



OPEN ACCESS

EDITED BY

Sara Uccella,
University of Genoa, Italy

REVIEWED BY

María Magán-Maganto,
Universidad de Salamanca, Spain
Gabriel Gonçalves Da Costa,
Federal University of Rio de Janeiro, Brazil

*CORRESPONDENCE

Magdalena Siegel
✉ magdalena.siegel@univie.ac.at

RECEIVED 27 November 2024

ACCEPTED 09 July 2025

PUBLISHED 28 July 2025

CITATION

Siegel M, Fanninger S, Riedel J and
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.
Front. Psychol. 16:1535606.
doi: 10.3389/fpsyg.2025.1535606

COPYRIGHT

© 2025 Siegel, Fanninger, Riedel and Zemp.
This is an open-access article distributed
under the terms of the [Creative Commons
Attribution License \(CC BY\)](#). The use,
distribution or reproduction in other forums is
permitted, provided the original author(s) and
the copyright owner(s) are credited and that
the original publication in this journal is cited,
in accordance with accepted academic
practice. No use, distribution or reproduction
is permitted which does not comply with
these terms.

Transparency, quality, and statistical consistency of meta-analytic systematic reviews in clinical child and adolescent psychology (2022–2024): study protocol for a meta-review

Magdalena Siegel*, Selina Fanninger, Julia Riedel and
Martina Zemp

Department of Clinical and Health Psychology, University of Vienna, Vienna, Austria

Meta-analytic systematic reviews are crucial for advancing research and practice in Clinical Child and Adolescent Psychology (CCAP). Despite their importance, there has been no systematic investigation into transparency- and quality-related aspects of these reviews in leading CCAP journals. This study protocol (<https://osf.io/qhrau/>) proposes a meta-review to assess the transparency, methodological quality, and statistical consistency of recent meta-analytic systematic reviews (2022–2024) published in leading journals from CCAP, aiming to improve future practices in the field. We will include meta-analytic systematic reviews from seven leading journals publishing CCAP-related content between 2022 and 2024 (estimated sample size based on piloting = 60). Eligible systematic reviews need to have conducted a frequentist meta-analysis, define eligible populations as children or adolescents between 0 and 20 years (ideally based on primary study sample mean), may include a clinical psychological or psychotherapeutic intervention, and need to focus on clinical psychological outcomes (no comparators defined). We will search *Web of Science (Core Collection)* by combining journal names (fully indexed within this database) and systematic review-related keywords. Eligible meta-analytic systematic reviews will be assessed for transparency (PRISMA-adaptation; newly developed set of items for CCAP-related content), methodological quality (AMSTAR 2), and statistical consistency (statcheck). Descriptive analyses will include overall and domain-based scores, as well as exploratory analyses assessing associations with transparency-promoting factors on review and journal level. This meta-review can shed light on and enhance the transparency, quality, and statistical consistency within meta-analytic systematic reviews from the field of CCAP. In doing so, it may provide guidance for researchers, reviewers, and editors, while laying the groundwork for future meta-studies in this field.

KEYWORDS

metascience, research synthesis, clinical child and adolescent psychology, reproducibility, meta-review

1 Introduction

Meta-analytic systematic reviews are paramount to research and practice within the field of Clinical Child and Adolescent Psychology (CCAP; Bellato et al., 2023; Weisz et al., 2023). Commonly placed at the top of the “evidence pyramid” (Murad et al., 2016), they ideally synthesize evidence for researchers, guide practitioners, and inform policy stakeholders through a systematic collation, critical appraisal, and statistical summary of relevant evidence (Cooper et al., 2019; Siddaway et al., 2019; Borenstein et al., 2021). CCAP is concerned with critical developmental periods—namely infancy, childhood, and adolescence—for the (early) onset (Solmi et al., 2022), prognosis, and treatment (Dekkers et al., 2023; Weisz et al., 2023) of mental disorders. Thus, syntheses within this field are not only highly relevant for public health policy, but ultimately also for children and adolescents affected by mental health burdens as well as their caregivers.

The trust placed in these syntheses by various stakeholders along with scientific developments, such as growing primary evidence, novel meta-analytic methods necessitating data re-analysis, and the growing popularity of meta-meta-analyses (Lakens et al., 2017; Polanin et al., 2020; Nguyen et al., 2022), highlight the need for transparent and methodologically sound meta-analytic systematic reviews. This is particularly relevant for reviews published in prestigious, widely disseminated journal outlets, as these are widely read (and cited) by practitioners and researchers. However, a systematic investigation of transparency and methodological quality-related aspects of meta-analytic systematic reviews within leading CCAP journals is currently lacking. Such an undertaking can best be realized through a meta-scientific approach. Meta-science (or “science of science”) examines the practice of science itself in order to analyze and optimize its processes and methods (Ioannidis et al., 2015; Ioannidis, 2018). Meta-reviews, i.e., methodological overviews of reviews (Biondi-Zoccai, 2016; López-Nicolás et al., 2022) are an important tool in this respect, as they can be used to evaluate the transparency, reproducibility, and quality of systematic reviews in a particular field.

This study protocol outlines our plans for a meta-review to examine the transparency, methodological quality, and consistency of statistical reporting in recent meta-analytic systematic reviews from seven leading journals from the field of CCAP. As a first meta-scientific venture into this field, we aim to provide an impression of current practices (and shortcomings) within these domains, and provide guidance for authors, reviewers, and editors.

1.1 Transparency of meta-analytic systematic reviews

A first key factor in assessing the trustworthiness of meta-analytic systematic reviews in CCAP is their transparency. Concerns about transparent reporting and ultimately reproducibility (i.e., the same analysis based on the same data yields the same results; Epskamp, 2019) and replicability (i.e., results at a later time support these conclusions; Epskamp, 2019) have accompanied the field of research synthesis from its inception (e.g., Mulrow, 1987). In part, these concerns sparked the development of the currently ubiquitously used (Caulley et al., 2020) and recommended (Kolaski et al., 2023) PRISMA guidelines (Page et al., 2021a) and their extensions that cover state-of-the-art reporting of protocols (Moher et al., 2015) or

search strategies (Rethlefsen et al., 2021) among others. In their current version (March 2021; Page et al., 2021a) the PRISMA guidelines provide detailed guidance on transparent reporting of (meta-analytic) systematic reviews, including on the structure and content of the background information and study aims; the eligibility criteria; systematic search, screening and coding procedures; bias and certainty assessments; meta-analytic procedures; the structure and content of the discussion; conflict of interest statements; protocol preregistration; public availability of materials such as data and code; and using automation tools during the systematic review process (Page et al., 2021a, 2021b). Importantly, PRISMA does not advise on methodological quality (Kolaski et al., 2023). For example, PRISMA recommends reporting the number of authors involved in study screening or coding (Page et al., 2021a), but it does not recommend double coding or interrater reliability thresholds, which indicate sufficient methodological quality (or conduct) within dedicated tools (Shea et al., 2017).

In Psychology and related biopsychosocial disciplines, the ongoing replication crisis (Open Science Collaboration, 2015; Wiggins and Christopherson, 2019) has also increased attention to the transparency and reproducibility of meta-analytic systematic reviews. This has culminated in several meta-scientific investigations within Psychology (Lakens et al., 2017; Hohn et al., 2020; Maassen et al., 2020; Polanin et al., 2020; Steil et al., 2022), Clinical Psychology (López-Nicolás et al., 2022), forensic science (Chin et al., 2022), cognitive training (Sandoval-Lentisco et al., 2024), or biopsychosocial interventions (Page et al., 2018, 2021c; Nguyen et al., 2022) among others. These studies, typically focusing on transparency-related aspects in line with PRISMA, show a relatively consistent pattern: Broad aspects of transparency and those related to narrative sections were most commonly reported, including the background information (Steil et al., 2022), eligibility criteria (Hohn et al., 2020; Polanin et al., 2020; López-Nicolás et al., 2022; Nguyen et al., 2022; Steil et al., 2022; Sandoval-Lentisco et al., 2024), databases searched (Chin et al., 2022; López-Nicolás et al., 2022; Steil et al., 2022; Sandoval-Lentisco et al., 2024) and number of hits retrieved (Nguyen et al., 2022), or the meta-analytic model used (e.g., fixed-effect or random-effects model; Polanin et al., 2020; López-Nicolás et al., 2022; Nguyen et al., 2022; Sandoval-Lentisco et al., 2024). However, detailed information required for reproducibility were often missing, including full and replicable search strategies for all databases searched (López-Nicolás et al., 2022; Nguyen et al., 2022; Steil et al., 2022; Rethlefsen et al., 2024), specific screening and data extraction procedures (Polanin et al., 2020) (but see Sandoval-Lentisco et al., 2024), meta-analytic weighting procedures and estimators of variance components (López-Nicolás et al., 2022; Nguyen et al., 2022; Sandoval-Lentisco et al., 2024), or formulae, information, and raw data to calculate effect sizes (Lakens et al., 2017; Maassen et al., 2020; Polanin et al., 2020; López-Nicolás et al., 2022). Regarding open science practices, the availability of preregistrations or study protocols was rare (Tsujimoto et al., 2017; Polanin et al., 2020; López-Nicolás et al., 2022; Nguyen et al., 2022; Sandoval-Lentisco et al., 2024) and, if existent, deviations from them were often not declared (Tricco et al., 2016; Koensgen et al., 2019; Hu et al., 2021; Sandoval-Lentisco et al., 2025). In a related vein, the sharing of data in interoperable format as well as statistical syntax to reproduce analyses was virtually absent (Polanin et al., 2020; Chin et al., 2022; López-Nicolás et al., 2022; Nguyen et al., 2022; Sandoval-Lentisco et al., 2024).

Currently, it is unclear to what extent these findings also apply to meta-analytic systematic reviews within the field of CCAP. The closest investigation is a meta-review on transparency and reproducibility within meta-analytic systematic reviews of clinical psychological interventions as cited above (López-Nicolás et al., 2022), which concluded that the adherence to these practices was somewhat higher than in other fields (López-Nicolás et al., 2022). Supporting this, three meta-meta-analyses in CCAP, focusing on externalizing behavior problems (Mingebach et al., 2018; Weber et al., 2019) and ADHD (Türk et al., 2023) used a now outdated version of PRISMA guidelines (Moher et al., 2009) as a proxy for study quality (not transparency). They concluded that the adherence to these guidelines was generally adequate (Mingebach et al., 2018; Weber et al., 2019; Türk et al., 2023). However, their scoring approach prevented identifying areas of high or low transparency, and current PRISMA guidelines would allow for a closer examination of finer-grained transparency practices as mentioned above.

Additional transparency considerations may be needed for meta-analytic systematic reviews in CCAP. This notion has been put forward for research syntheses involving pediatric populations, leading to the planned (Kapadia et al., 2016) but, as of November 2024, not yet published PRISMA guideline for such syntheses (PRISMA-C). This guideline is primarily geared toward systematic reviews for pediatric medical intervention research. However, preparatory works related to the development of this guideline (Kapadia et al., 2016; Farid-Kapadia et al., 2017a, 2017b) suggest several reporting domains relevant to CCAP: These include explicitly mentioning the pediatric population in the title, referencing (and justifying) this population within the study objectives, providing replicable age-related eligibility criteria, the reporting of age-specific search terms or tested pediatric search filters, and addressing the implications for children, adolescents, and caregivers specifically within the discussion section (Kapadia et al., 2016; Farid-Kapadia et al., 2017a, 2017b). Given these CCAP-specific considerations, it is crucial to assess their current uptake within meta-analytic systematic reviews in leading journals in the field.

1.2 Methodological quality of meta-analytic systematic reviews

A second key factor in assessing the trustworthiness of meta-analytic systematic reviews in CCAP, besides transparency, is their methodological quality (Kolaski et al., 2023). Its assessment (but not the quality itself) necessarily hinges upon transparent and reproducible reporting, but they are not equivalent, as described above. AMSTAR 2 (Shea et al., 2017), the recommended tool within psychology and health-related research (Kolaski et al., 2023), assesses methodological quality within several domains. These include providing a preregistered protocol; the adequacy of the eligibility criteria, literature search, data extraction, and meta-analytic procedures; a duplicate study selection and data extraction procedure; justifying the exclusion of individual studies; conducting risk and publication bias assessments; using appropriate meta-analytic methods, and reporting on potential conflicts of interest within primary studies and the review itself (Shea et al., 2017).

Contrasting the lack of evidence on transparency, the methodological quality of (meta-analytic) systematic reviews in CCAP has been assessed more frequently, likely due to the increased

popularity of overviews of reviews in psychology and health-related research that require a quality assessment (Gates et al., 2022). These quality-related assessments as part of overviews of reviews have examined diverse areas of CCAP, such as treatments for anxiety (Byrne et al., 2023; Galán-Luque et al., 2023; Zbukvic et al., 2024), depression (Espada et al., 2023; Zbukvic et al., 2024), eating disorders (García-Fernández et al., 2023), conduct problems (Romero et al., 2023), or consequences of sexual abuse (Sánchez De Ribera et al., 2020); mental health correlates of cyberbullying (Kwan et al., 2020); mental health burdens among sexual and gender minority youth (O'Shea et al., 2024); or the impact of paternal mental health on child development (Scarlett et al., 2024).

Except for one study (Zbukvic et al., 2024), these investigations documented poor methodological quality within most (Kwan et al., 2020; Sánchez De Ribera et al., 2020; Byrne et al., 2023; Romero et al., 2023) or all (Espada et al., 2023; Galán-Luque et al., 2023; García-Fernández et al., 2023; Scarlett et al., 2024; O'Shea et al., 2024) included systematic reviews (notably not all of them meta-analytic). Domains of consistently low quality included the lack of preregistration (Kwan et al., 2020; Sánchez De Ribera et al., 2020; Byrne et al., 2023; Espada et al., 2023; Galán-Luque et al., 2023; García-Fernández et al., 2023; Romero et al., 2023; Scarlett et al., 2024), failure to justify excluding individual studies (Kwan et al., 2020; Sánchez De Ribera et al., 2020; Byrne et al., 2023; Espada et al., 2023; Galán-Luque et al., 2023; García-Fernández et al., 2023; Romero et al., 2023; Scarlett et al., 2024), and no assessment of risk of bias (Kwan et al., 2020; Sánchez De Ribera et al., 2020; Byrne et al., 2023; Galán-Luque et al., 2023; Romero et al., 2023), publication bias (Sánchez De Ribera et al., 2020; Byrne et al., 2023), or primary studies' funding sources (Kwan et al., 2020; Espada et al., 2023; Romero et al., 2023; Scarlett et al., 2024). These investigations paint a rather bleak picture of the methodological quality of systematic reviews within certain domains and areas of CCAP. They also highlight the need for larger-scaled investigations into this matter (most overviews included fewer than 15 systematic reviews) and clearer guidance on improving methodological quality.

