

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/131016/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Sánchez de Ribera, Olga, Trajtenberg, Nicolás and Christensen, Larissa S. 2020. Evaluating the quality of meta-analytical reviews using the AMSTAR-2: A systematic review of meta-analytical reviews regarding child sexual abuse interventions. *Child Abuse and Neglect* 104 , 104463. 10.1016/j.chiabu.2020.104463

Publishers page: <http://dx.doi.org/10.1016/j.chiabu.2020.104463>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Evaluating the quality of meta-analytical reviews using the AMSTAR-2: A systematic review
of meta-analytical reviews regarding child sexual abuse interventions

Abstract

Background: Effects of treatment for child sexual abuse (CSA) victims have important implications. Assessing Risk of Bias (RoB) is a vital step to inform interpretations of treatment effects for these victims. The AMSTAR-2 offers a comprehensive critical appraisal, allowing users to distinguish high quality reviews.

Objective: The aim of this article is two-fold: 1) to provide an up-to-date systematic review of treatment program meta-analytical reviews on interventions for CSA victims; and 2) to evaluate the quality of meta-analytical reviews using the AMSTAR-2. This is the first systematic review to examine the quality of meta-analyses on the effectiveness of CSA interventions using the AMSTAR-2.

Method: Eight electronic databases were searched for articles published up to April 2019. Meta-analytical reviews that assessed the effectiveness of any treatment modality for sexually abused children and adolescents (up to 18 years old) were considered. Outcome measures included physical and mental symptoms, and disorders, measured through validated instruments. Of 2,794 articles, nine meta-analyses met the eligibility criteria. There was a variety of interventions, including: trauma-focused cognitive-behavioral therapy (CBT), psychodrama, play therapy, and eclectic interventions. The most common outcomes measured were post-traumatic stress disorder/trauma, externalizing, internalizing, and sexualized behaviors.

Results: Although effect sizes were moderately significant, with treatment having a positive effect, all meta-analyses showed a high RoB.

Conclusions: To use the best available evidence in clinical decision-making for CSA victims, reviewers should conduct meta-analyses that employ RoB tools.

Keywords: child sexual abuse; treatment; meta-analysis; systematic review; quality assessment; cognitive-behavioral therapy.

Introduction

Child sexual abuse (CSA) is a global public health problem (World Health Organization [WHO], 2004). Several negative factors frequently documented in the literature on the impacts of CSA on children and adolescents include: poor physical health; disrupted emotional wellbeing; poor mental health; internalizing behaviors; externalizing behaviors; disrupted interpersonal relationships; socioeconomic effects; and vulnerability to re-victimization (see Fisher, Goldsmith, Hurcombe, & Soares, 2017; Harvey & Taylor, 2010; Lev-Wiesel, 2008; van Toledo & Seymour, 2013). Interventions that attempt to address the impacts of CSA include: cognitive-behavioral therapy (CBT); exposure therapy; psychodynamic therapy; narrative therapy; supportive counseling; and eye-movement desensitization and reprocessing. For CSA victims, these interventions have mainly focused on reducing trauma (Trask, Walsh, & DiLillo, 2011). CBT has been shown to have the greatest efficacy reducing post-traumatic stress disorder (PTSD) symptoms (e.g., Cohen, Deblinger, Mannarino, & Steer, 2004). Although some preliminary conclusions have been achieved regarding the effectiveness of treatment, several methodological limitations have been highlighted across studies, including Risk of Bias (RoB) (Macdonald et al., 2012). Disagreement across the effectiveness of these treatments comes from the nature itself; CSA is not a syndrome or disorder, rather, it is the experience of CSA which may lead to syndromes and disorders (Finkelhor & Berliner, 1995).

In translating the research evidence into practical programs, it is critical to ensure that the intervention components (i.e., behavior change techniques and strategies) and characteristics (e.g., setting, delivery mode, and intervention provider) that are most strongly associated with effectiveness are included. Evidence-based agenda has a direct impact on the decision-making process of practitioners and policy makers, urging them to seek empirical evidence of intervention effects (Littell, 2005). Empirical evidence is generally summarized using the statistical technique called a meta-analysis. Several standards exist to assist the

production of meta-analytical reviews such as The Cochrane Collaboration, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA), and the Campbell Collaboration (for others see Banzi et al., 2018). Despite the meta-analysis being considered to improve precision, minimize bias (Higgins & Green, 2011), and is considered the best source of evidence (Moher, Tetzlaff, Tricco, Sampson, & Altman, 2007), it does not provide a valid estimation if the quality of the studies included in the meta-analysis are low (La Torre et al., 2006). As the aim of science is to garner valid, truthful knowledge, the emphasis on meta-analyses should be on the quality assessment of the studies (La Torre et al., 2006).

To interpret results appropriately, several quality instruments have been developed to assess the study design and implementation included in systematic reviews (Farrington, 2003; and for a review see Olivo et al., 2008). Moreover, not all systematic reviews are as systematic as they should be, thus decreasing their reliability (Moher et al., 2007). Analyses of overviews have shown limited rigor (Pieper, Buechter, Jerinic, & Eikermann, 2012) and there is limited methodological guidance available (Pollock, Fernandes, Becker, Featherstone, & Hartling, 2016), making the evidence-based decision process in healthcare difficult, conflicting, and complicated (Kung et al., 2010). Although the number of overviews of reviews are increasing, methods for conducting overviews are still in their infancy (Lunny, Brennan, McDonald, & McKenzie, 2017). No clear guidelines for overviews had been developed unlike those for meta-analyses and systematic reviews. Recently, rigorous guidelines were developed to guide reviewers to conduct their reviews (see Lunny, Brennan, McDonald, & McKenzie, 2017, 2018; Smith, Devane, Begley, & Clarke, 2011). The current study followed such guidelines.

Methodological quality assessment refers to whether the systematic review has been performed and reported according to the current standards (i.e., Cochrane and PRISMA) (Gates et al., 2018). While methodological quality assessment of the included meta-analytical reviews is important, there is conflicting guidance regarding the instrument that should be used (Pollock

et al., 2016; Shea et al., 2009). Several instruments have been developed and validated to evaluate the methodological quality of systematic reviews and to make the process more efficient (see Kung et al., 2010). The validated and reliable tool most commonly used by researchers is A MeaSurement Tool to Assess systematic Reviews (AMSTAR) (Pollock et al., 2016). This 11-item checklist covers methods used in the different stages of a systematic review. The AMSTAR has been reported as having good measurement properties in terms of validity, reliability, and applicability (Pieper, Buechter, Li, Prediger, & Eikermann, 2015; Shea et al., 2009) and has been utilized and widely accepted by policy institutions and professional healthcare associations (Kung et al., 2010). Recently, due to criticisms (see Shea et al., 2017), the AMSTAR-2 has been developed, which consists of 16 items to evaluate reviews that include both randomized controlled trials (RCTs) and non-RCTs. The items include the selection of the study designs, search strategy, justifications of exclusion criteria, RoB, and source of funding.

