	Infants (0-24 months); Children (2-10 years); Adolescents (11-20 years)	Externalizing; ASD	No
Kopp et al. (2024)	European Child & Adolescent Psychiatry	46	NR	—	Infants (0-24 months); Children (2-10 years)	Internalizing; externalizing	No
Lång et al. (2022)	Journal of the American Academy of Child and Adolescent Psychiatry	16	436	—	Infants (0-24 months); Children (2-10 years); Adolescents (11-20 years)	Psychosis	No
Lee et al. (2025)	Development and Psychopathology	127	163,767	—	Infants (0-24 months); Children (2-10	General mental health and psychopathology	No

Citation	Journal	*k* Studies	*n* Children	*n* Parents	Age Group(s)	Child outcome(s)	Intervention Focus
					years); Adolescents (11-20 years)		
Li et al. (2024)	European Child & Adolescent Psychiatry	42	266,888	—	Children (2-10 years); Adolescents (11-20 years)	Depression; suicidality; non-suicidal self-injury	No
Lin et al. (2024)	Journal of Child Psychology and Psychiatry	49	25,524	—	Children (2-10 years); Adolescents (11-20 years)	Emotional dysregulation; internalizing	No
Loheide-Niesmann et al. (2024)	European Child & Adolescent Psychiatry	42	NR	—	Children (2-10 years); Adolescents (11-20 years)	Externalizing	No
Maciel et al. (2023)	Clinical Psychology Review	20	NR	—	Children (2-10 years); Adolescents (11-20 years)	Internalizing; externalizing	Yes
Manuele et al. (2023)	Clinical Psychology Review	47	NR	—	Children (2-10 years); Adolescents (11-20 years)	Internalizing; anxiety; depression	No

Citation	Journal	*k* Studies	*n* Children	*n* Parents	Age Group(s)	Child outcome(s)	Intervention Focus
McKern et al. (2023)	Psychological Bulletin	31	1,561	—	Children (2-10 years); Adolescents (11-20 years)	ASD	No
Moltrecht et al. (2024)	European Child & Adolescent Psychiatry	66	NR	—	Children (2-10 years); Adolescents (11-20 years)	Internalizing; externalizing	Yes
Newlove-Delgado et al. (2023)	Journal of Child Psychology and Psychiatry	51	NR	—	Children (2-10 years); Adolescents (11-20 years)	General mental health and psychopathology	No
Nieto-Retuerto et al. (2024)	Development and Psychopathology	42	NR	—	Infants (0-24 months); Children (2-10 years); Adolescents (11-20 years)	Externalizing	No
O'Halloran et al. (2022)	Clinical Psychology Review	47	37,879	—	Adolescents (11-20 years)	Suicidality; ASD	No
Perlstein et al. (2023)	Journal of Child Psychology and Psychiatry	60	9,405	—	Children (2-10 years);	Disruptive behavior	Yes

Citation	Journal	*k* Studies	*n* Children	*n* Parents	Age Group(s)	Child outcome(s)	Intervention Focus
					Adolescents (11-20 years)		
Restoy et al. (2024)	Clinical Psychology Review	55	NR	—	Children (2-10 years); Adolescents (11-20 years)	Emotional dysregulation; ASD	Yes
Robberegt et al. (2023)	Journal of the American Academy of Child and Adolescent Psychiatry	18	582	—	Adolescents (11-20 years)	Depression, anxiety	Yes
Sacco et al. (2024)	European Child & Adolescent Psychiatry	17	50,605	—	Children (2-10 years); Adolescents (11-20 years)	Anxiety; depression; ADHD; conduct problems; oppositional defiant disorder; ASD; eating disorders; substance use	No
Shen et al. (2023)	Clinical Psychology Review	37	NR	—	Children (2-10 years); Adolescents (11-20 years)	Internalizing; externalizing	Yes
Shephard et al. (2022)	Journal of the American Academy of Child	32	3,848	—	Infants (0-24 months); Children (2-10 years)	ADHD	Yes

Citation	Journal	*k* Studies	*n* Children	*n* Parents	Age Group(s)	Child outcome(s)	Intervention Focus
	and Adolescent Psychiatry						
Siciliano et al. (2022)	Clinical Psychology Review	44	NR	—	Children (2-10 years); Adolescents (11-20 years)	Internalizing; externalizing; general mental health and psychopathology	No
Somers et al. (2024)	European Child & Adolescent Psychiatry	28	1,931	—	Children (2-10 years); Adolescents (11-20 years)	PTSD	Yes
Tang et al. (2024)	Clinical Psychology Review	22	10,682	—	Adolescents (11-20 years)	Depression; anxiety	No
Tindall et al. (2024)	European Child & Adolescent Psychiatry	24	NR	—	Children (2-10 years); Adolescents (11-20 years)	Depression	Yes
Trompeter et al. (2024)	European Child & Adolescent Psychiatry	54	NR	—	Adolescents (11-20 years)	Anxiety symptoms; eating disorders	No
Tung et al. (2024)	Psychological Bulletin	55	NR	—	Children (2-10 years); Adolescents (11-20 years)	Externalizing	No

Citation	Journal	*k* Studies	*n* Children	*n* Parents	Age Group(s)	Child outcome(s)	Intervention Focus
Van IJzendoorn et al. (2023)	Child Development	25	NR	—	Unclear	Externalizing	Yes
Van Meter et al. (2023)	Journal of the American Academy of Child and Adolescent Psychiatry	365	3,136,151	—	Children (2-10 years); Adolescents (11-20 years)	Suicidality	No
Venturo-Conerly et al. (2023)	Journal of the American Academy of Child and Adolescent Psychiatry	34	NR	—	Children (2-10 years); Adolescents (11-20 years)	Internalizing; externalizing; anxiety; depression; conduct problems	Yes
Walsh et al. (2022)	Journal of Child Psychology and Psychiatry	12	33,155	—	Adolescents (11-20 years)	Suicidality	Yes
Wang et al. (2022)	Psychological Bulletin	114	9,364	—	Infants (0-24 months); Children (2-10 years); Adolescents (11-20 years)	ASD	No

Citation	Journal	*k* Studies	*n* Children	*n* Parents	Age Group(s)	Child outcome(s)	Intervention Focus
Wang et al. (2024)	Journal of Child Psychology and Psychiatry	38	344,860	—	Children (2-10 years); Adolescents (11-20 years)	Internalizing; externalizing	No
Wendel et al. (2024)	European Child & Adolescent Psychiatry	13	NR	—	Children (2-10 years); Adolescents (11-20 years)	Internalizing; externalizing	Yes
Wergeland et al. (2025)	European Child & Adolescent Psychiatry	44	3,251	—	Adolescents (11-20 years)	Eating disorders	Yes
Woolgar et al. (2022)	Journal of the American Academy of Child and Adolescent Psychiatry	18	1,941	—	Children (2-10 years)	PTSD	No
Yang et al. (2022)	Clinical Psychology Review	167	66,119	—	Children (2-10 years); Adolescents (11-20 years)	Internalizing; externalizing	No
Zhang and Ji (2024)	Development and Psychopathology	52	38,561	—	Adolescents (11-20 years)	Depression; anxiety; internalizing	No

Note. ADHD = Attention deficit hyperactivity disorder; ASD = Autism spectrum disorder; PTSD = Post-traumatic stress disorder, NR = Not reported.

Index of Tables

Index of Figures