1.3 Statistical consistency of meta-analytic systematic reviews

A third key factor in assessing the trustworthiness of meta-analytic systematic reviews is the consistency of statistical reporting within null hypothesis significance testing (NHST; Sterling, 1959). Typically, this involves checking whether reported *p*-values are consistent with their accompanying test statistics and degrees of freedom (Bakker and Wicherts, 2011; Nuijten et al., 2016). Such inconsistencies can indicate various underlying issues, ranging from simple mistakes over misunderstandings of statistics to engaging in questionable research practices, such as incorrectly and deliberately rounding *p*-values below 0.05 (John et al., 2012; Nuijten et al., 2016). Regardless of their cause (and of discussions about whether to abandon NHST), such reporting inconsistencies can lead to false conclusions about the degree to which the data support the presence or absence of an effect (e.g., an effect is erroneously deemed to be significant), hinder reproducibility, and preclude calculating correct effect sizes based on test statistics (Nuijten et al., 2016; Polanin and Nuijten, 2018; Nuijten and Polanin, 2020).

Research has mainly focused on statistical reporting inconsistencies within primary studies (e.g., Bakker and Wicherts, 2011; Nuijten et al., 2016; Olsen et al., 2019; Nuijten and Wicherts, 2024), indicating a disconcertingly high prevalence of statistical reporting errors. For example, a re-analysis of over 250,000 *p*-values from eight major psychology journals (1985–2013) using the R package {statcheck} (Nuijten and Epskamp, 2024) revealed at least one inconsistency between test statistics, degrees of freedom, and *p*-values in about 50% of publications as well as at least one gross inconsistency (i.e., altering conclusions based on statistical significance) in about 13% of publications (Nuijten et al., 2016). Recently, statcheck has been adapted to include test statistics used in meta-analyses (e.g., the *Q* statistic), enabling investigations of meta-analytic systematic reviews (Nuijten and Polanin, 2020).

Some might argue that the primary purpose of a meta-analysis is to provide an estimate of a meta-analytic summary effect (or an exploration of the heterogeneity around this effect), and hence, NHST is of little interest within meta-analytic systematic reviews, including those within CCAP. However, NHSTs in meta-analytic studies are common (previous works documented averages between 29 and 68 NHSTs per article; Polanin and Pigott, 2015; Polanin and Nuijten, 2018) and, for example, routinely used to assess whether the data support the presence or absence of moderator effects, statistical heterogeneity, or, in small samples, even meta-analytic summary effects (Polanin and Pigott, 2015; Nuijten and Polanin, 2020). Furthermore, arguments about inconsistent statistical reporting as litmus tests for deeper underlying issues within a study (Nuijten et al., 2016), as outlined above, similarly apply to meta-analytic investigations (Nuijten and Polanin, 2020). Meta-scientific studies have revealed that 39% of meta-analyses within the social sciences included at least one inconsistent result, whereby 8% contained a gross inconsistency (Polanin and Nuijten, 2018; Nuijten and Polanin, 2020). However, no systematic investigation has yet focused on meta-analytic systematic reviews in CCAP.

1.4 The current study

Transparency, quality, and statistical consistency of reporting are crucial for the trust that researchers, clinicians, policymakers, and other stakeholders within the field of CCAP place in meta-analytic systematic reviews. Further, in seeking guidance for future studies, applied meta-analysts within the field of CCAP may want to know in which areas the reporting transparency or quality could be improved to guide their own efforts.

Within this meta-review we aim to answer three research questions (RQs) that all refer to meta-analytic systematic reviews from the field of CCAP within leading journals of this discipline published between 2022 and 2024.

RQ 1: How transparent is the reporting of current meta-analytic systematic reviews published within leading journals from the field of CCAP (overall and within broad domains, such as eligibility criteria, systematic search and screening procedure, or effect sizes and statistical synthesis)?

RQ 2: How high is the methodological quality of current meta-analytic systematic reviews published within leading journals from the field of CCAP (overall and within domains)?

RQ 3: How consistent is the reporting of statistical significance testing in current meta-analytic systematic reviews published within leading journals from the field of CCAP?

As a secondary aim, we plan to gain further insights into transparency-promoting factors by exploring associations between reporting transparency as well as methodological quality, statistical consistency, and other transparency-related variables within our review sample, as meaningful associations between these or related factors have been documented within the literature (Page et al., 2016, 2018; Chin et al., 2022; López-Nicolás et al., 2022; Nguyen et al., 2022). We plan to assess associations between:

- Reporting transparency and methodological quality
- Reporting transparency and statistical consistency
- Reporting transparency and journals' transparency and openness promotion factor (TOP; Nosek et al., 2015)
- Reporting transparency and self-reported adherence to PRISMA guidelines (yes vs. no)
- Reporting transparency and page limit of journal (no or long page limit vs. short page limit)

As a tertiary aim, we plan to provide practical recommendations for researchers within CCAP by highlighting “what works” (i.e., areas where reporting transparency and methodological quality are high) and “what could work better” (i.e., areas where more transparent reporting or higher quality is advisable for the field), which we will incorporate into our discussion section.

2 Methods and analysis

See Figure 1 for a complete flowchart of this meta-review.

2.1 Open science practices, guideline adherence, and transparency statement regarding generative AI

All study and Supplementary materials (S1–S5) for this study protocol can be found in Supplementary material as well as via <https://osf.io/qhrau/> (including preprint publication). We will make all data (in machine-readable format) and all materials (including R code to reproduce all analyses, figures, and tables) for the final publication available via this project on OSF, alongside a preprint publication on PsyArXiv. Regarding guideline adherence, there are currently no specific guidelines available for conducting meta-reviews (Dal Santo et al., 2023). Thus, we adhered to the PRIOR statement for overviews of reviews (Gates et al., 2022) together with the PRISMA (Page et al., 2021a), PRISMA-P (Moher et al., 2015) (checklist provided within Supplementary material S1) and PRISMA-S guidelines where applicable (Rethlefsen et al., 2021). The final publication will similarly adhere to these guidelines where applicable. Note that, in contrast to PRIOR, we did not include “overview of reviews” in our title (Gates et al., 2022), as this would imply a content focus rather than a methodological focus inherent to meta-reviews by our definition (Biondi-Zoccai, 2016; López-Nicolás et al., 2022).

During preparation of this manuscript, ChatGPT (versions 3.5 and 4.0; OpenAI, 2024) was used on author-generated text [MS] within narrative sections (Introduction, Discussion) to edit for

Meta-Review Flowchart

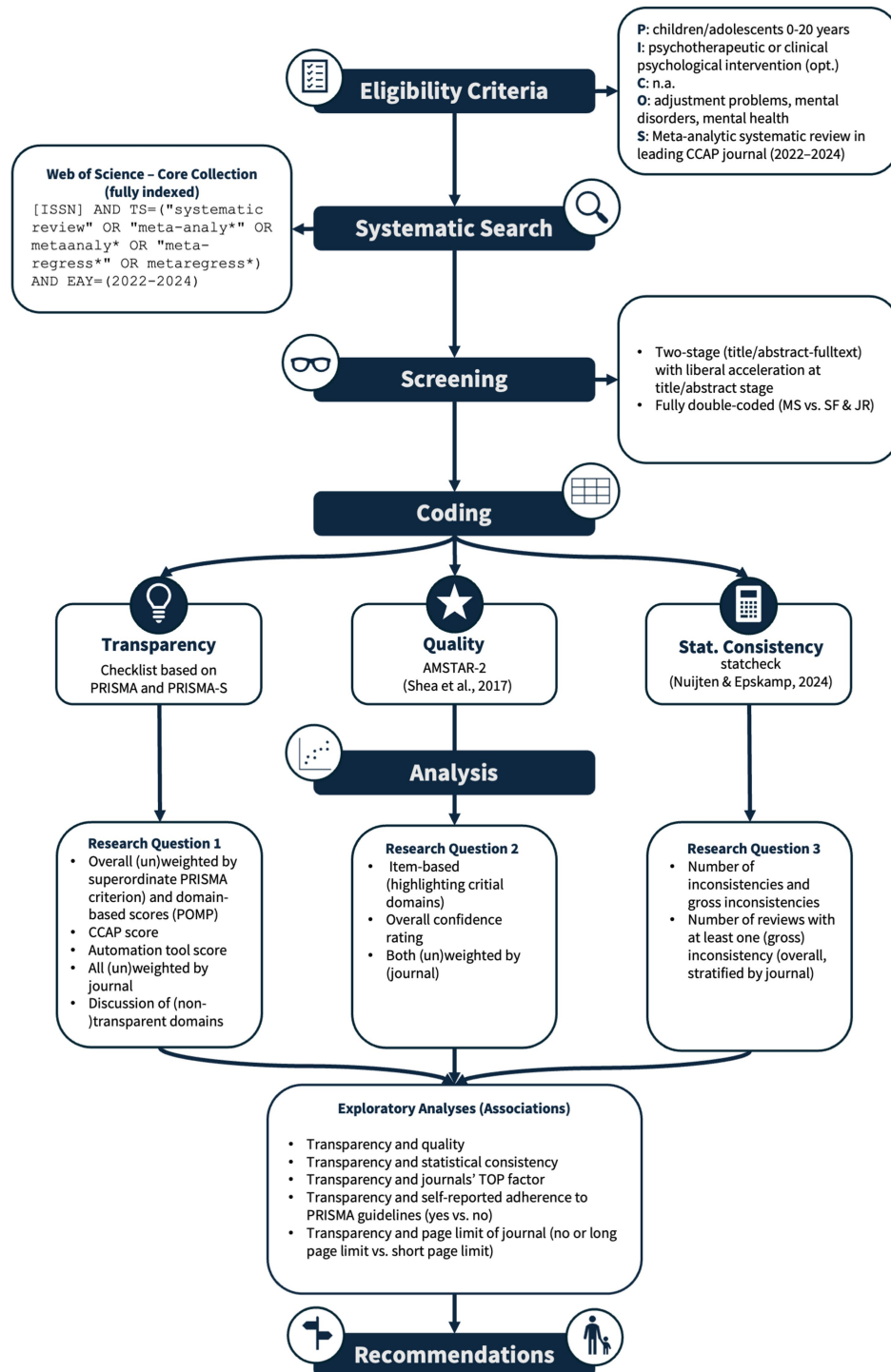


FIGURE 1
Flowchart of the meta-review.

typography, grammar, clarity, and brevity only, akin to a human co-author (Nakagawa and Lagisz, 2024). Prompts included varieties of the statement “Please edit the following paragraph for brevity and clarity and point out what was corrected: [author-generated text copied in full].” MS incorporated these suggestions as seen fit and, together with the other authors, takes full responsibility for the content of this manuscript. No new text segments, ideas, or references were generated with the help of ChatGPT, any other large language model, or any other generative artificial intelligence assistance.

2.2 Conceptual scope regarding the field of CCAP

This meta-review focuses on meta-analytic systematic reviews from the field of CCAP. To translate a definition of this field into precise eligibility criteria, we will first present our definition of CCAP for the purpose of this meta-review and highlight conceptual differences from related disciplines.

In line with the American Psychological Association (Division 53), we define CCAP as a subdiscipline of psychology that examines mental (emotional, behavioral, and developmental) disorders in children and adolescents, including research on the development, maintenance, diagnosis, and clinical-psychological as well as psychotherapeutic interventions/treatment of disorders with a focus on infants, children, and adolescents across all age groups (Brown et al., 2023). We consider developmental psychopathology to be a subarea of CCAP, as it deals with developmental processes that contribute to the etiology, onset, and chronification of mental disorders in children and adolescents, mostly focusing on examining risk and protective factors and trajectories of internalizing and externalizing problems throughout childhood and adolescence in the general population, i.e., including subclinical phenotypes (Barry et al., 2023).

The field of CCAP shares considerable conceptual overlap with Child and Adolescent Psychiatry as well as (Pediatric) Clinical Neuroscience, which are not within the scope of this meta-review. Regarding Child and Adolescent Psychiatry, we base our definition on American Academy of Child and Adolescent Psychiatry (2021) and Schlesinger et al. (2023), which defines it as a medical specialty area that primarily deals with medical aspects of mental disorders and drug/psychopharmacological therapies among children and adolescents. We define the field of (Pediatric) Clinical Neuroscience as an interdisciplinary field combining aspects of neuroscience and clinical medicine: It encompasses the study and treatment of diseases of the nervous system, focusing on the application of neurobiological findings to clinical problems (Kandel et al., 2021) with a typical focus on biological markers, such as genetic (e.g., DNA analyses), neuroimaging (e.g., fMRI), electrophysiological (e.g., EEG), biochemical (e.g., salivary cortisol), and cardiovascular (e.g., heart rate variability) markers.

2.3 Eligibility criteria

A concise summary of our eligibility criteria is reported in Table 1. We give additional clarifying information below. We [MS, MZ] conceptualized a first draft of our eligibility criteria based on

theoretical and methodological considerations. As recommended (Foo et al., 2021), we then refined these criteria through piloting the screening of a random sample of 55 records (10%) retrieved by our search strategy (see below), as well as through full-text assessments of published meta-analytic systematic reviews from the field of CCAP (Alozkan-Sever et al., 2023; Davis et al., 2023; Manuele et al., 2023; Harris et al., 2024; Lin et al., 2024; Tung et al., 2024; note that these did not need to fulfill final eligibility criteria).

2.3.1 Publication type and language

We will include meta-analytic systematic reviews (see section ‘Study Type’ for our definition of a meta-analytic systematic review) published as a peer-reviewed article in one of the seven journals as outlined below (see section ‘Information Sources’) within the years 2022 to 2024. We consider the publication year as indexed within *Web of Science (Core Collection)*; category: early access). Thus, it will be possible that we include articles that were later assigned a volume and issue number in 2025, because their early access publication date was 2024. Note that upon immediate assignment of a volume and issue number, the early access date is identical to this publication date within *Web of Science*. Articles will be excluded if they were received as a first submission at the journal prior to March, 2021, the month were the updated PRISMA guidelines were published, because we deem it unlikely that authors were asked to “update” their guideline adherence from PRISMA 2009 (Liberati et al., 2009) to PRISMA 2020 (Page et al., 2021a) during the review process. From our journal-based inclusion criterion it follows that the publication language will be English (all included journals publish English-language articles exclusively). We will exclude abstracts of conference proceedings published within these journals (e.g., abstracts of the Annual Meeting of the American Academy of Child and Adolescent Psychiatry, as published within the respective journal).