RoB in a systematic review refers to the internal validity, that is, whether the results are free from bias (Higgins et al., 2011). False positive conclusions (or false negatives) can result if less rigorous studies in the review overestimate (or underestimate) the effect of an intervention (Detsky, Naylor, O'Rourke, McGeer, & L'Abbé, 1992). It is important that reviewers assess the methodological quality of the individual studies included in the review, including the execution and design to identify the RoB (Higgins et al., 2011). The concepts of RoB and quality assessment differ in the sense that poor quality does not always imply a high RoB (Gates et al., 2018). Some sources of bias include selection bias (biased allocation to interventions), attrition bias (systematic differences in withdrawals across groups and how incomplete outcome data is handled), and detection bias (outcome assessors knowledge of the allocated interventions) (Higgins et al., 2011).

The Risk Of Bias In Systematic reviews (ROBIS, Whiting et al., 2016) is a newly developed tool that assesses the level of bias presented in the systematic review across four domains: study eligibility criteria (whether the criteria is clear, prespecified, and suitable to the review question); identification and selection of studies (whether a sensitive search has been used to identify as many eligible studies as possible); data collection and study appraisal (whether data collection is rigorous and has involved a structured, piloted data collection form and whether validity has been assessed using suitable criteria); and synthesis and findings (if data has been combined, whether appropriate approaches have been used). The final phase of the tool considers the RoB for the systematic review as a whole (this includes whether the interpretation of findings factors in all concerns identified across the four domains) (Whiting et al., 2016). Possible judgements to the questions for each domain are: low, high, or unclear. Pieper, Puljak, González-Lorenzo, and Minozzi (2019) compared the AMSTAR-2 and the ROBIS tools in terms of validity, reliability, and applicability and found no differences between them. They concluded that both instruments can be applied to assess the methodological quality and RoB in systematic reviews including both RCTs and non-RCTs. However, while ROBIS exclusively measures RoB, the AMSTAR-2 assesses both the methodological quality of the systematic review (which can impact on findings) and RoB, with specific items to tackle the RoB of the included studies (Shea et al., 2017). Further, compared with ROBIS, the AMSTAR-2 has been found to be clearer, more specific, and simpler (Pieper et al., 2019).

In 2015, Benuto and O'Donohue published a review of meta-analyses on CSA interventions, which identified seven meta-analyses. They highlighted methodological issues such as the different inclusion criteria across the reviews. It was mainly restricted to published studies, raising concerns about the file drawer problem.¹ Also, there was a large heterogeneity

¹ This is also referred to as publication bias: where statistically significant findings are more likely to be published than results that are not significant.

of variability across the samples and interventions within each meta-analysis, raising concerns about the veracity of the effect size reported (i.e., Type 2 error). They found different evidence of beneficial effects for treatment which ranged from small to large. Benuto and O'Donohue (2015) noted the most effective interventions were those administered in an agency setting and with long duration. Play therapy was effective for social functioning problems and CBT was superior compared with other interventions; however, the effects were only moderate to treat behavior problems, PTSD, self-concept, and caregiver outcomes. Finally, it was unclear what format was more effective (i.e., individual, group, or family). However, no quality assessment of the individual reviews was performed, even though one meta-analysis did so (Macdonald et al., 2012) which reported low quality of the single studies, study designs, and interventions.

The aim of this article is two-fold: 1) to provide an up-to-date systematic review of treatment program meta-analytical reviews concerning interventions for victims of child sexual abuse; and 2) to evaluate the quality of available meta-analytical reviews using the AMSTAR-2. To our knowledge, this will be the first study to assess the quality of meta-analyses on the effectiveness of treatment programs of CSA utilizing the AMSTAR-2. Doing so, will allow us to illustrate the use of the tool in this discipline. While systematic reviews in healthcare are considered to be the most reliable evidence (Higgins et al., 2011), they are subject to biases (Shea et al., 2017). The AMSTAR-2 offers a comprehensive critical appraisal, allowing users to distinguish high quality reviews (Shea et al., 2017). This study should be a valuable resource for researchers in the healthcare field, individuals interested in meta-analytical review standards as well as research and practice professionals in the CSA field.

Methods

Although there are clear guidelines for reporting meta-analytical reviews (see PRISMA, Moher et al., 2009; see The Cochrane Collaboration Handbook; see the Campbell Collaboration), guidance for reporting overviews of systematic reviews and meta-analyses is

scarce (Lunny et al., 2018). The overview in the current study is reported according to the recently developed rigorous guidelines (see Lunny et al., 2017, 2018; Smith et al., 2011).

Search methods. Systematic literature searches were conducted to identify meta-analyses on the effectiveness of CSA interventions. Eight online databases, including Web of Science, Google Scholar, Scopus, PsychINFO, Cochrane Library, PubMed, DARE, and WHO Library were searched using Boolean operators and the following search and Mesh terms in the title and the abstract: “systematic review” OR “meta-analysis”; AND “child*” OR “sexual abuse” OR “child sexual abuse” OR “child sexual victimization”; AND “intervention” OR “treatment” OR “therap*” OR “cognitive-behavior*” OR “play therapy” OR “art therapy” OR “group therapy” OR “family therapy” AND “randomized control* trial”. These searches covered the period from all years to April 2019. Furthermore, digital searches of journals in the area (i.e., *Clinical Psychology Review*; *Child Abuse and Neglect*; *Child Maltreatment*; *Journal of Abnormal Child Psychology*; *Journal of Child Sexual Abuse*) during this period were also conducted. Reference lists of selected reviews were searched. The search was limited to systematic reviews and meta-analyses. These searches returned 2,794 results, the abstracts of which were screened for eligibility.

Eligibility criteria. We developed an a priori unpublished protocol. In general, we included unpublished and published meta-analytical reviews concerning interventions of CSA. The following inclusion criteria were employed to identify eligible studies for the present review:

1. *Type of reviews:* only systematic reviews with a meta-analysis were included. Comprehensive, systematic, and critical reviews were excluded because we were interested on the treatment effect (i.e., effect size) to inform policy and practice. Primary individual studies were also excluded. If a systematic review was updated (e.g., Macdonald et al., 2006, 2012), the last version was included (Macdonald et al., 2012).

2. *Type of participants:* meta-analytical reviews that used trials involving human subjects who were victims of sexual abuse were included. Other types of maltreatment such as physical, emotional abuse, neglect, traumatic events such as individual/mass, intentional/unintentional, or manmade/natural traumatic events were excluded. Age was restricted to children and adolescents (up to 18 years of age) but sex was not restricted. Interventions carried out on adult survivors of sexual abuse were excluded.
3. *Type of interventions:* meta-analytical reviews of effects of any of the following therapies were included: cognitive-behavioral (CBT), abuse-specific (trauma), pharmacological, exposure-based, psychodynamic, narrative, supportive counseling, eye-movement desensitization and reprocessing, play, art, and any other type of conventional or unconventional/alternative therapies. These therapies were not mutually exclusive. For example, a person could receive CBT and abuse-specific group therapy. No restrictions for modalities (i.e., individual, group, and family), dosage, intensity, or duration of the intervention were applied (for a description of these therapies, see Saunders, Berliner, & Hanson, 2003). School based education programs were excluded due to a recent systematic review and meta-analysis which included a RoB assessment (Walsh, Zwi, Woolfenden, & Shlonsky, 2015).
4. *Type of comparators:* placebo, no treatment, treatment-as-usual, or waitlist control groups were included.
5. *Type of outcomes:* any review that included studies that reported validated measures of PTSD/trauma, externalizing symptoms (i.e., ADHD, Oppositional Defiant Disorder/Conduct Disorder, sexual behavior problems, and aggression), emotional wellbeing, mental health and internalizing symptoms (i.e., anxiety, depression), and interpersonal relationships were included.

6. *Review characteristics:* the meta-analytical review could include unpublished studies along with published studies in English and Spanish in peer-reviewed journals. Book chapters and conference abstracts were excluded.

Study selection. The first two authors independently reviewed the title, abstracts, and full text for their potential inclusion against the eligibility criteria. Any review appearing to meet the inclusion criteria based on the abstract was retrieved as a full-text article. No disagreement was reported.

Data extraction and coding. The first author extracted data using a form adapted from Smith et al. (2011, p.5) on the effectiveness of interventions and on the relationship of effectiveness and intervention components. These were as follows: authors (publication year), aim and participants, search strategy, number of included studies, total number of participants, RoB, intervention type, outcome, design, effect size (after intervention and at follow up), and moderator analysis. The data extraction template was checked by the second author with reference to the full text of the article. No disagreement was reported.

The two raters had different levels of expertise in conducting systematic reviews. One rater had performed 7 meta-analyses and has knowledge on quality assessment; and the other rater has experience reading meta-analyses and had performed one meta-analysis. Before starting the assessment, both raters read the AMSTAR-2's guidelines and discussed those that were unclear until an agreement was reached. Finally, a pilot quality assessment with a meta-analysis was performed and discussed between the two raters to reduce bias.

Statistical analyses. All data was collected in Excel sheets. The first two authors summarized the descriptive characteristics of the meta-analyses using frequencies. Cohen's Kappa (κ) coefficients were calculated as measures of inter-rater reliability (IRR) between the first two authors for each study, as well as for each question. The cut-offs for Kappa values were: <0.20 = no agreement; 0.21 to 0.39 = minimal; 0.40 to 0.59 = weak; 0.60 to 0.79 =

moderate; 0.80 to 0.90 = strong; and > 0.90 = almost perfect agreement (McHugh, 2012). Some authors have pointed out the potential bias for Kappa coefficients (see Hallgren, 2012). In addition, the intraclass correlation coefficient (ICC) was calculated. The cut-offs for the ICC values were: <0.40 = poor; 0.40 to 0.59 = fair; 0.60 to 0.74 = good; and 0.75 to 1 = excellent (Cicchetti, 1994). Statistical analyses were performed using the SPSS version 21.0 (IBM, Corp., 2012).

Quality assessment of meta-analyses with AMSTAR-2 instrument. The methodological quality of each meta-analysis was independently assessed by the first two authors using the AMSTAR-2 checklist (for detail, visit <https://amstar.ca/index.php>). The quality of the reviews was calculated using the checklist form (see https://amstar.ca/Amstar_Checklist.php). The possible answers on the checklist are “Yes”, “No”, and “Partial Yes”. The overall score is calculated by giving 1 point to “Yes” (i.e., criteria was met) and 0 points to “No” (i.e., criteria was not met) and “Partial Yes” answers (i.e., criteria was partially met). However, the authors strongly recommend the identification of critical domains (i.e., items 2, 4, 7, 9, 11, 13, 15) rather than an overall score because “an overall score may disguise critical weaknesses that should diminish confidence in the results of a systematic review” (Shea et al., 2017, p.6)².

The AMSTAR-2 provides four levels of quality: high (none or one non-critical weakness), moderate (more than one non-critical weakness), low (one critical flaw with or without non-critical weaknesses) and critically low (more than one critical flaw with or without non-critical weaknesses). Disagreement between the first two authors was reported, especially for items 9b (whether non-RCTs used a comprehensive instrument) and 11 (whether

² Other researchers have highlighted that the focus should be on the individual methodological aspects, allowing for identification of key domains of concealment (e.g., dropout, handing of withdrawals) (Jüni, Witschi, Bloch, & Egger, 1999).

appropriate methods were used for the statistical combination of results). These discrepancies were resolved through discussion.

Deviation from the protocol. In our prior unpublished protocol, we said that we would include meta-analytical reviews that examined child maltreatment but reported effect sizes for sexual abuse independently. However, it was decided that we would only include CSA studies as it was complex to use the AMSTAR-2 tool for meta-analyses that included other outcomes not included in those specific for CSA. Furthermore, we included only meta-analytical reviews that considered a control group. However, this criterion reduced the number of meta-analyses considerably, so we decided to include meta-analyses that included single pre- and post-test studies.

Results

Search results. Figure 1 shows the flowchart of this overview selection. We found 2,794 potentially relevant studies. After excluding duplicate articles as well as screening titles and abstracts, 50 studies were eligible for full-text review (see reasons for exclusion in Table 2S). Finally, nine meta-analyses were included in the present study. All studies were published in English.

[INSERT FIGURE 1 HERE]

Review characteristics. Nine meta-analyses comprising a maximum sample size of 1,839 participants about a broad range of interventions for sexually abused children and adolescents between 1997 and 2014 years were identified (see Table 1). Eight were published; only one unpublished PhD dissertation was located (Tehrani, 2014).

[INSERT TABLE 1 HERE]

The number of studies included in each meta-analysis varied, ranging from 7 to 58 studies. The attrition rate was reported in only three meta-analyses. Only one article was published in the Campbell Collaboration (Macdonald et al., 2012). Five meta-analyses

investigated multiple interventions (CBT and CBT-related, drama, play, supportive, psychodynamic, humanistic, narrative, mixture, and other), and modalities (individual, group, and family), with four common primary outcomes in all reviews: internalizing behaviors, externalizing behaviors, sexualized behaviors, and post-traumatic stress symptoms (see Table 2). Effect sizes for these outcomes were moderately significant, that is, treatment had a positive effect. Three studies exclusively included CBT therapy. Regarding the design, one review only included quasi-experimental designs and RCTs, five reviews included single group (pre- and post-test) in combination with quasi-experimental and RCTs, and two meta-analyses exclusively included a pre- and post-design. All meta-analyses included in our study conducted moderator analyses except for two, due to the small number of studies included in these meta-analyses. The main moderators included across the meta-analyses were: sex and age of participants, informant (e.g., teacher, parent, or child), outcome type, study design, type, setting, and intensity of intervention (see Table 2).