2.3.2 Study type

We will include meta-analytic systematic reviews. Explicit definitions of systematic reviews in overviews of systematic reviews (including meta-reviews) are currently rare and, when reported, poorly described (Faggion and Diaz, 2019; Gates et al., 2022). In our definition of a systematic review, we follow the current PRISMA guidelines that define a systematic review based on the Cochrane Handbook (Higgins et al., 2024) as “a review that uses explicit, systematic methods to collate and synthesize findings of studies that address a clearly formulated research question” (Page et al., 2021a, p. 3). We translated this broad definition into the following explicit eligibility criteria (see Table 1): The publication must (i) present a clearly formulated research question or study aim, (ii) formulate (and follow) explicit eligibility criteria against which potentially eligible studies are assessed (regardless of whether a common scheme, such as PICO, was used), (iii) describe a systematic literature search approach, operationalized as searching at least one (named) database and providing at least examples of keywords used within this search, and (iv) describe a systematic approach to data extraction, with the minimum standard of enumerating or indicating some (key) variables that were extracted.

Systematic reviews that rely on a non-traditional, but still systematic literature search approach (e.g., a forward search of a key reference instead of a keyword search) will be judged on a case-by-case basis within the research team. Decisions regarding eligibility will

TABLE 1 Eligibility criteria.

Criterion	Inclusion criteria	Notable exclusion criteria
Publication type and language	<ul style="list-style-type: none">Published within one of the following journals: <i>Child Development</i>, <i>Clinical Psychology Review</i>, <i>Development and Psychopathology</i>, <i>European Child & Adolescent Psychiatry</i>, <i>Journal of Child Psychology and Psychiatry</i>, <i>Journal of the American Academy of Child and Adolescent Psychiatry</i>, <i>Psychological Bulletin</i>Published between 2022 and 2024 (early access date if no volume and issue assigned upon first online publication)Published in English (consequence of journal selection)	<ul style="list-style-type: none">Conference abstracts published within these journalsReceived as first submission prior to March, 2021 due to possible adherence to PRISMA 2009
Study type	<ul style="list-style-type: none">Meta-analytic systematic review; i.e., including (i) a clearly formulated research question or aim, (ii) explicit eligibility criteria, (iii) a systematic literature search, (iv) a systematic approach to data extraction, (v) a formal frequentist meta-analysis of at least two effect sizes from two different primary studies.	<ul style="list-style-type: none">Systematic reviews evaluating diagnostic test accuracy or psychometric features of an assessment instrument, systematic reviews of qualitative studiesNarrative systematic reviews (i.e., without a meta-analysis)Network meta-analyses, Bayesian meta-analyses, individual participant meta-analyses, coordinate-based meta-analyses or other meta-analyses specific to the field of neuroimaging, psychometric meta-analyses, or machine learning-based meta-analysesOther forms of research syntheses, including scoping reviews, rapid reviews, and overviews of reviews
Population	<ul style="list-style-type: none">Population of interest (specified within eligibility criteria of the meta-analytic systematic review) are children and/or adolescents between 0 and 20 years (ideally defined as primary study mean age <20 years, see text for fallback criteria)	<ul style="list-style-type: none">Adult populations providing retrospective accounts on relevant outcomes within childhood or adolescence.
Intervention	<ul style="list-style-type: none">For meta-analytic systematic reviews assessing intervention effects only: assessment of at least one clinical-psychological or psychotherapeutic intervention	<ul style="list-style-type: none">For meta-analytic systematic reviews assessing intervention effects only: exclusive assessments of non-psychological interventions, including medical and psychopharmacological
Comparator	–	–
Outcome	<ul style="list-style-type: none">Synthesized effect sizes include any kind of child adjustment problems in the domain of (i) externalizing (behavior) and internalizing (emotional) problems, (ii) other mental disorders, or (iii) any indicator of children and adolescents’ overall mental health	–

be transparently reported and systematically applied to all further, similar cases. We consciously kept these eligibility criteria as broad as possible to not systematically exclude systematic reviews of low quality. We will exclude the following study types: Scoping reviews (as defined within [Tricco et al., 2018](#)); overviews of (systematic) reviews ([Gates et al., 2022](#)), including studies labeled as umbrella reviews, meta-meta-analyses, meta-reviews, or second-order meta-analyses; and systematic reviews evaluating diagnostic test accuracy or psychometric features of an assessment instrument, the latter due to differing reporting guidelines.

Because our aim is to analyze the reporting characteristics of *meta-analytic* systematic reviews, the systematic review also (v) needs to report on a frequentist meta-analysis of effect sizes of at least two primary studies based on the systematic literature search. Here, we again follow the PRISMA guidelines and define 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). This broad definition of a meta-analysis translates into the following eligibility criterion: Use

of a fixed-effect or random-effects meta-analytic model within a frequentist framework (as outlined, for example, in [Borenstein et al., 2021](#)), where models with several variance components (to account for nesting of effect sizes within studies, most commonly a three-level meta-analysis, e.g., [Assink and Wibbelink, 2016](#); [Pustejovsky and Tipton, 2022](#)) will be included. We do not place any restriction on the type of meta-analytic effect size or whether moderator analyses were conducted. Following argumentation within previous meta-reviews with a similar aim ([López-Nicolás et al., 2022](#); [Sandoval-Lentisco et al., 2024](#)), we will exclude systematic reviews that apply network meta-analyses, Bayesian meta-analyses, individual participant meta-analyses, coordinate-based meta-analyses or other meta-analyses specific to the field of neuroimaging, psychometric meta-analyses, or machine learning-based meta-analyses because these types of meta-analyses require different criteria to assess their quality and transparency of reporting and/or they do not require reporting of statistical information as evaluated within statcheck ([Nuijten and Polanin, 2020](#); [Nuijten and Epskamp, 2024](#)), the tool we use to assess statistical reporting consistency.

2.3.3 Population

We will include meta-analytic systematic reviews that define their population of interest within their eligibility criteria as children and adolescents between 0 and 20 years, following [World Health Organization \(2024\)](#). Ideally, this would entail a definition of the mean sample age of <20 years for primary studies (included in the meta-analytic systematic review) as specified within the meta-analytic systematic reviews' eligibility criteria. If a more ambiguous criterion is reported within these meta-analytic systematic reviews, such as "(...) included at least one measure of offspring externalizing behavior assessed during childhood or adolescence [defined as ages 2–20]" ([Tung et al., 2024](#), p. 6; 18 as upper age limit replaced with 20 for the sake of the example given here), where it is unclear whether this refers to a measure of central tendency or range, we will include the meta-analytic systematic review in case of at least 90% of primary studies with a mean age <20 years. We will proceed the same way in cases where no clear age-related eligibility criteria are defined but it is evident that the meta-analytic systematic review focuses on children or adolescents. We will tally this manually from information within the meta-analytic systematic review if reported (without reviewing primary studies of the respective meta-analytic systematic review by ourselves). We will exclude meta-analytic systematic reviews based on retrospective accounts, i.e., adults reporting on eligible outcomes within childhood or adolescence. If this information (i.e., primary studies' mean sample age) is not reported within the meta-analytic systematic review, but the (ambiguous) age range is within 0–20 years, we will include the systematic review (without reviewing the primary studies). If the systematic review specifies the population of interest as "children or adolescents" without reporting on age-related criteria or the population of interest is ambiguously specified (e.g., "youth"), we will decide on inclusion on a case-by-case basis within the research team and report the decision for inclusion or exclusion transparently. The implemented approach will be systematically applied to all further, similar cases.

If the systematic review includes both children/adolescent and adult samples, we will include the systematic review in case of at least 90% of primary studies with a mean age <20 years. If this information is not reported, we will exclude the systematic review (without reviewing the primary studies). Note that this eligibility criterion does not imply that self-report data for this population need to be included within the meta-analytic systematic review (see section 'Outcomes').

2.3.4 Intervention

We will include both meta-analytic systematic reviews of studies without examining treatments/interventions (e.g., epidemiological and prevalence studies, studies focusing on developmental psychopathology) as well as intervention studies. For meta-analytic systematic reviews that do not synthesize information on intervention effects, this eligibility criterion does not apply. For meta-analytic systematic reviews that synthesize information on intervention effects, we will include those that assess the effect of at least one clinical-psychological or psychotherapeutic intervention within the conceptual scope of CCAP (see section 'Conceptual Scope Regarding the Field of CCAP'). We do not place any restriction on the recipient of the intervention (e.g., children, caregivers), but the outcome must be assessed for children or adolescents (through any form of rating or assessment, including, e.g., self-reports, other-reports, or clinical assessments). Systematic reviews that only assess effects of

non-psychological treatments (e.g., psychopharmacological or other medical interventions) will be excluded, as these would belong to the field of psychiatry or pediatrics. Meta-analytic systematic reviews that include both clinical-psychological or psychotherapeutic interventions together with psychopharmacological interventions will be included.

2.3.5 Comparator

No eligibility criteria defined.

2.3.6 Outcome

Eligible for inclusion are meta-analytic systematic reviews where synthesized effect sizes (and thus, at least one meta-analytic summary effect reported) are based on any kind of child adjustment problems in the domain of externalizing (behavior) and internalizing (emotional) problems, other mental disorders, or any indicator of children and adolescents' overall mental health. First, internalizing and externalizing problems are defined either as a (1) subclinical phenomenon (for internalizing problems: e.g., anxiety symptoms, depressive symptoms, or a sum score of internalizing problems; for externalizing problems: e.g., noncompliance, hyperactivity, aggressive and disruptive behaviors, delinquency, or a sum score of externalizing problems) or as a (2) clinical diagnosis (e.g., anxiety disorders, depressive disorders, conduct disorders [CD], oppositional defiant disorders [ODD], attention-deficit/hyperactivity disorders [ADHD]) based on the ICD-10 or ICD-11 ([World Health Organization, 1993, 2019](#)), or on the DSM-IV or DSM-5 criteria ([American Psychiatric Association, 1994, 2013](#)). We will also include outcomes indicating a combination of externalizing and internalizing problems.

Second, synthesized effect sizes related to symptoms or syndromes of other mental disorders included in the two latest revisions of the DSM and ICD (chapter F for ICD-10 or section 06 for ICD-11, respectively) that cannot be clearly categorized as internalizing and externalizing problems will be included (e.g., posttraumatic stress disorder [PTSD], autism spectrum disorders [ASD], obsessive-compulsive disorders [OCD], substance-related and addictive disorders, bipolar disorders, eating disorders, schizophrenia spectrum and other psychotic disorders). They could be assessed as clinical diagnoses based on the above-mentioned classification manuals or as continuous measures of selected symptoms related to these disorders (e.g., in disorder-specific questionnaires or scales).

Third, overall mental health includes any global indicators of psychological distress, level of functioning (global assessment or in different areas of life), perceived (chronic) stress, and indicators of a wide range of (non-specific) psychological symptoms (e.g., self-injurious behavior, suicidal ideation/thoughts, behaviors or attempts) and their severity. We will base our inclusion decision on (1) relevant descriptions within meta-analytic systematic reviews' eligibility criteria, or, in case it is not sufficiently specified, based on (2) descriptions of measures used within primary studies (if reported within the meta-analytic systematic review). If it is unclear based on (1) and (2), the systematic review will be excluded.

Meta-analytic systematic reviews exclusively synthesizing effect sizes based on child adjustment problems in other domains, such as i.e., academic functioning, physical health chronic diseases, coping skills, social support, life satisfaction, and overall or health-related quality of life will be excluded, as they do not fit into our specific outcome categories defined above. Similarly, as a consequence of our

considerations outlined above (see section ‘Conceptual Scope Regarding the Field of CCAP’), we will exclude systematic reviews where effect sizes are exclusively based on biological outcome markers, such as genetic markers (e.g., DNA analyses), neuroimaging markers (e.g., fMRI), electrophysiological markers (e.g., EEG), biochemical marker (e.g., salivary cortisol), and cardiovascular marker (e.g., heart rate variability).

2.4 Information sources

We focus on meta-analytic systematic reviews from leading journals (based on content and citation metrics criteria) within the field of CCAP to examine reporting practices of the (potentially) most impactful publications within this field. This may represent an upper threshold of the uptake of good reporting practices, but substantial reporting non-transparency has also been identified within top-ranking journals in Psychological Science (e.g., Polanin et al., 2020).

This is the first (notably unfunded) investigation into transparency, methodological quality, and statistical consistency of reporting within meta-analytic systematic reviews within CCAP. Thus, we deem this approach reasonable to gain a first impression of the field and deduce practical recommendations for researchers, as well as to lay the foundation for future works within this field (e.g., meta-reviews assessing time trends or prevalence estimates on reporting practices within the field). Since there is not one single, non-invitational flagship journal dedicated to publishing systematic reviews within CCAP that we could sample from (a commonly used approach within other psychological disciplines, e.g., Polanin et al., 2020; Siegel et al., 2022), and we deemed a keyword search as too broad (and therefore impractical, if not impossible) to identify eligible systematic reviews, we identified eligible journals through a three-step procedure based on expert opinion and journal metrics, as detailed in [Supplementary material S2](#).

Ultimately, this yielded a list of seven journals we considered to be representative of the top-ranking journals publishing meta-analytic systematic reviews within the field of CCAP (in alphabetical order): *Child Development*, *Clinical Psychology Review*, *Development and Psychopathology*, *European Child & Adolescent Psychiatry*, *Journal of Child Psychology and Psychiatry*, *Journal of the American Academy of Child and Adolescent Psychiatry*, and *Psychological Bulletin*. Together, these journals cover a broad spectrum of clinical topics and areas of application relevant to CCAP and encompass all age groups specified in our eligibility criteria, i.e., from infancy to adolescence/young adults. They publish widely disseminated research syntheses (and primary studies) from around the world using a diverse range of methods and are highly relevant to an international audience (with a broad ethnic background), both in terms of readership and authorship. All journals are indexed in *Web of Science (Core Collection)*, as subscribed by the University of Vienna). Thus, we rely on this database as our single information source.

2.5 Search strategy

Our full search string for *Web of Science (Core Collection)*, as subscribed by the University of Vienna) is reported in [Table 2](#). We combined (1) all seven unique journal ISSNs (see also [Chin et al.,](#)

[2022](#), for a similar approach and [Table 3](#) for journal ISSNs) with (2) keywords relating to systematic reviews or meta-analyses (appearing in titles, abstracts, or keywords). These records were then (3) limited to publication years (early access) 2022, 2023, or 2024. Within sets (1) and (2), search terms were combined using the OR-operator, whereas sets (1), (2), and (3) were combined using the AND-operator. No other filters or limits were used. The search string was constructed by the first author [MS] and peer-reviewed for accuracy by the second author [SF]. We did not validate the search string against a pre-specified set of studies, because the keywords used within our search were limited and eligible articles largely defined through their publishing outlets. On September 25, 2024, this search strategy yielded 550 records. After peer review for this protocol has been concluded, we will re-run this search. If peer review has been completed prior to 2025, we will update our search after January 1, 2025, for records from 2024.

To derive search terms for set (2), we screened previous meta-reviews ([Koffel and Rethlefsen, 2016](#); [Cortese et al., 2019](#); [Wayant et al., 2019](#); [Hohn et al., 2020](#); [Chow et al., 2021](#); [Chin et al., 2022](#); [López-Nicolás et al., 2022](#); [Nelson et al., 2022](#); [Nguyen et al., 2022](#); [Steil et al., 2022](#); [Wong and Bouchard, 2023](#); [Pellegrini et al., 2024](#); [Rethlefsen et al., 2024](#); [Sandoval-Lentisco et al., 2024](#)) for keywords and deemed variants of “systematic review” and “meta-analysis” useful for the purpose of this study. We supplemented these with variants of the term “meta-regression” to also include meta-analytic systematic reviews that primarily focused on exploring heterogeneity within effect sizes. We piloted including the term “quantitative review” (as used, for example, in [López-Nicolás et al., 2022](#)) as well as “research syntheses” (own considerations), which yielded no additional records and were dropped.

We deliberately chose to search within titles, abstracts, and keywords (vs. a title-only search) and to include variants of the keyword “meta-analysis” (vs. “systematic review” only) because we wanted to avoid limiting our search to those meta-analytic systematic reviews that fulfill the first PRISMA—and thus transparency—criterion, namely to include the term “systematic review” within the title ([Page et al., 2021a](#)). Our journal- and study type-based search strategy will also retrieve systematic reviews unrelated to children and adolescents (e.g., from general clinical psychology journals such as *Clinical Psychology Review*) and/or clinical psychology (e.g., from developmental journals such as

TABLE 2 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

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 September 25, 2024, 550 hits.

TABLE 3 Characteristics of included journals.

Journal	IF22	IF23	TOP	APA format	PRISMA adherence	Page limit	ISSN
Child Development	4.6	3.9	7	Yes	No	Long (50 pages)	0009-3920
Clinical Psychology Review	12.8	13.7	1	No	No	Long (50 pages)	0272-7358
Development and Psychopathology	3.3	3.1	0	Yes	Yes	None	0954-5794
European Child & Adolescent Psychiatry	6.4	6.0	4	No	Yes	Long (12,000 words)	1018-8827
Journal of Child Psychology and Psychiatry	6.5	7.6	5	Yes	Yes	Short (5,000 words)	0021-9630
Journal of the American Academy of Child and Adolescent Psychiatry	13.3	9.2	5	No	Yes	Short (5,000 words)	0890-8567
Psychological Bulletin	22.4	17.3	15	Yes	Yes	None	0033-2909

IF = Impact Factor (Clarivate Analytics). ISSN = International Standard Serial Number (Print). APA-formatting, PRISMA adherence, and page limit assessed through author guidelines from journal websites (August 26, 2024). TOP = Transparency and Openness Promotion Score, assessed through <https://topfactor.org/> (August 24, 2024). Page-word conversion as per standard APA formatting recommendation (12 pt, double-spaced) calculated by assuming 250 words/page.

Child Development or multidisciplinary journals such as the *Psychological Bulletin*). We will screen out these reviews manually (see below), a strategy that is feasible given the moderate number of records.

2.6 Selection process and expected sample size

We opt for a classic two-stage screening procedure (Page et al., 2021a), including screening of titles and abstracts, followed by full-text screening. We piloted this approach based on the same sample of 55 records that was also used to pilot our eligibility criteria. Title and abstract screening will be undertaken by three authors [MS, SF, JR], who will assess titles and abstracts of every record independently. Records will be advanced to the full-text stage via liberal acceleration (Page et al., 2018), i.e., every record deemed as eligible for full-text assessment by one (or more) author(s) will be assessed as a full-text. For descriptive purposes, we will report our interrater reliability (percentage agreement) between all three authors and between pairs of authors.

Full-text assessment will also be undertaken by [MS, SF, JR], in such that we will assess full-texts' final eligibility prior to data extraction. Final eligibility will be based on consensus among all three authors. The senior author [MZ] will arbitrate in cases where no consensus can be reached or additional expertise is needed. For descriptive purposes (our inclusion decision is consensus-based in all cases, Shea et al., 2017), we will report initial interrater reliability (percentage agreement between all three authors and pairs of authors). References of records that could not be retrieved in full or that were excluded at this stage will be published together with the reason for their exclusion. For inclusion decisions that we made on a case-by-case basis, we will provide a separate table (possibly within [Supplementary material](#)) including a detailed reasoning for each case. The number of excluded records by category will also be made available within the PRISMA flow chart.

For both steps, we will not use any automation tools or rely on crowdsourcing or previously “known” assessments (Page et al., 2021a). We do not expect needing to translate abstracts or articles because of our journal-based search strategy, including English-language articles only. We also do not expect having to contact authors for information necessary to determine eligibility. We will publish a PRISMA flow chart.

We estimate that our final sample of eligible meta-analytic systematic reviews will include about 60 reviews. We base this estimate on our piloting procedure, where we [MS, SF] piloted our eligibility criteria as well as our systematic search procedure, as recommended (Foo et al., 2021). Based on a random sample of 55 out of 550 retrieved records (10%; September 25, 2024), we deemed six as eligible for inclusion, thus totaling $6 * 10 = 60$ studies based on extrapolation.

2.7 Data collection process

Data collection for research questions (RQ 1; transparency) and (RQ 2; quality) will be undertaken by [MS, SF, JR]. We will randomly split eligible records into two groups. The first author [MS] will code all records once and the second authors [SF, JR] will each code records from one of the two groups. Thus, every eligible record will be coded twice and independently. We will resolve discrepancies through discussion and/or arbitration by another author (SF and JR for records they did not code; MZ for all records). We piloted our data collection process as well as our data collection sheets (see Section ‘Data Items’) on two meta-analytic systematic reviews that were deemed eligible within piloting of the screening procedure (England-Mason et al., 2023; Farkas et al., 2024). In addition, we will code three meta-analytic reviews jointly to reach a common understanding of the codebook and coding materials, in particular regarding quality- and transparency-related items. Through regular meetings between the raters regarding overarching coding questions, we will further strive

to ensure a common understanding of the codebook and minimize systematic rater bias or discrepancies.

We will assess and report interrater reliability for every variable (percentage agreement as recommended within AMSTAR-2; [Shea et al., 2017](#)) and report on the mean (*M*), standard deviation (*SD*), and range across all variables. Because our assessment of transparency-related variables is based on a self-developed, PRISMA-based checklist (see section 2.8.1.2), we will also report the prevalence-adjusted, bias-adjusted Kappa (PABAK; as calculated within [Stevenson and Sergeant, 2025](#)) for all categorical transparency-related variables and pairs of raters and the intraclass correlation (ICC) for continuous variables (as calculated within [Gamer et al., 2019](#)). We will not contact authors for additional information, because we are interested in the transparency and quality of reporting as is. We will not use any automation tools for data collection regarding these two research questions.

Data collection for research question 3 (RQ 3; consistency of statistical reporting) will be carried out through statcheck ([Nuijten and Polanin, 2020](#); [Nuijten and Epskamp, 2024](#)). Specifically, we will rely on the CRAN version at first date of analysis for the R package {statcheck} ([Nuijten and Epskamp, 2024](#)). Statcheck assesses whether reported *p*-values are consistent with their accompanying test statistics by recalculating them ([Nuijten and Polanin, 2020](#)). As basis for further analysis via statcheck, we will download .html (preferable) or .pdf-versions of all eligible studies within our meta-review, which will then be read into R and analyzed through functions provided within the package (see section 'Analysis'). Statcheck can only reliably detect reporting inconsistencies within (correctly) APA-formatted texts ([Nuijten and Epskamp, 2024](#)). At time of writing, four out of seven journals included within our meta-review require APA-formatting ([Table 3](#)), although this requirement often refers to formatting of references specifically (and not test statistics). We will still include all articles in our analysis (i.e., we will attempt to analyze all downloaded articles through statcheck functions), because it is possible that test statistics from other journals might be formatted in a way that is consistent with APA style and thus analyzable. However, it could be possible that we can only report on results from four journals.

2.8 Data items

We will code three groups of items and generate a fourth group of items via statcheck: (1) items related to general publication and study information, (2) transparency-related items (necessary for RQ 1), and (3) quality-related items (necessary for RQ 2) in the form of the AMSTAR 2 tool for assessing methodological quality ([Shea et al., 2017](#)). The full codebook is available as a Word document within [Supplementary material S3](#), and as the raw coding sheet (Excel; excepting AMSTAR 2 items which will be coded verbatim within a separate sheet not included for copyright reasons) within [Supplementary material S4](#). We will also analyze items from group (4), which are not manually coded but generated through statcheck.

2.8.1 Item content and development

2.8.1.1 Items related to general publication and study information (group 1)

For descriptive purposes, we will code the following information: Journal, topic of the systematic review (freetext; post-hoc grouping),

(sub-)clinical outcome investigated in children or adolescents (freetext; post-hoc grouping), whether or not the systematic review assessed intervention effects (yes/no), and the eligible age group(s) of the systematic review (infancy [0–24 months], childhood [2–10 years], adolescence [11–20 years]), number of included publications (preferred over number of studies due to easier coding access to this information), number of included participants (if reported). In addition, we also code further qualifying information for some transparency items (e.g., type of estimator for variance components).

2.8.1.2 Transparency-related items (group 2)

We focus on transparency-related criteria as reported within the PRISMA guidelines, given their ubiquitous use and current best-practice recommendation within health-related systematic reviews ([Kolaski et al., 2023](#)), including those from Psychology ([Steil et al., 2022](#)). To this end, we used transparency-related indicators as formalized within the core PRISMA guidelines, version 2020, within their explanation and elaboration document ([Page et al., 2021b](#)) as well as the PRISMA-S guidelines for search strategies ([Rethlefsen et al., 2021](#)), which were published within the same year and can thus be assumed to have a similar level of uptake.

While the PRISMA guidelines provide detailed guidance for review authors, they cannot be readily used for assessing transparency of existing systematic reviews, as the items are typically broad and consist of several transparency-related recommendations, which would need to be assessed (and scored) individually. To adapt the PRISMA indicators of transparent reporting for our purposes, we proceeded in the following steps: First, we manually checked codebooks of published meta-reviews within the social and medical sciences gathered within an unsystematic literature search ([Page et al., 2016, 2018](#); [Polanin et al., 2020](#); [Chow et al., 2021](#); [Chin et al., 2022](#); [López-Nicolás et al., 2022](#); [Nguyen et al., 2022](#); [Silva et al., 2023](#); [Wong and Bouchard, 2023](#); [Rethlefsen et al., 2024](#); [Sandoval-Lentisco et al., 2024](#)) and grouped items according to their superordinate PRISMA items (e.g., eligibility criteria, information sources). Second, we cross-checked the content of these items against the “essential items” list as reported within the PRISMA explanation and elaboration document ([Page et al., 2021b](#)) and adapted them where necessary. Third, for all essential items not covered, we created new items within the research team.

We did not create new items or adapted existing items for PRISMA essential items that would require comparison with primary studies (e.g., whether assumptions were made regarding missing or unclear information within primary studies; whether all effect sizes possible were extracted) or the review protocol (e.g., whether all planned analyses were conducted), as these analyses are excessively time-consuming and typically conducted within separate meta-scientific investigations, ([Tricco et al., 2016](#); see, e.g., [Lakens et al., 2017](#); [Maassen et al., 2020](#)). PRISMA is geared toward systematic reviews of interventions. Thus, we also adapted items to encompass a broader spectrum of reviews where necessary (e.g., a meta-analysis investigating prevalence rates does not need to define eligibility criteria for comparators and interventions). For items related to the transparency of the search procedure, we combined items from PRISMA with the more detailed items from PRISMA-S ([Rethlefsen et al., 2021](#)). Following ([Wong and Bouchard, 2023](#)), we opted against coding abstracts of systematic reviews (as formalized within the respective PRISMA guidelines for abstracts), because we deemed

length and content of abstracts as too journal-dependent for a meaningful analysis. For transparency of reporting of effect sizes, we also added an item (adapted from López-Nicolás et al., 2022) related to reporting of formulae for effect size calculation. While this is not mentioned explicitly within PRISMA guidelines, it is an important aspect of meta-analytic reproducibility (Lakens et al., 2017; Maassen et al., 2020). We also added items related to the reporting of interrater reliability of the screening and coding procedure (irrespective of its value, as would be assessed via AMSTAR 2), as well as regarding thresholds for reporting bias indication with statistical methods (Siegel et al., 2022). To better track the rising use of generative AI within the systematic review process as well as manuscript preparation (Ge et al., 2024) we also included an item on the use of generative AI within the disclosure section (note that some forms of generative AI use are also coded within the automation tools section). In all, we created (143) or adapted (25) 168 items related to transparency of meta-analytic systematic reviews (see Table 4 and Supplementary materials S3, S4).

As outlined within the introduction, scholars have voiced a repeated interest in and need for PRISMA (i.e., transparency) guidelines specifically for systematic reviews involving pediatric populations (i.e., the PRISMA-C guidelines; Kapadia et al., 2016). Presently (November 2024), these guidelines have not been published, but a proposition of possible adaptations exists (Farid-Kapadia et al., 2017a). We used these propositions to develop seven transparency-related items unique to systematic reviews from the field of CCAP. These involve: Mentioning the pediatric population of interest within the title (1) and study objectives (2; two separate items); providing justification for the age group(s) studied (3); providing replicable age-related eligibility criteria (4); providing a rationale for search terms related to the eligible age group (5); providing a field-specific classification of the strength of effect sizes (6); and discussing practical implications of findings for children, adolescents, and/or their caregivers (7). For descriptive purposes, we further developed items to assess whether the authors used search terms or filters related to the eligible age group(s), whether they used tested pediatric search filters (both not included in transparency scores), and whether they considered (within the eligibility criteria) or coded (within the data items) ethics-related criteria regarding child and adolescent populations within the reviewed primary studies (e.g., provision of IRB approval, waiver of parental consent).

2.8.1.3 Quality-related items (group 3)

To assess methodological quality, we will use AMSTAR 2 (Shea et al., 2017), which is the currently recommended and widely used tool for assessing quality of systematic reviews within meta-scientific investigations (Kolaski et al., 2023). With AMSTAR 2 it is possible to assess the methodological quality of (meta-analytic) systematic reviews irrespective of content domain within 16 items, with seven of them being deemed as critical (Shea et al., 2017). The reliability and validity of AMSTAR 2 has been found to be acceptable (Kolaski et al., 2023). We chose to not adapt AMSTAR 2 in any way to facilitate comparisons with other meta-scientific studies interested in the methodological quality of systematic reviews. Thus, we will use AMSTAR 2 as recommended within its accompanying documentation. [MS, SF, JR] will use AMSTAR 2 to assess the methodological quality of meta-analytic systematic reviews with MS assessing all eligible reviews and SF and JR each assessing half (corresponding to their coding of transparency-related items). We will pilot our shared

TABLE 4 Correspondence between transparency domains, PRISMA items, and items within our transparency checklist.