[INSERT TABLE 2 HERE]

Results using AMSTAR-2. AMSTAR-2 scores are shown in Table 3 and the justification statements for AMSTAR-2 are presented in Table 1S (Supplement Material). As a general score is not recommended (Shea et al., 2017), Table 4 shows the results of critical domains. Eight meta-analyses did not adhere to a priori well-designed protocols (Question 2, Q2). None used a comprehensive literature search strategy (Q4) and six reported some features of a comprehensive literature search (Q4). Two included a list of excluded studies and justified the exclusions (Q7). Only one meta-analysis assessed RoB in the individual studies using the Cochrane RoB instrument for RCTs and two partially assessed the RoB (Q9a); that is, they developed their own quality tool. However, five meta-analyses did not assess RoB (Q9a). None of the reviews that included non-RCTs (seven) used a comprehensive instrument to assess RoB (Q9b). The four meta-analyses that included RCTs used appropriate methods for the statistical

combination of results (Q11a), but of the studies that included non-RCTs (Q11b), none used appropriated methods for statistical combination. The results of RoB assessment was not discussed in seven studies (Q13). Finally, five of the reviews did not perform publication bias analyses and discussed its impact (Q15), whereas four of the reviews did.

[INSERT TABLE 3 HERE]

[INSERT TABLE 4 HERE]

Inter-rater reliability. The overall inter-rater agreement for scoring the meta-analyses was strong (ICC = 0.852, [95% CI: 0.73, 0.93], $p < .001$). Considering each meta-analysis, the Corcoran and Pillai (2008) obtained the lowest agreement between raters ($\kappa = 0.43$, $p < .001$), whereas Hetzel-Riggin, Brausch, and Montgomery (2007), Peltonen (2014), Sánchez-Meca, Rosa-Alcazar, and Lopez-Soler (2011), and Trask et al. (2011) obtained total level agreement ($\kappa = 1$, $p < .001$) (see Table 3). Perfect agreement between raters was observed for Q1, Q2, Q3, Q4, Q5, Q9b, Q10, Q11b, Q12, Q13, Q15, and Q16 ($\kappa = 1$, $p < .01$). For Q6 ($\kappa = 0.64$, $p < .05$), Q8 ($\kappa = 0.84$, $p < .001$), and Q9a ($\kappa = 0.85$, $p < .001$) the agreement was substantial. The lowest levels of agreement were obtained for Q11a ($\kappa = 0.58$, $p < .01$) and Q7 ($\kappa = 0.44$, $p < .05$), indicating only a slight agreement. Finally, no agreement was observed for Q14 ($\kappa = -0.10$, $p = 0.74$).

Discussion

To the best of our knowledge, this is the first systematic review to examine the quality of meta-analyses on the effectiveness of CSA interventions using the AMSTAR-2, highlighting the use of the tool in this discipline. Although effect sizes were moderately significant (treatment had a positive effect), all meta-analyses showed a high RoB and poor methodological quality. Due to the potential of meta-analyses informing decision-making for practitioners and policy makers, it is crucial that they are conducted and reported using high quality methodology. Below we discuss our findings prior to suggesting the need for all meta-

analyses to adhere to systematic review protocols (such as the CONSORT-SPI protocol). Finally, we offer some recommendations for improvement to the AMSTAR-2.

One of the meta-analyses (Macdonald et al., 2012) reported alarming scores, specifically, a low quality of single studies, study designs, and interventions. Macdonald et al. (2012) was a Campbell Collaboration review, which follows structured guidelines for quality of reporting and RoB (using the modified Cochrane Collaboration Risk of Bias tool). Promisingly, this meta-analysis scored “Yes” on most of the items and it was the only meta-analysis to assess the RoB, however, Macdonald et al. (2012) did not provide justifications for language restrictions; did not report sources of funding for the studies included in the meta-analysis; could not assess the potential impact of the RoB in individual studies because all studies reported a high RoB (i.e., the score was “No”); and did not perform a statistical test for publication bias. It seems that the Campbell Collaboration could benefit by including some new items in the systematic review protocols.

Recently, based on the Consolidated Standards of Reporting Trials (CONSORT 2010) statement, a group of scientists developed the Consolidated Standards of Reporting Trials for Social and Psychological Interventions (CONSORT-SPI) statement to report these interventions transparently. Based on the evidence, they added new items (compared to the CONSORT 2010) that can affect the results including: the online access to the trial protocol, sources of funding and the role of funders in the trial, stakeholder involvement, and incentives offered as part of the trial (see Grant et al., 2018 for more detail). The introduction of CONSORT-SPI would mitigate discrepancies between raters as the adherence and transparency is reported by the authors of the respective meta-analyses. Through introducing such a tool for systematic review protocols, not only quality assessment studies (like this one) would be unnecessary, but more importantly, the tool would allow for an easier decision

process for research and practice professionals in the CSA field. In turn, we suggest the Campbell Collaboration could integrate a similar tool into their systematic review protocols.

Several challenges have been reported for AMSTAR (Pollock, Fernandes, & Hartling, 2017), which have been considered for the development of AMSTAR-2. While the AMSTAR-2 was an efficient and clear tool to use, there are some areas where it could be enhanced. To provide one example, Harvey and Taylor's (2010) review was comprehensive and scored "Yes" for questions important to be considered (i.e., 7, 13, and 15) on the AMSTAR-2. On Q12, it scored "No", because the impact of the RoB in individual studies was not assessed but it reported a power analysis instead, which is also a very important threat to statistical conclusion validity (see Jackson & Turner, 2017; Turner, Bird, & Higgins, 2013). In Q4, they achieved a "Partial Yes" (similar to most of the reviewed meta-analyses in the current study), because publication restrictions were not provided. Moreover, Q8 was coded as "No" because they did not report comparators even though they reported all of the rest of the conditions. These results suggest that the inclusion of some additional options to "Yes" and "Partial Yes" on the AMSTAR-2 may remove a level of rigidity from responses.

Another recommendation concerns Q1 where it considers "Yes" for a comparison group. However, there were some meta-analyses which included both types of studies (i.e., with and without a comparison group), making it difficult to respond to this question dichotomously. In turn, we suggest the AMSTAR-2 provides the rater with an option for reviews which do not solely include comparison groups. Another area for enhancement is Q4; "Partial Yes" includes "*justified publication restrictions (e.g. language)*" (among other options). We agree that to be able to replicate the search we need to know which restrictions were applied. However, we are unsure to what extent the justification for language restrictions are important, especially when in most cases the assumed restriction is lack of resources to pay a translator and/or the team members are not bilingual. Perhaps it could be considered as

“optional (recommended)”. Finally, Q11 could include an option similar to Q9 (RCTs, non-randomized studies of interventions [NRSI]) (“Includes only NRSI” and “Includes only RCTs”, respectively) because the option “No meta-analysis included” resulted in some confusion.

Finally, we found it difficult to deal with items that generated some ambiguity. For instance, in Q12 the reviewer is expected to assess the impact of RoB within the individual studies of the meta-analysis. However, in some cases (e.g., Macdonald et al. 2012) it was not possible to perform an analysis to assess the impact of the RoB (all studies had a high RoB). Adding a “Not possible” response would overcome this limitation. Another example is assessing the effect of moderators. In some studies (e.g., Corcoran & Pillai, 2008) it was not possible to assess the effect of moderators because of the small number of studies included. Alarming, no agreement was observed between the raters on Q14 (whether heterogeneity was investigated). The guidance document states that “*there are many potential causes of heterogeneity... Both the PICO elements and the domains of bias listed in Item 9 should also be considered as important potential sources of heterogeneity*”. If a reviewer is familiar with meta-analyses, they will likely consider the statement of the scale in a broad sense (e.g., inclusion of moderator analyses), whereas if a reviewer is less familiar with meta-analyses and focus exclusively on the PICO elements and Q9, there will be disagreements across raters (similar to this study). Moreover, we think the guidance document is too strict when considering both PICO AND³ the domains bias listed in Q9. While the implementation of decision rules has helped to overcome some of the challenges in AMSTAR, it would be better to add specific answers (as mentioned above) to the AMSTAR-2 to decrease ambiguity and rigidity, while increasing the reliability of the tool.