Section	PRISMA item number	Number (Max. Points) or range of number of items (Max. Points) within category
Domain		
Title	1	1 (1)
Introduction	3, 4	4–6 (4–6)
Eligibility criteria	5	6 (3)
Systematic search procedure	6, 7, 8 ^a , 16; all PRISMA-S items	Overall: 8–27 (8–24) Methods: 6: 1–11 (1–11), 7: 2–7 (2–7); 8: 3–6 (3–4) Results: 16: 2–3 (2–3)
Data collection	9 ^b , 10, 17	Overall: 6–12 (6–11) Methods: 5–11 (4–9) Results: 1 (2)
Risk of bias	11, 18	Overall: 5–9 (5–8) Methods: 3–7 (3–6) Results: 2 (2)
Effect measures & statistical synthesis	12 ^b , 13, 19, 20	Overall: 13–26 (11–22) Methods: 12: 3–4 (3–4), 13: 6–11 (4–7) Results: 19: 2 (2); 20: 2–9 (2–9)
Reporting bias assessment	14 ^c , 21	Overall: 2–10 (2–8) Methods: 1–8 (1–6) Results: 1–2 (1–2)
Certainty assessment	15, 22	Overall: 7–12 (5–9) Methods: 4–10 (3–7) Results: 2 (2)
Discussion	23	4 (4)
Open science practices ^d	24, 27	2–5 (2–4)
Disclosure practices	25, 26	2–3 (2–3)
Supplementary scores		
Automation tools	8, 9, 11, 15, 16	1–10 (1–10)
CCAP	–	6–7 (6–7)

^aItem about reporting of interrater agreement (in case of more than one rater) added.

^bItem about reference to formulae to calculate effect sizes added.

^cItem about thresholds for statistical bias detection methods added.

^dTo be analyzed in detail within separate investigations (preregistration to be amended, see below). Range items (excepting supplementary score): 59–121; range maximum points (excepting supplementary score): 53–104.

understanding of AMSTAR 2 based on two eligible reviews and, if necessary, code a third to reach consensus.

We deliberately did not include further CCAP-specific items related to methodological quality as we did for transparency. The reason is that there is currently, to the best of our knowledge, no broader consensus of experts within the field (or discussion thereof) what these CCAP-specific quality criteria could entail. We believe it is more conducive to the field that such a consensus is reached among different groups of experts and researchers rather than providing our idiosyncratic assessment of CCAP-related methodological quality

here. Note that, for transparency, while a fully fleshed out PRISMA-C guideline has not been published, these deliberations between experts have already been taking place (Kapadia et al., 2016; Farid-Kapadia et al., 2017a, 2017b) and could be adapted here.

2.8.1.4 Items generated by statcheck (group 4)

For our assessment of statistical consistency, we will use the open-source tool *statcheck* (Nuijten and Epskamp, 2024) in its version as an R package (CRAN release at date of analysis; currently version 1.5). *Statcheck* searches for results of NHSTs within manuscripts, recomputes p -values from reported test statistics (e.g., t -statistics, degrees of freedom), and compares the recomputed with the reported p -value (Nuijten et al., 2016; Nuijten and Epskamp, 2024). In case of a mismatch between the reported and recomputed p -values (with rounding and, possibly, one-sided testing being taken into account), the result is flagged as an inconsistency. If one p -value would be regarded as statistically significant (based on a chosen threshold, e.g., $\alpha = 0.05$) and the other would not (e.g., the reported but not the recomputed p -value is significant), the result is flagged as a gross inconsistency (Nuijten and Epskamp, 2024). It is important to note that it does not necessarily need to be the p -value that is incorrectly reported. Also, a reporting error in test statistics or degrees of freedom (e.g., a typo) could result in a recomputed p -value that is different from the one that was reported (Nuijten and Epskamp, 2024).

Technically, *statcheck* first converts a .pdf- or .html-file to plain text, whereby .html-files are preferable due to fewer issues in plain text conversion (Nuijten et al., 2016). Second, *statcheck* detects and then extracts test statistics, degrees of freedom, and p -values based on predefined string templates in APA format that are matched against the plain text. Third, *statcheck* recomputes the p -values and, fourth, compares it with the reported p -value (Nuijten et al., 2016). More information regarding technical details of *statcheck* can be found elsewhere (Nuijten et al., 2016; Nuijten and Polanin, 2020; Nuijten and Epskamp, 2024).

Several requirements need to be met for *statcheck* to function properly based on its mode of working (Nuijten and Epskamp, 2024): First, the results must be reported according to APA style. Second, the result needs to be reported in full, i.e., the test statistic, the degrees of freedom (if applicable) and the p -value need to be reported. Third, the result must be reported within the text and not within tables due to its reliance on predefined sets of strings to be detected (Nuijten and Epskamp, 2024). A validation study (based on primary studies) comparing manual extraction of results and *statcheck* found that *statcheck* was able to extract 67.5% of reported results (Nuijten et al., 2016). Thus, we expect only a subsample of all reported test statistics to be analyzable via *statcheck*, namely those reported in APA format in the main text. As a comparison, only 21.6% of meta-analytic systematic reviews within the social sciences have been found to contain at least one fully APA-formatted result from a NHST (Polanin and Nuijten, 2018).

We plan to use *statcheck::statcheck()* (Nuijten and Epskamp, 2024) in its default settings, including: (i) extraction of all test statistics; (ii) assumption of two-tailed tests; (iii) an alpha level of 0.05; (iv) $p \leq \alpha$ (vs. $p < \alpha$) is considered to be significant; (v) " $p = 0.000$ " (as opposed to " $p < 0.001$ ") is considered to be an error; (vi) correction for one-sided tests if mentioned within text; (vii) only p -values that comply fully with APA-formatting are included in the output. We will also use default settings for classifications of inconsistencies (i.e.,

mismatch between reported and recomputed p -values not explained via rounding or one-sided testing) and gross inconsistencies (the reported p -value is < 0.05 and the recomputed is not, or vice versa).

Statcheck's performance in detecting statistical inconsistencies (given necessary preconditions) has been assessed elsewhere (Nuijten et al., 2016, 2017) and is not the focus of this paper. To gauge the accuracy and completeness of the data extraction procedure within our study (*statcheck*'s main source of errors or non-detection), we will randomly draw 10% (rounded to the next integer) of eligible meta-analytic systematic reviews where *statcheck* was able to detect at least one NHST reported in APA-style. For example, if *statcheck* was able to detect at least one NHST reported in APA-style in 13 meta-analytic reviews (21.6% of the estimated 60 studies), we will randomly select two meta-analytic systematic reviews. From these, we will manually extract all reported test statistics (including from tables) and compare the number of extracted statistics to those extracted by *statcheck*. Reasons for failure of detection (e.g., incomplete test statistic, reported in a table) will be noted.

By default, *statcheck* in its current CRAN version 1.5 (Nuijten and Epskamp, 2024) provides the following output per analyzed document (here: review), excepting variables extracted to facilitate manual post-hoc checks related to the correct import: type of test statistic (e.g., Q), first and second degrees of freedom (if applicable), test statistic, reported p -value, recomputed p -value, errors (true/false; i.e., whether reported and recomputed p -values are incongruent), decision errors (true/false; i.e., whether either the reported or recomputed p -value is significant and the other is not), whether or not text indicated one-sided testing was detected (true/false), and the proportion of correctly APA-formatted test statistics (as an indication of possibly missed results due to reporting issues). *Statcheck* also alerts users to possible manual checks (e.g., checking the significance level in case a result would have been correct in case of $\alpha = 0.01$). We will respond to these alerts by checking full-texts and adapting the *statcheck* input for this review if necessary (e.g., $\alpha = 0.01$). Following (Nuijten and Polanin, 2020), we will also conduct manual checks of errors classified as decision errors.

2.9 Planned analyses

Our aim is to assess meta-analytic systematic reviews from seven leading journals within the field but not to generalize to a wider population (which would, ideally, include a random sample). Thus, we will generally present our results descriptively and will not, for example, provide confidence intervals (e.g., for proportions). Notable exceptions include our exploratory analyses, where we will provide p -values and/or confidence intervals for the sake of completeness, but emphasize the descriptive nature of our analyses. We assume a two-tailed p -value of < 0.05 as our threshold for significance for these purposes.

2.9.1 Descriptive sample statistics

We will report descriptive sample statistics (counts and proportions for categorical variables, M , SD , range for continuous variables) for the following variables, possibly stratified by journal if deemed valuable: Journal, type and topic of the meta-analytic systematic reviews, (sub-)clinical conditions studied within pediatric populations, pediatric age groups, type of effect size metric, type of

meta-analytic model (fixed-effect/equal-effects model; two-level random-effects model; three-or-more-level random-effects model), number of included publications, number of participants.

2.9.2 Research question 1 (transparency)

2.9.2.1 Scoring and primary metric

In general, we will assign 0 to all transparency-related items where necessary information was not reported and 1 to those where necessary information was reported. In some cases, answer options allow for scoring of a “partial” reporting of information, for which we will assign 0.5 points. Evidently, not all items are applicable for every meta-analytic systematic review. For example, for a review that did not report undertaking a risk of bias assessment, we will code all respective items related to transparent reporting of risk of bias assessment (e.g., specification of tool; reporting of number of authors involved in assessment) as “not applicable” (i.e., as a missing value). Importantly, this means that we do not code these items as a lack of transparency (i.e., by awarding 0 points). Rather, the lack of the risk of bias assessment is related to and reflected within study quality, which is assessed via AMSTAR 2 items. Notable exceptions include, for example, the reporting of studies excluded during full-text screening together with a justification, where transparent reporting is a quality criterion within AMSTAR 2. Additionally, some essential items within the PRISMA guidelines were explicitly phrased as “if-statements” (Page et al., 2021a) and thus only applicable to reviews where these conditions were met (e.g., if the systematic review assesses an intervention, it should be briefly described how this intervention works within the introduction section). In these cases, we also code items as “not applicable” if the condition is not met.

This means that every systematic review within our sample will possibly have a different maximum possible score, because different sets of items will apply to different reviews. We solve this through calculating the percentage of maximum possible (POMP) score (Cohen et al., 1999) as our primary metric, where the transparency score is expressed as the tallied score for a given review divided by the maximum possible score for this review. Thus, as 0 is the minimum score for all reviews, the POMP score of review i can be expressed as: $\text{POMP}_i = (\text{Score}_i / \text{Max.Score}_i) * 100$. Table 4 reports the possible range of items and points for all broad transparency-related domains.

2.9.2.2 Scores

We will form several scores based on our coding of transparency-related items as well as based on different weighting schemes for PRISMA domains. As outlined above, all scores will be expressed as POMP (see above).

First, we will calculate overall transparency scores by forming POMP scores based on all transparency items excepting those related to automation tools and CCAP (which will be scored separately, see below). We will form two different version of these overall scores: First, we form an unweighted version where it is not considered that some PRISMA items comprise more individual essential items. For example, the PRISMA item related to a review’s title (item #1) comprises one essential item and is thus awarded only one point, whereas item #12 (effect measures) comprises three essential items (and we added a fourth related to the provision of formulae for calculation of effect sizes, see above) and is thus awarded four points. Second, we will thus also calculate a weighted

version, where each PRISMA item is weighted equally (i.e., contributing equally to the overall score, regardless of the number of essential items). For our example above, this would mean that the essential item for PRISMA item #1 would still be scored with one point, but our essential items for PRISMA item #12 would be awarded $1/4 = 0.25$ points each.

Second, this broad score will be supplemented by a finer-grained, domain-based POMP score to elucidate transparent reporting practices within different domains of a review. These domains include the following sections (see Table 4 for correspondence between these domains, the PRISMA items, as well as our items based on PRISMA essential items within the broad items): Title, introduction, eligibility criteria, systematic search procedure, data collection, risk of bias, statistical synthesis (i.e., meta-analysis), reporting bias assessment, certainty assessment, discussion, open science practices, disclosure practices. Domains will be scored as missing in case all transparency-related items were coded as “not applicable” for this domain (e.g., a review that did not conduct a certainty assessment will not be scored within this domain).

Third, we will compute a separate automation tool score for those reviews that used automation tools within sections systematic search procedure (PRISMA item #8), data collection (#9), and risk of bias (#11), as outlined within the PRISMA guidelines (Page et al., 2021a). As the uptake of automation tools within research synthesis is an on-going process (Van Altena et al., 2019), we deemed it reasonable to score and report usage of these tools separately from reporting of other, more established, components of a review.

Fourth, we will compute a separate, CCAP-related transparency score that focuses on transparent reporting as it relates to CCAP-specific information. Because we derived seven CCAP-related transparency items, we will report adherence to transparent reporting for every item separately and will also form an overall score.

2.9.2.3 Presentation of results and answer to RQ 1

We will report results per item within Supplementary material (overall and stratified by journal). Within the main text, we will present M and SD of POMP scores (overall; by domain; automation tools; CCAP) across all reviews in two version: First, an unweighted version (i.e., mean score across all reviews) and second, a weighted version where differing numbers of reviews across journals are considered. For this weighted version, we will form a grand mean of the mean scores by journal. For the overall score, this will be crossed with the weighted and unweighted scoring related to PRISMA items (see above), resulting in $2 \text{ (PRISMA)} \times 2 \text{ (journal)} = 4$ scores. Results by journal will similarly be made available within Supplementary material. Within the main text, we will focus on emphasizing broad patterns found within the results and highlight certain domains (e.g., particularly transparent or non-transparent parts of reviews) as deemed insightful. Findings will be presented in tabular form, narratively, and through visualization (e.g., stacked bar plots). To answer RQ 1, we will focus on overall and domain-based scores, which we will contextualize within the discussion both with regard to between-journal variability as well as with regard to findings from other (sub-) disciplines. Our aim is not to provide readers with a single percentage score, but rather to elucidate areas of high and low transparency as well as how the latter could be improved.

2.9.3 Research question 2 (quality)

2.9.3.1 Scoring

We will score AMSTAR 2 based on its accompanying information. For our overall rating of the confidence in a review, we will adhere to recommendations as given within the original AMSTAR 2 publication, where high confidence is assigned to reviews with none or only one non-critical weakness (i.e., in a domain deemed as non-critical); moderate confidence where more than one non-critical weakness has been identified, low in cases where at least one flaw in a critical domain has been identified, and critically low when more than one critical flaw has been identified (Shea et al., 2017).

AMSTAR 2 does not provide an overall score to avoid concealing critical flaws within a review (Shea et al., 2017). We agree with this notion and will thus focus on presenting results for single items (see below) as well as an overall assessment of confidence within the review. However, for our exploratory analyses (see below), we will need to assign numeric values to AMSTAR 2 ratings. For the purpose of these analyses, we will assign “Yes” ratings with a score of 1, “Partial Yes” ratings with a score of 0.5, and “No” ratings with a score of 0 and create a sum score of all items. If, similar to our transparency ratings, we encounter instances where an AMSTAR 2 item is not codable for a given review, we will revert to POMP scores if necessary.

To answer RQ 2, we will focus on results from single items as well as overall confidence ratings. As for RQ 1, both will be contextualized within the discussion section.

2.9.3.2 Presentation of results

We will present results per item (items deemed as critical within AMSTAR 2, i.e., #2, #4, #7, #9, #11, #13, #15, will be highlighted) as well as for the overall confidence rating as recommended (Shea et al., 2017). Similar to results from our transparency assessment, we will report unweighted and weighted results that do not (unweighted) and do (weighted) take differing numbers of reviews within journals into account. We will present results in tabular form, narratively, and visually (e.g., by using stacked bar plots) and within main text and [Supplementary material](#) based on the nature of the results.

2.9.4 Research question 3 (consistency)

For RQ 3, we will report the number of errors (i.e., inconsistencies between reported and re-computed p -values) as well as the number of gross decision errors (i.e., the assessment of significance differs based on reported and re-computed p -values). Both will be reported overall (unweighted and weighted) as well as stratified by journal. To answer RQ 3, we will focus on these two types of errors and contextualize them within the discussion section in comparison with other statcheck-based evaluations of corpora of primary studies and meta-analytic systematic reviews.

2.9.5 Exploratory analyses and effect strength interpretation

We will conduct a series of exploratory analyses, all of which are expressed as associations between two variables (Pearson r or Cohen d); the latter calculated according to default settings within `effectsize:cohens_d()` (Ben-Shachar et al., 2020), corresponding to formulae 4.2–4.6 as reported within Borenstein et al. (2021), where nesting within journal is not taken into account due to the descriptive nature of our analysis. Specifically, we will test: (1) Associations

between reviews' transparency and quality, as expressed by transparency and quality overall scores; (2) associations between reviews' transparency and inconsistent reporting, as expressed by overall transparency scores and dummy-coded variables related to number of errors (i.e., at least one error within the review) and number of decision errors (i.e., at least one decision error within the review); (3) associations between reviews' transparency scores and the journal's TOP factor; (4) associations between reviews' transparency and self-reported adherence to PRISMA guidelines (yes vs. no), (5) associations between a reviews' transparency and journal page limit (no page limit vs. long page limit; no page limit vs. short page limit; see [Table 3](#)). Transparency scores will always be used twice for analysis (weighted/unweighted by PRISMA criterion).

Regarding effect size interpretation, there are currently no established guidelines for what constitutes meaningful effect strength within meta-reviews such as this. Thus, we resort to an *ad hoc* approach based on previous meta-scientific findings within the literature and our own considerations. For effects expressed as Cohen d (analyses 2, 4, 5), we assume a difference of 5% of transparency scores between groups as practically meaningful. Based on findings within Psychology (Polanin et al., 2020), which align with standard deviations of PRISMA scores found in other fields (Dhillon et al., 2022; Peña et al., 2022; Streck et al., 2022), we assume a pooled standard deviation of 14%, which would result in a $d = 5/14 = 0.35$, which we round down to 0.30 to be in line with well-established guidelines for typical, lower thresholds of small-to-medium-sized effects (Cohen, 1988). For effects expressed as Pearson r (analyses 1, 3), we turn to previous associations found between PRISMA and AMSTAR 2 scores (to our knowledge, PRISMA scores have not been correlated with journals' TOP scores) that have been documented in the literature, (e.g., Rainkie et al., 2020; Peña et al., 2022; Javidan et al., 2023; Sehmbi et al., 2023), ranging between 0.33 (Rainkie et al., 2020) and 0.78 (Sehmbi et al., 2023). Here, we use the lower limit of $r = 0.30$ as our assumed threshold for a meaningful effect.

2.10 Data management and availability plan

See [Supplementary material S5](#).

2.11 Status and timeline

Upon submission to the journal and concurrent preregistration as a preprint (November, 2024), the status of our project was the following: Preliminary searches as well as piloting of the eligibility criteria and the screening procedure in a randomly drawn sample of 55 records had been conducted. The coding procedure had been piloted on two meta-analytic systematic reviews. We had not started the main search, screening, and coding procedures.

Upon reception of reviews (June, 2025), the status of the project is as follows: We have completed the systematic search and title-and-abstract screening and established a final sample of eligible meta-analytic systematic reviews. Coding for all three main pillars of this meta-review (transparency, quality, and statistical consistency) is currently underway. Reviewers' comments did not address any steps completed in the meantime that would need to be repeated.

2.12 Potential follow-up investigations

We plan to conduct further investigations based on the review sample obtained from this study. First, we plan to analyze adherence to open science practices of eligible meta-analytic systematic reviews in more detail within a separate report, including practical recommendations to improve adherence to open science principles within meta-analytic systematic reviews. Second, we plan to assess the computational reproducibility of meta-analyses within our sample, similar to López-Nicolás et al. (2024). Third, we plan to compare review protocols and preregistrations with published reviews, similar to Sandoval-Lentisco et al. (2025). All of these investigations will be preregistered.

3 Discussion

This study protocol for a meta-review outlines how we plan to assess transparency, methodological quality, and statistical consistency of recent (2022–2024) meta-analytic systematic reviews published in seven leading CCAP journals. Our goal is to identify strengths in these areas and highlight opportunities for improvement. In doing so, we ultimately hope to provide further guidance for applied meta-analysts, reviewers, and editors within the field. To our knowledge, this is the first meta-scientific investigation focused specifically on meta-analytic systematic reviews within CCAP (see Sifers, 2002; Raad et al., 2008, for slightly dated investigations at the primary study level) and one of the first within Clinical Psychology (López-Nicolás et al., 2022). As such, it will contribute to a better understanding of the trustworthiness of meta-analytic systematic reviews within these fields (and whether and how these could be improved) and provide insights into how they compare to other subdisciplines of Psychology in terms of reporting transparency, quality, and statistical consistency.

Our transparency assessment tool is based on previous literature (Page et al., 2016, 2018; Polanin et al., 2020; Chow et al., 2021; Chin et al., 2022; López-Nicolás et al., 2022; Nguyen et al., 2022; Silva et al., 2023; Wong and Bouchard, 2023; Rethlefsen et al., 2024; Sandoval-Lentisco et al., 2024) and aligns with PRISMA essential items (vs. the broader PRISMA checklist), which allows for a nuanced, comprehensive, systematic, and field-independent assessment. We also include field-specific transparency items for CCAP, which are literature-based (Kapadia et al., 2016; Farid-Kapadia et al., 2017a, 2017b). Methodological quality will be evaluated by using AMSTAR 2 (Shea et al., 2017), enabling comparisons with previous studies. By applying statcheck (Nuijten and Epskamp, 2024), we also provide new evidence on the prevalence of reporting errors within research syntheses, adding to a predominantly primary study-based evidence base. Finally, we will assess data and code sharing practices as part of our transparency assessment, disconcertingly uncommon within other (sub-)disciplines (Polanin et al., 2020; López-Nicolás et al., 2022; Nguyen et al., 2022). This will help to shed light on open science practices within CCAP and Clinical Psychology more broadly, where uptake of these practices has been found to be low (Howard et al., 2024).

Several limitations of this planned meta-review are foreseeable at this stage. First, our sample will consist of recent meta-analytic systematic reviews from leading, high-impact CCAP journals, not a

random (and thus representative) selection. Therefore, our results will likely represent an upper threshold of transparency, quality, and statistical reporting consistency and generalizations to the field cannot be drawn. For this reason, results will also be presented descriptively. We made this sampling choice for several reasons: For one, this is the first, notably unfunded, meta-scientific investigation of this kind within the field of CCAP, where an assessment of this “upper threshold” of reporting and quality characteristics might be particularly insightful for future applied and meta-scientific works in this area. Applied meta-analysts might turn to meta-analytic systematic reviews published within these outlets to guide their own work and thus domains of reporting non-transparency or lack of quality could be further perpetuated, because examples from leading journals would indicate that a review is nonetheless publishable. Similarly, reviewers and editors, tasked with assessing potentially highly impactful reviews for the field, can focus their attention on domains where reporting has been found to be lacking quality or transparency. In addition, as there is no dedicated category for CCAP (as, for example, compared to Clinical Psychology as a whole) within large-scaled scientific databases (e.g., *Web of Science*, *Scopus*), it is difficult to establish the sampling frame for such a randomly drawn sample. As for meta-reviews of somewhat narrower focus, the best option would be a keyword search, but it is currently unclear on how a search string encompassing all outcome domains eligible for our (or similar) investigations could look like. Thus, we opted for a journal-based strategy, as is also common within meta-scientific investigations (e.g., Polanin et al., 2020; Siegel et al., 2022). However, the sample from this meta-review might help to establish such a search string in the future, possibly enabling a more comprehensive assessment of reporting practices within CCAP.

Second, our comprehensive assessment of reporting transparency is not a published and validated tool, even though it is grounded in previous literature and aligns with current PRISMA (Page et al., 2021a) and PRISMA-S (Rethlefsen et al., 2021) guidelines. However, such a tool does (to our knowledge) not exist so far. Similarly, our CCAP-specific items are based on preparatory works for PRISMA-C guidelines, but ultimately reflect our subjective judgments about transparency-related issues relevant for the field of CCAP. Ideally, consensus among experts would guide these judgments, which exceeds the scope of this study. Thus, we view the inclusion of these items as exploratory and possibly informative for future, consensus-based work in this regard.

Third, we do not assess reporting practices requiring access to primary studies or study protocols (if existent), such as the correct extraction and computation of effect sizes, comparisons of primary outcomes within protocols and final publications, or meta-analytic reproducibility. These assessments have been found to be excessively time-consuming (Tricco et al., 2016; Lakens et al., 2017; Maassen et al., 2020) and typically warrant their own investigation. Our sample could serve as the basis for focused investigations into these domains.

Fourth, while the use of NHSTs within meta-analyses is common (averaging between 28 and 60 per publication; Polanin and Pigott, 2015; Nuijten and Polanin, 2020), the number of NHSTs reported in sufficient detail for re-analysis via statcheck is much lower ($Md = 4$; Polanin and Nuijten, 2018). This limitation is common to all investigations into statistical reporting consistency using statcheck (the only dedicated tool to date) and might indicate a broader problem of insufficient or unsystematic reporting of *p*-values and accompanying

test statistics within meta-analyses. Still, our conclusions to be drawn from our statcheck-based analysis might be limited, also considering that statcheck is not able to identify all NHSTs provided within a paper, for example those reported within tables, and relies on results to be reported in APA style (Nuijten and Epskamp, 2024).

Fifth, we want to acknowledge that our investigation of quality focuses on methodological quality as assessed within AMSTAR 2. Thus, we do not assess aspects of quality that would require content-knowledge (e.g., the suitability of eligibility criteria or the search string) because this is not possible for such a diverse set of systematic reviews.

In conclusion, our planned meta-review holds the potential to shed light on transparency, quality, and statistical consistency of meta-analytic systematic reviews within the field of CCAP. In doing so, we hope to provide guidance for researchers, reviewers, and editors within this field and lay the foundation for future, larger-scaled works in this field.

Author contributions

MS: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Visualization, Writing – original draft. SF: Conceptualization, Methodology, Software, Visualization, Writing – review & editing. JR: Investigation, Software, Visualization, Writing – review & editing. MZ: Conceptualization, Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. Funding for open access publication was provided by the University of Vienna. The authors declare that no other funding was received for this research.

Acknowledgments

This study protocol was presented as a poster at the META-REP conference 2024 (October 28–31) held in Munich, Germany. We would like to thank Madlene Radosavljevic for her assistance in identifying leading journals within the field.

References

- Alozkan-Sever, C., Uppendahl, J. R., Cuijpers, P., De Vries, R., Rahman, A., Mitterdorfer-Rutz, E., et al. (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. *J. Child Psychol. Psychiatry* 64, 1776–1788. doi: 10.1111/jcpp.13891
- American Academy of Child and Adolescent Psychiatry (2021). Practice parameters for child and adolescent psychiatric practice. Available online at: https://www.aacap.org/AACAP/Practice/Clinical%20Practice%20Guidelines/AACAP/Resources_for_Primary_Care/Practice_Parameters_and_Resource_Centers/Practice_Parameters.aspx (November 26, 2024).
- American Psychiatric Association (1994). Diagnostic and statistical manual of mental disorders. 4th Edn. Washington, DC: American Psychiatric Publishing Inc.
- American Psychiatric Association (2013). Diagnostic and statistical manual of mental disorders. 5th Edn. Washington, DC: American Psychiatric Association.
- Assink, M., and Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: a step-by-step tutorial. *Quant. Methods Psychol.* 12, 154–174. doi: 10.20982/tqmp.12.3.p154
- Bakker, M., and Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behav. Res. Methods* 43, 666–678. doi: 10.3758/s13428-011-0089-5
- Barry, C. T., Kim, H., and Halter, B. M. (2023). “Developmental psychopathology” in Handbook of clinical child psychology. ed. J. L. Matson (Cham: Springer International Publishing), 19–34.
- Bellato, A., Cristea, I. A., Giovane, C. D., Fazel, S., Polanczyk, G. V., Solmi, M., et al. (2023). Evidence-based child and adolescent mental health care: the role of high-quality and transparently reported evidence synthesis studies. *JCPP Adv.* 3:e12197. doi: 10.1002/jcv2.12197
- Ben-Shachar, M., Lüdtke, D., and Makowski, D. (2020). Effectsize: estimation of effect size indices and standardized parameters. *J. Open Source Softw.* 5:2815. doi: 10.21105/joss.02815
- Biondi-Zoccai, G. (Ed.) (2016). Umbrella reviews: evidence synthesis with overviews of reviews and meta-epidemiologic studies. Cham: Springer International Publishing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that Gen AI was used in the creation of this manuscript. [This statement is copied verbatim from the methods section of the paper]: During preparation of this manuscript, ChatGPT (versions 3.5 and 4.0; OpenAI, 2024) was used on author-generated text [MS] within narrative sections (Introduction, Discussion) to edit for typography, grammar, clarity, and brevity only, akin to a human co-author (Nakagawa and Lagisz, 2024). Prompts included varieties of the statement “Please edit the following paragraph for brevity and clarity and point out what was corrected: [author-generated text copied in full]”. MS incorporated these suggestions as seen fit and, together with the other authors, takes full responsibility for the content of this manuscript. No new text segments, ideas, or references were generated with the help of ChatGPT, any other large language model, or any other generative artificial intelligence assistance.