³ Capital letters are added by authors of this study, but not by the authors of the AMSTAR-2 guidance document (Shea et al., 2017).

Limitations. This review is not without limitations. The current study did not include a large number of meta-analytical reviews so generalizability of the data is not possible. Also, while we provided an up-to-date systematic review of reviews, the most recent study was from 2014. As aforementioned, the AMSTAR-2's authors do not recommend an overall score, even though the electronic version of the form calculates a final score (see https://amstar.ca/Amstar_Checklist.php). Conversely, we think that a final score is useful for practitioners to better understand the results. As the AMSTAR-2 measures both methodological quality and RoB, we suggest it would be useful to have an overall score for each section. Finally, we restricted the search to Spanish and English languages and did not include databases that could have given access to studies from low and middle-income countries. However, we searched in Scielo, a database that includes studies from the Latin America region and could not locate any meta-analyses on this topic (only one systematic review from Brazil). Veenema, Thornton, and Corley (2015) conducted an integrative review concerning CSA in developing societies (i.e., perceptions, prevalence, risk factors, and barriers) but did not report interventions and searched in some databases similar to ours (i.e., PsycInfo, Web of Knowledge, PubMed). In turn, future systematic reviews on this subject may need to consider including lower quality evidence from developing societies (e.g., Habigzang et al., 2009, 2013; van Westrhenen, Fritz, Vermeer, Boelen, & Kleber, 2019).

Implications of results. As previous studies highlighted (Benuto & O'Donohue, 2015; Macdonald et al., 2012), there is a need for larger and more methodologically sound RCTs to be conducted on the effectiveness of interventions for CSA. While CBT had a few RCT trials, these RCTs reported a high RoB. Although some authors (e.g., Kim, Noh, & Kim, 2016; Sánchez-Meca et al., 2011) elucidate the importance of considering ethical issues when carrying out research into CSA, adopting RCTs using a treatment as usual comparison would be feasible such as the case of Danielson et al. (2012).

Recent systematic reviews have been published considering specific interventions such as group therapy (Miffitt, 2014) and psychosocial interventions (Kim et al., 2016). Researching specific types of interventions for specific outcomes can help to achieve more accurate results and reduce heterogeneity. According to the previous overview (Benuto & O'Donohue, 2015), there are 77 RCTs across the reviews. A meta-analysis including these studies is recommended to evaluate the RoB and the effect size. Further to this, it was surprising that the most recent meta-analysis in our review that fit the inclusion criteria was conducted over five years ago. We suggest the need for an updated study to provide research and practice professionals in the CSA field with the latest evidence-base. Finally, while the properties of ROBIS and AMSTAR-2 are similar (Pieper et al., 2019), more research is needed for further validation of the AMSTAR-2. Thus, future studies should use both tools to test possible differences in terms of validity, reliability, and applicability.

Conclusions

The AMSTAR-2 offers a comprehensive critical appraisal, which allows users to distinguish high quality reviews. The use of this tool was highlighted in the current systematic review of meta-analytical reviews for child sexual abuse interventions. Our review suggests CBT is a promising intervention for treating post-traumatic symptoms, internalizing, externalizing, and sexualized behaviors. However, no firm conclusions can be drawn for the existing evidence since the quality of research showed that the studies included reported a high RoB and low methodological quality. Despite the importance and magnitude of the topic, there is an alarming paucity in the research. More methodologically high-quality, large-scale randomized controlled trials are needed, along with the tool for systematic review protocols on this topic, as well as studies from developing societies.

Role of Funding Sources: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Contributors: The first two authors designed the study, wrote the protocol, and conducted the analysis. The third author provided intellectual input throughout the drafts. All authors contributed to the final manuscript.

Declaration of Interest: None.

References

- Banzi, R., Cinquini, M., Gonzalez-Lorenzo, M., Pecoraro, V., Capobussi, M., & Minozzi, S. (2018). Quality assessment versus risk of bias in systematic reviews: AMSTAR and ROBIS had similar reliability but differed in their construct and applicability. *Journal of Clinical Epidemiology*, *99*, 24–32. <https://doi.org/10.1016/j.jclinepi.2018.02.024>
- Benuto, L.T., & O'Donohue, W. (2015). Treatment of the sexually abused child: Review and synthesis of recent meta-analyses. *Children and Youth Services Review*, *56*, 52–60. <https://doi.org/10.1016/j.chidyouth.2015.06.009>
- Cicchetti, D.V. (1994) Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290.
- Cohen, J. A., Deblinger, E., Mannarino, A. P., & Steer, R. A. (2004). A multisite, randomized controlled trial for children with sexual abuse–related PTSD symptoms. *Journal of the American Academy of Child & Adolescent Psychiatry*, *43*(4), 393–402. <https://doi.org/10.1097/00004583-200404000-00005>.
- Corcoran, J., & Pillai, V. (2008). A meta-analysis of parent-involved treatment for child sexual abuse. *Research on Social Work Practice*, *18*(5), 453–464. <https://doi.org/10.1177/1049731507313980>
- Danielson, C. K., McCart, M. R., Walsh, K., De Arellano, M. A., White, D., & Resnick, H. S. (2012). Reducing substance use risk and mental health problems among sexually assaulted adolescents: A pilot randomized controlled trial. *Journal of Family Psychology*, *26*(4), 628– 635. <https://doi.org/10.1037/a0028862>
- Detsky, A. S., Naylor, C. D., O'Rourke, K., McGeer, A. J., & L'Abbé, K. A. (1992). Incorporating variations in the quality of individual randomized trials into meta-analysis. *Journal of clinical epidemiology*, *45*(3), 255–265.
- Farrington, D. P. (2003). Methodological quality standards for evaluation research. *Annals of American Academy of Political and Social Science*, *587*(1), 49–68. <https://doi.org/10.1177/0002716202250789>
- Finkelhor, D., & Berliner, L. (1995). Research on the treatment of sexually abused children: A review and recommendations. *Journal of the American Academy of Child & Adolescent Psychiatry*, *34*(11), 1408–1423. <http://doi.org/10.1097/00004583-199511000-00007>
- Fisher, C., Goldsmith, A., Hurcombe, R., & Soares, C. (2017). *The impacts of child sexual abuse: A rapid evidence assessment*. UK: Independent Inquiry into Child Sexual Abuse. Retrieved from: <https://www.iicsa.org.uk/key-documents/1534/view/iicsa-impacts-child-sexual-abuse-rapid-evidence-assessment-full-report-english.pdf>
- Gates, A., Gates, M., Duarte, G., Cary, M., Becker, M., Prediger, B., Vandermeer, B., Fernandes, R. M., Pieper, D., & Hartling, L. (2018). Evaluation of the reliability, usability, and applicability of

- AMSTAR, AMSTAR 2, and ROBIS: Protocol for a descriptive analytic study. *Systematic Reviews*, 7(1), 85-91. <https://doi.org/10.1186/s13643-018-0746-1>
- Grant, S., Mayo-Wilson, E., Montgomery, P., Macdonald, G., Michie, S., Hopewell, S., Moher, D. & the CONSORT-SPI Group. (2018). CONSORT-SPI 2018 Explanation and elaboration: Guidance for reporting social and psychological intervention trials. *Trials*, 19, 406-423. <https://doi.org/10.1186/s13063-018-2735-z>
- Habigzang, L.F., Damásio, B.F., & Koller, S.H. (2013). Impact evaluation of a cognitive behavioral group therapy model in Brazilian sexually abused girls. *Journal of Child Sexual Abuse*, 22(2), 173-90.: <http://doi.org/10.1080/10538712.2013.737445>
- Habigzang, L.F., Stroehrer, F.H., Hatzenberger, R., Cunha, R.C., Ramos Mda, S., & Koller, S.H. (2009). Cognitive behavioral group therapy for sexually abused girls. *Revista de Saude Publica*, 43(1), 70-8. <http://doi.org/10.1590/S0034-89102009000800011>
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23–34. doi:10.20982/tqmp.08.1.p023
- Harvey, S., & Taylor, J. (2010). A meta-analysis of the effects of psychotherapy with sexually abused children and adolescents. *Clinical Psychology Review*, 30, 517-535. <http://doi.org/10.1016/j.cpr.2010.03.006>.
- Hetzel-Riggin, M., Brausch, A., & Montgomery, B. (2007). A meta-analytic investigation of therapy modality outcomes for sexually abused children and adolescents: An exploratory study. *Child Abuse & Neglect*, 31, 125–141. <http://doi.org/10.1017/j.chiabu.2006.10.007>
- Higgins, J.P.T., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., ... & Sterne, J. A. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal*, 343(7829), 889-897. <https://doi.org/10.1136/bmj.d5928>
- Higgins, J.P.T., & Green, S. (Eds.), 2011. *Cochrane Handbook for Systematic Reviews of Interventions 5.1.0*. The Cochrane Library, vol. 5. John Wiley & Sons, Ltd, Chichester, UK. Available at <http://handbook.cochrane.org/>
- Jackson, D., & Turner, R. (2017). Power analysis for random-effects meta-analysis. *Research synthesis methods*, 8(3), 290-302. <https://doi.org/10.1002/jrsm.1240>
- Jüni, P., Witschi, A., Bloch, R., & Egger, M. (1999). The hazards of scoring the quality of clinical trials for meta-analysis. *Journal of the American Medical Association*, 282(11), 1054-1060. <https://doi.org/10.1001/jama.282.11.1054>
- Kim, S., Noh, D., & Kim, H. (2016). A summary of selective experimental research on psychosocial interventions for sexually abused children. *Journal of Child Sexual Abuse*, 25(5), 597-617, <https://doi.org/10.1080/10538712.2016.1181692>
- Kung, J., Chiappelli, F., Cajulis, O. O., Avezova, R., Kossan, G., Chew, L., & Maida, C. A. (2010). From systematic reviews to clinical recommendations for evidence-based health care: Validation

- of revised assessment of multiple systematic reviews (R-AMSTAR) for grading of clinical relevance. *The Open Dentistry Journal*, 4, 84-91. <https://doi.org/10.2174/1874210601004020084>
- La Torre, G., Chiaradia, G., Gianfagna, F., Boccia, S., De Laurentis, A., & Ricciardi, W. (2006). Quality assessment in meta-analysis. *Italian Journal of Public Health*, 3(2), 44-50. <https://doi.org/10.2427/5937>
- Lev-Wiesel, R. (2008). Child sexual abuse: A critical review of intervention and treatment modalities. *Children and Youth Services Review*, 30(6), 665-673. <https://doi.org/10.1016/j.chilyouth.2008.01.008>
- Littell, J. H. (2005). Lessons from a systematic review of effects of multisystemic therapy. *Children and Youth Services Review*, 27(4), 445-463. <https://doi.org/10.1016/j.chilyouth.2004.11.009>
- Lunny, C., Brennan, S. E., McDonald, S., & McKenzie, J. E. (2017). Toward a comprehensive evidence map of overview of systematic review methods: Paper 1-purpose, eligibility, search and data extraction. *Systematic Reviews*, 6(1), 231. <https://doi.org/10.1186/s13643-017-0617-1>
- Lunny, C., Brennan, S. E., McDonald, S., & McKenzie, J. E. (2018). Toward a comprehensive evidence map of overview of systematic review methods: Paper 2-risk of bias assessment; synthesis, presentation and summary of the findings; and assessment of the certainty of the evidence. *Systematic Reviews*, 7(1), 159. <https://doi.org/10.1186/s13643-018-0784-8>
- Macdonald, G., Higgins, J.P.T., Ramchandani, P. (2006). Cognitive-behavioural interventions for children who have been sexually abused. *Cochrane Database of Systematic Reviews*, 4(CD001930). <https://doi.org/10.1002/14651858.CD001930.pub3>
- Macdonald, G., Higgins, J.P.T., Ramchandani, P., Valentine, J.C., Bronger, L.P., Klein, P., Taylor, M. (2012). Cognitive-behavioural interventions for children who have been sexually abused. *Cochrane Database of Systematic Reviews*, 5(CD001930). <https://doi.org/10.1002/14651858.CD001930.pub3>
- McHugh, M.L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276-282. <https://hrcak.srce.hr/89395>
- Miffitt, L.A. (2014). State of the science: Group therapy interventions for sexually abused children. *Archives of Psychiatric Nursing*, 28(3), 174-179. <https://doi.org/10.1016/j.apnu.2013.09.004>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D.G., & the PRISMA Group (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA Statement. *PLoS Medicine*, 6(7), e1000097. <https://doi.org/10.1371/journal.pmed.1000097>
- Moher, D., Tetzlaff, J., Tricco, A. C., Sampson, M., & Altman, D. G. (2007). Epidemiology and reporting characteristics of systematic reviews. *PLoS Medicine*, 4(3), e78. <https://doi.org/10.1371/journal.pmed.0040078>
- Olivo, A.S., Macedo, L.G., Gadotti, I.C., Fuentes, J., Stanton, T., & Magee, D.J. (2008). Scales to assess the quality of randomized controlled trials: A systematic review. *Physical Therapy*, 88(2), 156-175. <https://doi.org/10.2522/ptj.20070147>