Publisher’s note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2025.1535606/full#supplementary-material>

- Borenstein, M., Hedges, L. V., Higgins, J., and Rothstein, H. R. (2021). Introduction to meta-analysis. 2nd Edn. Hoboken: Wiley.
- Brown, M., Matson, J. L., and Tevis, C. (2023). "History of clinical child psychology" in Handbook of clinical child psychology. ed. J. L. Matson (Cham: Springer International Publishing), 3–17.
- Byrne, S., Cobham, V., Richardson, M., and Imuta, K. (2023). Do parents enhance cognitive behavior therapy for youth anxiety? An overview of systematic reviews over time. *Clin. Child. Fam. Psychol. Rev.* 26, 773–788. doi: 10.1007/s10567-023-00436-5
- Caulley, L., Cheng, W., Catalá-López, F., Whelan, J., Khoury, M., Ferraro, J., et al. (2020). Citation impact was highly variable for reporting guidelines of health research: a citation analysis. *J. Clin. Epidemiol.* 127, 96–104. doi: 10.1016/j.jclinepi.2020.07.013
- Chin, J. M., Grown, B., Sebastian, J., Page, M. J., and Nakagawa, S. (2022). The transparency and reproducibility of systematic reviews in forensic science. *Forensic Sci. Int.* 340:111472. doi: 10.1016/j.forsciint.2022.111472
- Chow, J. C., Sjogren, A. L., and Zhao, H. (2021). Reporting and reproducibility of meta-analysis in speech, language, and hearing research. *J. Speech Lang. Hear. Res.* 64, 2786–2793. doi: 10.1044/2021_JSLHR-21-00047
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences. 2nd Edn. Hillsdale, NY: Lawrence Erlbaum.
- Cohen, P., Cohen, J., Aiken, L. S., and West, S. G. (1999). The problem of units and the circumstance for POMP. *Multivar. Behav. Res.* 34, 315–346. doi: 10.1207/S15327906MBR3403_2
- Cooper, H., Hedges, L. V., and Valentine, J. C. (2019). The handbook of research synthesis and meta-analysis. 3rd Edn. New York: Russell Sage Foundation.
- Cortese, S., Tomlinson, A., and Cipriani, A. (2019). Meta-review: network meta-analyses in child and adolescent psychiatry. *J. Am. Acad. Child Adolesc. Psychiatry* 58, 167–179. doi: 10.1016/j.jaac.2018.07.891
- Dal Santo, T., Rice, D. B., Amiri, L. S. N., Tasleem, A., Li, K., Boruff, J. T., et al. (2023). Methods and results of studies on reporting guideline adherence are poorly reported: a meta-research study. *J. Clin. Epidemiol.* 159, 225–234. doi: 10.1016/j.jclinepi.2023.05.017
- Davis, R. S., Meiser-Stedman, R., Afzal, N., Devaney, J., Halligan, S. L., Lofthouse, K., et al. (2023). Systematic review and meta-analysis: group-based interventions for treating posttraumatic stress symptoms in children and adolescents. *J. Am. Acad. Child Adolesc. Psychiatry* 62, 1217–1232. doi: 10.1016/j.jaac.2023.02.013
- Dekkers, T. J., Groenman, A. P., Cuijpers, P., Hoekstra, P. J., Luman, M., Orobio De Castro, B., et al. (2023). Commentary: why treatment is the best choice for childhood mental disorders – a commentary on Roest et al. (2022). *J. Child Psychol. Psychiatry* 64, 470–473. doi: 10.1111/jcpp.13715
- Dhillon, J., Khan, T., Siddiqui, B., Torgerson, T., Ottwell, R., Johnson, A. L., et al. (2022). Analysis of systematic reviews in clinical practice guidelines for head and neck cancer. *Laryngoscope* 132, 1976–1983. doi: 10.1002/lary.30051
- England-Mason, G., Andrews, K., Atkinson, L., and Gonzalez, A. (2023). Emotion socialization parenting interventions targeting emotional competence in young children: a systematic review and meta-analysis of randomized controlled trials. *Clin. Psychol. Rev.* 100:102252. doi: 10.1016/j.cpr.2023.102252
- Epskamp, S. (2019). Reproducibility and replicability in a fast-paced methodological world. *Adv. Methods Pract. Psychol. Sci.* 2, 145–155. doi: 10.1177/2515245919847421
- Espada, J. P., Sánchez-López, A., and Morales, A. (2023). Effectiveness of psychological treatments for depression in childhood and adolescence: a review of reviews. *Rev. Psicol. Clínica Con Niños Adolesc.* 10, 68–83. doi: 10.21134/rpcna.2023.10.1.6
- Faggion, C. M., and Diaz, K. T. (2019). Overview authors rarely defined systematic reviews that are included in their overviews. *J. Clin. Epidemiol.* 109, 70–79. doi: 10.1016/j.jclinepi.2019.01.004
- Farid-Kapadia, M., Askie, L., Hartling, L., Contopoulos-Ioannidis, D., Bhutta, Z. A., Soll, R., et al. (2017a). Do systematic reviews on pediatric topics need special methodological considerations? *BMC Pediatr.* 17:57. doi: 10.1186/s12887-017-0812-1
- Farid-Kapadia, M., Joachim, K. C., Balasingham, C., Clyburne-Sherin, A., and Offringa, M. (2017b). Are child-centric aspects in newborn and child health systematic review and meta-analysis protocols and reports adequately reported?—two systematic reviews. *Syst. Rev.* 6:31. doi: 10.1186/s13643-017-0423-9
- Farkas, B. F., Takacs, Z. K., Kollárovics, N., and Balázs, J. (2024). The prevalence of self-injury in adolescence: a systematic review and meta-analysis. *Eur. Child Adolesc. Psychiatry* 33, 3439–3458. doi: 10.1007/s00787-023-02264-y
- Foo, Y. Z., O'Dea, R. E., Koricheva, J., Nakagawa, S., and Lagisz, M. (2021). A practical guide to question formation, systematic searching and study screening for literature reviews in ecology and evolution. *Methods Ecol. Evol.* 12, 1705–1720. doi: 10.1111/2041-210X.13654
- Galán-Luque, T., Serrano-Ortiz, M., and Orgilés, M. (2023). Effectiveness of psychological interventions for child and adolescent specific anxiety disorders: a systematic review of systematic reviews and meta-analyses. *Rev. Psicol. Clínica Con Niños Adolesc.* 10, 31–41. doi: 10.21134/rpcna.2023.10.1.3
- Gamer, M., and Lemon, J. (2019). irr: Various coefficients of interrater reliability and agreement (version 0.84.1). Available online at: <https://CRAN.R-project.org/package=irr> (November 26, 2024).
- García-Fernández, G., Krotter, A., and Udeanu, A. (2023). Effectiveness of psychological interventions for eating disorders in adolescence: an overview of systematic reviews. *Rev. Psicol. Clínica Con Niños Adolesc.* 10, 116–126. doi: 10.21134/rpcna.2023.10.1.10
- Gates, M., Gates, A., Pieper, D., Fernandes, R. M., Tricco, A. C., Moher, D., et al. (2022). Reporting guideline for overviews of reviews of healthcare interventions: development of the PRIOR statement. *BMJ* 378:e070849. doi: 10.1136/bmj-2022-070849
- Ge, L., Agrawal, R., Singer, M., Kannapiran, P., De Castro Molina, J. A., Teow, K. L., et al. (2024). Leveraging artificial intelligence to enhance systematic reviews in health research: advanced tools and challenges. *Syst. Rev.* 13:269. doi: 10.1186/s13643-024-02682-2
- Harris, J. L., Swanson, B., and Petersen, I. T. (2024). A developmentally informed systematic review and meta-analysis of the strength of general psychopathology in childhood and adolescence. *Clin. Child. Fam. Psychol. Rev.* 27, 130–164. doi: 10.1007/s10567-023-00464-1
- Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., et al. (2024). Cochrane handbook for systematic reviews of interventions version 6.5 (updated august 2024). Available online at: www.training.cochrane.org/handbook (November 26, 2024).
- Hohn, R. E., Slaney, K. L., and Tafreshi, D. (2020). An empirical review of research and reporting practices in psychological meta-analyses. *Rev. Gen. Psychol.* 24, 195–209. doi: 10.1177/1089268020918844
- Howard, A. L., Neal, T. M. S., Kirtley, O. J., Urry, H. L., and Tackett, J. L. (2024). Open science at clinical psychological science: reflections on progress, lessons learned, and suggestions for continued improvement. *Clin. Psychol. Sci.* 13, 195–204. doi: 10.1177/21677026241255882
- Hu, K., Zhao, L., Zhou, Q., Mei, F., Gao, Q., Chen, F., et al. (2021). Inconsistencies in study eligibility criteria are common between non-Cochrane systematic reviews and their protocols registered in PROSPERO. *Res. Synth. Methods* 12, 394–405. doi: 10.1002/jrsm.1476
- Ioannidis, J. P. A. (2018). Meta-research: why research on research matters. *PLoS Biol.* 16:e2005468. doi: 10.1371/journal.pbio.2005468
- Ioannidis, J. P. A., Fanelli, D., Dunne, D. D., and Goodman, S. N. (2015). Meta-research: evaluation and improvement of research methods and practices. *PLoS Biol.* 13:e1002264. doi: 10.1371/journal.pbio.1002264
- Javidan, A., Alaichi, J., Nassar, Y., Li, A., Balta, K. Y., and Naji, F. (2023). Completeness of reporting in systematic reviews and meta-analyses in vascular surgery. *J. Vasc. Surg.* 78, 1550–1558.e2. doi: 10.1016/j.jvs.2023.04.009
- John, L. K., Loewenstein, G., and Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychol. Sci.* 23, 524–532. doi: 10.1177/0956797611430953
- Kandel, E. R., Koester, J., Mack, S., and Siegelbaum, S. (Eds.) (2021). Principles of neural science. Sixth Edn. New York: McGraw Hill.
- Kapadia, M. Z., Askie, L., Hartling, L., Contopoulos-Ioannidis, D., Bhutta, Z. A., Soll, R., et al. (2016). PRISMA-children (C) and PRISMA-protocol for children (P-C) extensions: a study protocol for the development of guidelines for the conduct and reporting of systematic reviews and meta-analyses of newborn and child health research. *BMJ Open* 6:e010270. doi: 10.1136/bmjopen-2015-010270
- Koensgen, N., Rombey, T., Allers, K., Mathes, T., Hoffmann, F., and Pieper, D. (2019). Comparison of non-Cochrane systematic reviews and their published protocols: differences occurred frequently but were seldom explained. *J. Clin. Epidemiol.* 110, 34–41. doi: 10.1016/j.jclinepi.2019.02.012
- Koffel, J. B., and 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:e0163309. doi: 10.1371/journal.pone.0163309
- Kolaski, K., Logan, L. R., and Ioannidis, J. P. A. (2023). Guidance to best tools and practices for systematic reviews. *Syst. Rev.* 12:96. doi: 10.1186/s13643-023-02255-9
- Kwan, I., Dickson, K., Richardson, M., MacDowall, W., Burchett, H., Stansfield, C., et al. (2020). Cyberbullying and children and young people's mental health: a systematic map of systematic reviews. *Cyberpsychol. Behav. Soc. Netw.* 23, 72–82. doi: 10.1089/cyber.2019.0370
- Lakens, D., Page-Gould, E., Van Assen, M. A. L. M., Spellman, B., Schönbrodt, F. D., Hasselman, F., et al. (2017). Examining the reproducibility of meta-analyses in psychology: a preliminary report. *MetaArXiv [Preprint]*. doi: 10.31222/osf.io/xbfj
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gotzsche, P. C., Ioannidis, J. P. A., et al. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *PLoS Med.* 6:e1000100. doi: 10.1371/journal.pmed.1000100
- Lin, S. C., Kehoe, C., Pozzi, E., Lontos, D., and Whittle, S. (2024). Research review: child emotion regulation mediates the association between family factors and internalizing symptoms in children and adolescents – a meta-analysis. *J. Child Psychol. Psychiatry* 65, 260–274. doi: 10.1111/jcpp.13894
- López-Nicolás, R., Lakens, D., López-López, J. A., Rubio-Aparicio, M., Sandoval-Lentisco, A., López-Ibáñez, C., et al. (2024). Reproducibility of published meta-analyses on clinical-psychological interventions. *Adv. Methods Pract. Psychol. Sci.* 7:25152459231202929. doi: 10.1177/25152459231202929