- Peltonen, K. (2014). Efficacy of interventions for sexually abused children -A literature review and meta-analysis of treatment of post-traumatic stress symptoms *International Journal of Clinical Psychiatry and Mental Health*, 2, 10-28.
- Pieper, D., Buechter, R., Jerinic, P., & Eikermann, M. (2012). Overviews of reviews often have limited rigor: A systematic review. *Journal of Clinical Epidemiology*, 65(12), 1267-1273. doi: <https://doi.org/10.1016/j.jclinepi.2012.06.015>
- Pieper, D., Buechter, R.B., Li, L., Prediger, B., & Eikermann, M. (2015). Systematic review found AMSTAR, but not R(evised)-AMSTAR, to have good measurement properties. *Journal of Clinical Epidemiology*, 68(5), 574-583. <https://doi.org/10.1016/j.jclinepi.2014.12.009>
- Pieper, D., Puljak, L., González-Lorenzo, M., & Minozzi, S. (2019). Minor differences were found between AMSTAR 2 and ROBIS in the assessment of systematic reviews including both randomized and nonrandomized studies. *Journal of Clinical Epidemiology*, 108, 26-33. <https://doi.org/10.1016/j.jclinepi.2018.12.004>
- Pollock, M., Fernandes, R. M., Becker, L. A., Featherstone, R., & Hartling, L. (2016). What guidance is available for researchers conducting overviews of reviews of healthcare interventions? A scoping review and qualitative meta summary. *Systematic Reviews*, 5(1), 190-204. <https://doi.org/10.1186/s13643-016-0367-5>
- Pollock, M., Fernandes, R. M., & Hartling, L. (2017). Evaluation of AMSTAR to assess the methodological quality of systematic reviews in overviews of reviews of healthcare interventions. *BMC Medical Research Methodology*, 17(1), 48-60. <https://doi.org/10.1186/s12874-017-0325-5>
- Reeker, J., Ensing, D., & Elliott, R. (1997). A meta-analytic investigation of group treatment outcomes for sexually abused children. *Child Abuse and Neglect*, 21(7), 669-680. [https://doi.org/10.1016/S0145-2134\(97\)00024-0](https://doi.org/10.1016/S0145-2134(97)00024-0)
- Sánchez-Meca, J., Rosa-Alcazar, A.I., & Lopez-Soler, C. (2011). The psychological treatment of sexual abuse in children and adolescents: A meta-analysis. *International Journal of Clinical and Health Psychology*, 11(1), 67-93. Retrieved from: <http://www.redalyc.org/articulo.oa?id=33715423005>
- Saunders, B.E., Berliner, L., & Hanson, R.F. (Eds.). (2003). Child physical and sexual abuse: Guidelines for treatment (Final Report: January 15, 2003). Charleston, SC: National Crime Victims Research and Treatment Center.
- Shea, B. J., Hamel, C., Wells, G. A., Bouter, L. M., Kristjansson, E., Grimshaw, J., Henry, D.A., & Boers, M. (2009). AMSTAR is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. *Journal of Clinical Epidemiology*, 62(10), 1013-1020. doi: <https://doi.org/10.1016/j.jclinepi.2008.10.009>
- 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. *British Medical Journal*, 358-j4008.
<https://doi.org/10.1136/bmj.j4008>
- Smith, V., Devane, D., Begley, C.M., & Clarke, M. (2011). Methodology in conducting a systematic review of systematic reviews of healthcare interventions. *BMC Medical Research Methodology*, 11(1), 15-20. <https://doi.org/10.1186/1471-2288-11-15>
- Statistical Package for Social Science (SPSS): IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.
- Tehrani, E. (2014). *Comparisons between Traditional and Contemporary Treatment Modalities for Sexually Abused Children and Adolescents: A Meta-Analysis* (Doctor of Philosophy of Psychology thesis, unpublished). The Chicago School, Chicago, United States.
- Trask, E., Walsh, K., & DiLillo, D. (2011). Treatment effects for common outcomes of child sexual abuse: A current meta-analysis. *Aggression and Violent Behavior*, 16(1), 6-19.
<http://doi.org/10.1016/j.avb.2010.10.001>
- Turner, R. M., Bird, S. M., & Higgins, J. P. (2013). The impact of study size on meta-analyses: Examination of underpowered studies in Cochrane reviews. *PloS ONE*, 8(3), e59202.
<https://doi.org/10.1371/journal.pone.0059202>
- van Toledo, A., & Seymour, F. (2013). Interventions for caregivers of children who disclose sexual abuse: A review. *Clinical Psychology Review*, 33(6), 772-781.
<https://doi.org/10.1016/j.cpr.2013.05.006>
- van Westrhenen, N., Fritz, E., Vermeer, A., Boelen, P., & Kleber, R. (2019). Creative arts in psychotherapy for traumatized children in South Africa: An evaluation study. *PLoS ONE* 14(2), e0210857. <https://doi.org/10.1371/journal.pone.0210857>
- Veenema, T.G., Thornton, C.P., & Corley, A. (2015). The public health crisis of child sexual abuse in low and middle income countries: An integrative review of the literature. *International Journal of Nursing Studies*, 52(4), 864–881. <https://doi.org/10.1016/j.ijnurstu.2014.10.017>
- Walsh, K., Zwi, K., Woolfenden, S., & Shlonsky, A. (2015). School-based education programs for the prevention of child sexual abuse. *Cochrane Database of Systematic Reviews*, 4(CD004380).
<https://doi.org/10.1002/14651858.CD004380.pub3>
- Whiting, P., Savović, J., Higgins, J.P., Caldwell, D.M., Reeves, B.C., Shea, B., Davies, P., Kleijnen, J., & Churchill, R., ROBIS group (2016). ROBIS: A new tool to assess risk of bias in systematic reviews was developed. *Journal of Clinical Epidemiology*, 69, 225-234.
<https://doi.org/10.1016/j.jclinepi.2015.06.005>
- World Health Organization. (2004). *Child sexual abuse: A silent health emergency. Regional Committee for Africa*, (AFR/ RC54/15 Rev. 1). Geneva: World Health Organization. Retrieved from <http://apps.who.int/iris/handle/10665/1878>

Table 1: Scope of the reviews included in this systematic review of reviews¹

Review (year)	Aim (participants)	Search strategy	No. of studies included	Total No. of participants (total attrition rate)	Length of intervention (No. of sessions) Follow up
Corcoran & Pillai (2008)	To determine the effects of parent-involved treatment in four major child symptom areas: internalizing, externalizing, sexualized behaviors, and post-traumatic stress when compared to another type of treatment or control group.	CINAHL, Infotrac, Medline, Psychinfo, Social Sciences Index, and Social Work Research and Abstracts (1980-summer 2005). Search terms provided but not full list to replicate the search. Authors contacted. No PRISMA flow diagram. Search restrictions: NR.	7	516 (186)	Mean of 12 weekly CBT sessions (but varied across studies) Follow up range from 12 weeks to 24 months (2 studies did not follow up)
Harvey & Taylor (2010)	To describe a more methodologically accurate meta-analysis of the effects of treatment for children and youth who have experienced sexual abuse.	PsycINFO, Social Services Abstracts, Medline, and Cochrane Central Register of Controlled Trials (through to 2009).	39	1258 (NR)	Between 6 and 12 sessions (for PTSD) Follow up (for PTSD): 1–3 months, 4–6 months, more than 6 months
Hetzel-Riggin, Brausch, & Montgomery (2007)	To investigate the independent effects of different treatment elements on several secondary problems related to childhood and adolescent sexual abuse (3-18 years of age). To investigate a number	PsychINFO, Social Science Abstracts, Medline Reference lists Review articles (from 1975 to 2004). Search terms provided but not full list to replicate the search.	28	1839 (NR)	NR NR