- López-Nicolás, R., López-López, J. A., Rubio-Aparicio, M., and Sánchez-Meca, J. (2022). A meta-review of transparency and reproducibility-related reporting practices in published meta-analyses on clinical psychological interventions (2000–2020). *Behav. Res. Methods* 54, 334–349. doi: 10.3758/s13428-021-01644-z
- Maassen, E., Van Assen, M. A. L. M., Nuijten, M. B., Olsson-Collentine, A., and Wicherts, J. M. (2020). Reproducibility of individual effect sizes in meta-analyses in psychology. *PLoS One* 15:e0233107. doi: 10.1371/journal.pone.0233107
- Manuele, S. J., Yap, M. B. H., Lin, S. C., Pozzi, E., and Whittle, S. (2023). Associations between paternal versus maternal parenting behaviors and child and adolescent internalizing problems: a systematic review and meta-analysis. *Clin. Psychol. Rev.* 105:102339. doi: 10.1016/j.cpr.2023.102339
- Mingebach, T., Kamp-Becker, I., Christiansen, H., and Weber, L. (2018). Meta-meta-analysis on the effectiveness of parent-based interventions for the treatment of child externalizing behavior problems. *PLoS One* 13:e0202855. doi: 10.1371/journal.pone.0202855
- Moher, D., Liberati, A., Tetzlaff, J., and Altman, D. G. The PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 6:e1000097. doi: 10.1371/journal.pmed.1000097
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., et al. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst. Rev.* 4:1. doi: 10.1186/2046-4053-4-1
- Mulrow, C. D. (1987). The medical review article: state of the science. *Ann. Intern. Med.* 106, 485–488. doi: 10.7326/0003-4819-106-3-485
- Murad, M. H., Asi, N., Alsawas, M., and Alahdab, F. (2016). New evidence pyramid. *Evid. Based Med.* 21, 125–127. doi: 10.1136/ebmed-2016-110401
- Nakagawa, S., and Lagisz, M. (2024). The ABC of academic writing: non-native speakers' perspective. *Trends Ecol. Evol.* 39, 307–310. doi: 10.1016/j.tree.2024.01.008
- Nelson, G., Park, S., Brafford, T., Heller, N. A., Crawford, A. R., and Drake, K. R. (2022). Reporting quality in math meta-analyses for students with or at risk of disabilities. *Except. Child.* 88, 125–144. doi: 10.1177/00144029211050851
- Nguyen, P.-Y., Kanukula, R., McKenzie, J. E., Alqaideem, Z., Brennan, S. E., Haddaway, N. R., et al. (2022). Changing patterns in reporting and sharing of review data in systematic reviews with meta-analysis of the effects of interventions: cross sectional meta-research study. *BMJ* 379:e072428. doi: 10.1136/bmj-2022-072428
- Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., et al. (2015). Promoting an open research culture. *Science* 348, 1422–1425. doi: 10.1126/science.aab2374
- Nuijten, M. B., and Epskamp, S. (2024). *statcheck*: Extract statistics from articles and recompute p-values. Available online at: <https://CRAN.R-project.org/package=statcheck> (November 26, 2024).
- Nuijten, M. B., Hartgerink, C. H. J., van Assen, M. A. L. M., Epskamp, S., and Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behav. Res. Methods* 48, 1205–1226. doi: 10.3758/s13428-015-0664-2
- Nuijten, M. B., and Polanin, J. R. (2020). “*statcheck*”: Automatically detect statistical reporting inconsistencies to increase reproducibility of meta-analyses. *Res. Synth. Methods* 11, 574–579. doi: 10.1002/jrsm.1408
- Nuijten, M. B., Van Assen, M. A. L. M., Hartgerink, C. H. J., Epskamp, S., and Wicherts, J. M. (2017). The validity of the tool “*statcheck*” in discovering statistical reporting inconsistencies. *PsyArXiv* [Preprint]. doi: 10.31234/osf.io/tcxaj
- Nuijten, M. B., and Wicherts, J. M. (2024). Implementing *statcheck* during peer review is related to a steep decline in statistical-reporting inconsistencies. *Adv. Methods Pract. Psychol. Sci.* 7:25152459241258945. doi: 10.1177/25152459241258945
- O'Shea, J., Jenkins, R., Nicholls, D., Downs, J., and Hudson, L. D. (2024). Prevalence, severity and risk factors for mental disorders among sexual and gender minority young people: a systematic review of systematic reviews and meta-analyses. *Eur. Child Adolesc. Psychiatry* 34, 959–982. doi: 10.1007/s00787-024-02552-1
- Olsen, J., Mosen, J., Voracek, M., and Kirchler, E. (2019). Research practices and statistical reporting quality in 250 economic psychology master's theses: a meta-research investigation. *R. Soc. Open Sci.* 6:190738. doi: 10.1098/rsos.190738
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science* 349:aac4716. doi: 10.1126/science.aac4716
- OpenAI (2024). ChatGPT (GPT-4). Available online at: <https://chatgpt.com/> (November 26, 2024).
- Page, M. J., Altman, D. G., Shamseer, L., McKenzie, J. E., Ahmadzai, N., Wolfe, D., et al. (2018). Reproducible research practices are underused in systematic reviews of biomedical interventions. *J. Clin. Epidemiol.* 94, 8–18. doi: 10.1016/j.jclinepi.2017.10.017
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021a). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Syst. Rev.* 10:89. doi: 10.1186/s13643-021-01626-4
- Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., et al. (2021b). PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews. *BMJ* 372, 1–26. doi: 10.1136/bmj.n160
- Page, M. J., Moher, D., Fidler, F. M., Higgins, J. P. T., Brennan, S. E., Haddaway, N. R., et al. (2021c). The REPRISSE project: protocol for an evaluation of REProducibility and replicability in syntheses of evidence. *Syst. Rev.* 10:112. doi: 10.1186/s13643-021-01670-0
- Page, M. J., Shamseer, L., Altman, D. G., Tetzlaff, J., Sampson, M., Tricco, A. C., et al. (2016). Epidemiology and reporting characteristics of systematic reviews of biomedical research: a cross-sectional study. *PLoS Med.* 13:e1002028. doi: 10.1371/journal.pmed.1002028
- Pellegrini, M., Pigott, T., Sutton Chubb, C., Day, E., Pruitt, N., and Scarbrough, H. F. (2024). Protocol for a meta-review on education meta-analyses: exploring methodological quality and potential significance for research use in practice. *Nord. J. Syst. Rev. Educ.* 2, 76–102. doi: 10.23865/njsre.v2.6169
- Peña, A. M., Ladd, C. C., Anderson, J. M., Torgerson, T., Hartwell, M., Johnson, B. S., et al. (2022). An analysis of the evidence underpinning the American urologic association clinical practice guidelines. *Urology* 161, 42–49. doi: 10.1016/j.urology.2021.12.019
- Polanin, J. R., Hennessy, E. A., and Tsuji, S. (2020). Transparency and reproducibility of meta-analyses in psychology: a meta-review. *Perspect. Psychol. Sci.* 15, 1026–1041. doi: 10.1177/1745691620906416
- Polanin, J. R., and 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 Syst. Rev.* 14, 1–36. doi: 10.4073/csr.2018.1
- Polanin, J. R., and Pigott, T. D. (2015). The use of meta-analytic statistical significance testing. *Res. Synth. Methods* 6, 63–73. doi: 10.1002/jrsm.1124
- Pustejovsky, J. E., and Tipton, E. (2022). Meta-analysis with robust variance estimation: expanding the range of working models. *Prev. Sci.* 23, 425–438. doi: 10.1007/s11211-021-01246-3
- Raad, J. M., Bellinger, S., McCormick, E., Roberts, M. C., and Steele, R. G. (2008). Brief report: reporting practices of methodological information in four journals of pediatric and child psychology. *J. Pediatr. Psychol.* 33, 688–693. doi: 10.1093/jpepsy/jsm130
- Rainkie, D. C., Abedini, Z. S., and Abdelkader, N. N. (2020). Reporting and methodological quality of systematic reviews and meta-analysis with protocols in diabetes mellitus type II: a systematic review. *PLoS One* 15:e0243091. doi: 10.1371/journal.pone.0243091
- Rethlefsen, M. L., Brigham, T. J., Price, C., Moher, D., Bouter, L. M., Kirkham, J. J., et al. (2024). Systematic review search strategies are poorly reported and not reproducible: a cross-sectional meta-research study. *J. Clin. Epidemiol.* 166:111229. doi: 10.1016/j.jclinepi.2023.111229
- Rethlefsen, M. L., Kirtley, S., Waffenschmidt, S., Ayala, A. P., Moher, D., Page, M. J., et al. (2021). PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. *Syst. Rev.* 10:39. doi: 10.1186/s13643-020-01542-z
- Romero, E., Álvarez-Voces, M., Díaz-Vázquez, B., and López-Romero, L. (2023). What works in the psychological treatment of child conduct problems? An umbrella review of meta-analytic studies. *Rev. Psicol. Clínica Con Niños Adolesc.* 10, 9–19. doi: 10.21134/rpcna.2023.10.1.1
- Sánchez De Ribera, O., Trajtenberg, N., and Christensen, L. S. (2020). Evaluating the quality of meta-analytic reviews using the AMSTAR-2: a systematic review of meta-analytic reviews regarding child sexual abuse interventions. *Child Abuse Negl.* 104:104463. doi: 10.1016/j.chiabu.2020.104463
- Sandoval-Lentisco, A., López-Nicolás, R., Tortajada, M., López-López, J. A., and Sánchez-Meca, J. (2024). Transparency in cognitive training meta-analyses: a meta-review. *Neuropsychol. Rev.* doi: 10.1007/s11065-024-09638-2
- Sandoval-Lentisco, A., Tortajada, M., Lopez-Nicolas, R., López-López, J. A., Wagenmakers, E.-J., Sánchez-Meca, J., et al. (2025). Preregistration of psychology meta-analyses: a cross-sectional study of prevalence and practice. *Adv. Methods Pract. Psychol. Sci.* 8, 1–12. doi: 10.1177/25152459241300113
- Scarlett, H., Moirangthem, S., and Van Der Waerden, J. (2024). The impact of paternal mental illness on child development: an umbrella review of systematic reviews and meta-analyses. *Eur. Child Adolesc. Psychiatry* 33, 3693–3706. doi: 10.1007/s00787-023-02261-1
- Schlesinger, A., Sengupta, S., Marx, L., Hilt, R., Martini, D. R., DeMaso, D. R., et al. (2023). Clinical update: collaborative mental health care for children and adolescents in pediatric primary care. *J. Am. Acad. Child Adolesc. Psychiatry* 62, 91–119. doi: 10.1016/j.jaac.2022.06.007
- Sehmbi, H., Retter, S., Shah, U. J., Nguyen, D., Martin, J., and Uppal, V. (2023). Methodological and reporting quality assessment of network meta-analyses in anesthesiology: a systematic review and meta-epidemiological study. *Can. J. Anesth.* 70, 1461–1473. doi: 10.1007/s12630-023-02510-6
- Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., et al. (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ* j4008:j4008. doi: 10.1136/bmj.j4008
- Siddaway, A. P., Wood, A. M., and 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. *Annu. Rev. Psychol.* 70, 747–770. doi: 10.1146/annurev-psych-010418-102803
- Siegel, M., Eder, J. S. N., Wicherts, J. M., and 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. *J. Appl. Psychol.* 107, 2013–2039. doi: 10.1037/apl0000991

Sifers, S. K. (2002). Reporting of demographics, methodology, and ethical procedures in journals in pediatric and child psychology. *J. Pediatr. Psychol.* 27, 19–25. doi: 10.1093/jpepsy/27.1.19

Silva, F., Rodrigues Amorim Adegboye, A., Lachat, C., Curioni, C., Gomes, F., Collins, G. S., et al. (2023). Completeness of reporting in diet- and nutrition-related randomized controlled trials and systematic reviews with meta-analysis: protocol for 2 independent meta-research studies. *JMIR Res. Protoc.* 12:e43537. doi: 10.2196/43537

Solmi, M., Radua, J., Olivola, M., Croce, E., Soardo, L., Salazar De Pablo, G., et al. (2022). Age at onset of mental disorders worldwide: large-scale meta-analysis of 192 epidemiological studies. *Mol. Psychiatry* 27, 281–295. doi: 10.1038/s41380-021-01161-7

Steil, A. V., Dias, N. M., Lopes, F. M., Silva, M. L. B. D., Bousfield, A. B. D. S., and De Luca Canto, G. (2022). Reporting characteristics of systematic reviews in psychology: a scoping review. *J. Health Psychol.* 27, 2964–2981. doi: 10.1177/13591053221074592

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance - or vice versa. *J. Am. Stat. Assoc.* 54, 30–34.

Stevenson, M., and Sergeant, E. (2025). epiR: Tools for the analysis of epidemiological data. Available online at: <https://CRAN.R-project.org/package=epiR> (November 26, 2024).

Streck, S., McIntire, R., Canale, L., Michael Anderson, J., Hartwell, M., Torgerson, T., et al. (2022). An evaluation of evidence underpinning management recommendations in tobacco use disorder clinical practice guidelines. *Nicotine Tob. Res.* 24, 847–854. doi: 10.1093/ntr/ntac012

Tricco, A. C., Cogo, E., Page, M. J., Polisen, J., Booth, A., Dwan, K., et al. (2016). A third of systematic reviews changed or did not specify the primary outcome: a PROSPERO register study. *J. Clin. Epidemiol.* 79, 46–54. doi: 10.1016/j.jclinepi.2016.03.025

Tricco, A. C., Lillie, E., Zarin, W., O'Brien, K. K., Colquhoun, H., Levac, D., et al. (2018). PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann. Intern. Med.* 169, 467–473. doi: 10.7326/M18-0850

Tsujimoto, Y., Tsujimoto, H., Kataoka, Y., Kimachi, M., Shimizu, S., Ikenoue, T., et al. (2017). Majority of systematic reviews published in high-impact journals neglected to register the protocols: a meta-epidemiological study. *J. Clin. Epidemiol.* 84, 54–60. doi: 10.1016/j.jclinepi.2017.02.008

Tung, I., Hipwell, A. E., Grosse, P., Battaglia, L., Cannova, E., English, G., et al. (2024). Prenatal stress and externalizing behaviors in childhood and adolescence: a systematic review and meta-analysis. *Psychol. Bull.* 150, 107–131. doi: 10.1037/bul0000407

Türk, S., Korfmacher, A.-K., Gerger, H., Van Der Oord, S., and Christiansen, H. (2023). Interventions for ADHD in childhood and adolescence: a systematic umbrella review and meta-meta-analysis. *Clin. Psychol. Rev.* 102:102271. doi: 10.1016/j.cpr.2023.102271

Van Altena, A. J., Spijker, R., and Olabarriaga, S. D. (2019). Usage of automation tools in systematic reviews. *Res. Synth. Methods* 10, 72–82. doi: 10.1002/jrsm.1335

Wayant, C., Page, M. J., and Vassar, M. (2019). Evaluation of reproducible research practices in oncology systematic reviews with meta-analyses referenced by National Comprehensive Cancer Network Guidelines. *JAMA Oncol.* 5, 1550–1555. doi: 10.1001/jamaoncol.2019.2564

Weber, L., Kamp-Becker, I., Christiansen, H., and Mingeback, T. (2019). Treatment of child externalizing behavior problems: a comprehensive review and meta-meta-analysis on effects of parent-based interventions on parental characteristics. *Eur. Child Adolesc. Psychiatry* 28, 1025–1036. doi: 10.1007/s00787-018-1175-3

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

Wiggins, B. J., and Christopherson, C. D. (2019). The replication crisis in psychology: an overview for theoretical and philosophical psychology. *J. Theor. Philos. Psychol.* 39, 202–217. doi: 10.1037/teo0000137

Wong, J. S., and Bouchard, J. (2023). Do meta-analyses of intervention/prevention programs in the field of criminology meet the tests of transparency and reproducibility? *Trauma Violence Abuse* 24, 1522–1542. doi: 10.1177/15248380211073839

World Health Organization (1993). The ICD-10 classification of mental and behavioural disorders: Diagnostic criteria for research. Geneva: World Health Organization.

World Health Organization (2019). International classification of diseases for mortality and morbidity statistics. 11th Edn. Geneva: World Health Organization.

World Health Organization (2024). 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. Available online at: <https://iris.who.int/bitstream/handle/10665/376852/9789240092198-eng.pdf?sequence=1> (November 26, 2024).

Zbukvic, I., McKay, S., Cooke, S., Anderson, R., Pilkington, V., McGillivray, L., et al. (2024). Evidence for targeted and universal secondary school-based programs for anxiety and depression: an overview of systematic reviews. *Adolesc. Res. Rev.* 9, 53–73. doi: 10.1007/s40894-023-00211-1