	of different moderators of treatment effectiveness.	No PRISMA flow diagram. Search restrictions (justified): published studies, English language.			
Macdonald et al. (2006, 2012)	To assess the efficacy of CBT approaches in addressing the immediate and longer-term sequelae of sexual abuse on children and young people up to 18 years of age.	Cochrane Central Register of Controlled Trials (CENTRAL) (2011 Issue 4); MEDLINE (1950 to 2011); EMBASE (1980 to 2011); CINAHL (1937 to 2011); PsycINFO (1887 to 2011); LILACS (1982 to 2011) and OpenGrey, previously Open SIGLE (1980 to 2011). ClinicalTrials.gov and the International Clinical Trials Registry Platform (ICTRP). References in previous reviews and studies. Contacted authors/experts. Established contacts in non-English speaking countries.	10	847 (for single studies)	Immediately post-intervention, 3 to 6 months later and at least 1 year later
Peltonen (2014)	To describe interventions' efficacy among sexually abused children, and analyze their effect in decreasing the post-traumatic stress symptoms.	Ebsco Host, Ovid (including PsycArticles) and Cochrane Database of Systematic Reviews (2000 to 2013). Reference lists of earlier reviews.	23	1318 (NR)	NR

Reeker, Ensing, & Elliott (1997)	To establish a quantitative measure of group treatment effectiveness for sexually abused children and adolescents from 3-17.	PsychInfo (1967 to 1996), reference lists, and review articles. Only one search descriptor provided (<i>sexual abuse (limited to) treatment</i>). No PRISMA flow diagram. Search restrictions (not justified): English language.	15	220 (NR)	Ranged from 3 to 25 (mean 14 sessions) NR
Sánchez-Meca et al. (2011)	The efficacy of the psychological treatment of children and adolescents that have suffered sexual abuse. Studying the influence of treatment, participant, and methodological variables on the effect sizes	PsycInfo and Medline (1970-2006). Reference list of studies and previous meta-analyses, SR, and theoretical reviews. Contacted several experts.	33	44 treatment groups and 7 control groups Attrition rate: 28% in the treatment groups and 17% in the control group in pretest	Mean for all studies: 12 weeks (range from 2 to 96). Mean for all studies: 18.70% (range from 0% to 58%).
Tehrani (2014)	To illustrate the efficacy of the diverse psychological treatments for children and adolescents who have experienced different forms of sexual abuse. Outlines the treatments and orientations that most benefit clients, depending	Ebsco, Google Scholar, other online sources, reference list of studies, and previous meta-analyses, SR, and theoretical reviews.	58	71 conditions (NR)	NR

	on the specific symptoms that they are presenting.				
Trask, Walsh, & DiLillo (2011)	Overall effectiveness of treatments for the negative consequences of CSA and the specific conditions under which treatments might be more or less effective.	Cochrane Database of Systematic Reviews, Dissertation Abstracts International, Eric, MEDLINE, ProQuest Dissertations and Theses, PsycINFO, the Social Science Citation Index, and Sociological Abstracts). Articles references and Journals relevant to CSA (<i>Journal of Child Sexual Abuse, Child Abuse and Neglect, Child Maltreatment, and Journal of Interpersonal Violence</i>) (1960 and December 2009).	35 19 single group 16 between group	1345 (NR)	6 weeks 3 months 9 months 1 year More than a year

NOTE: CBT: Cognitive behavioral therapy; NR: No reported; PTSD: Post traumatic stress disorder; SR: systematic review.

¹ Adapted from Smith et al. (2011, p.5)

Table 2: Summary of results reported¹

Review	Intervention type	Outcome	Design	Effect size (95% CI) ²	Follow up: Effect size (95% CI) ²	Moderator analysis
Corcoran & Pillai (2008)	CBT Modality: individual, group, joint parent-child session	Internalizing Externalizing Sexualized behaviors Post-traumatic stress	Experimental, randomization vs. treatment as usual, individual therapy, supportive therapies, waiting list control, other experimental conditions, compared with placebo or no treatment	Internalizing: $g=0.41 (0.21, 0.61)^{***}$ Externalizing: $g=0.32 (0.12, 0.52)^{**}$ Sexualized behaviors: $g=0.31 (1.00, 0.52)^{**}$ Post-traumatic stress: $g=0.36 (1.00, 0.54)^{**}$	Internalizing: $0.36 (-0.15, 0.86)$ Externalizing $0.19 (-0.20, 0.58)$ Sexualized behaviors $0.31 (1.00, 0.52)^{**}$ Post-traumatic stress $0.25 (-0.13, 0.63)$	Small number of studies precluded moderator analysis
Harvey & Taylor (2010)	Therapy: CBT (CBT, IRT, CCT, EMDR, RAP, SC, SIT, TF-CBT), insight oriented (individual therapy, narrative group therapy), eclectic, other Modality: individual, family, group, mixed	Overall Internalizing Externalizing Sexualized behaviors PTSD/trauma Self-appraisal Coping/functioning Social skills/competence	Experimental Quasi-experimental Uncontrolled Repeated measures	Overall: $g=1.37 (0.78, 1.95)^{NR}$ Internalizing: $g=0.74 (0.55, 0.94)^{NR}$ Externalizing: $g=0.52 (0.37, 0.67)^{NR}$ Sexualized behaviors: $g=0.49 (0.35, 0.63)^{NR}$ PTSD: $g=1.12 (0.76, 1.49)^{NR}$ Self-concept/esteem: $g=0.63 (0.37, 0.89)^{NR}$ Coping/functioning: $g=0.44 (0.20, 0.67)^{NR}$ Social skills: $g=0.38 (0.19-0.58)^{NR}$	Post-treatment: $g=0.76 (0.59, 0.93)^{NR}$ 1-3 months: $g=0.50 (0.26, 0.73)^{NR}$ 4-6 months: $g=0.79 (0.55, 1.04)^{NR}$ More than 6 months $g=0.95 (0.63, 1.27)^{NR}$	For global outcomes: Study design Outcome type Sample size Source of information Age Abuse length (for PSTD) Abuse number Therapy type Setting Modality No. of sessions Sessions length

						Therapist experience Family Context Therapy structure
Hetzel-Riggin, Brausch, & Montgomery (2007)	CBT, play, supportive, abuse-specific, no treatment, EMDR, only to parents Modality: group, individual, family	Global Behavior Psychological distress Self-concept Social functioning Other problems ³	Single- and between-group (pre- and post-test)	Global: $d=0.72$ (NR) ^{***} Behavior: $d=1.60$ (NR) Psychological distress: $d=1.05$ (NR) Self-concept: $d=0.71$ (NR) Social functioning: $d=0.48$ (NR) Other problems ³ : $d=1.49$ (NR)	NR	Children characteristics (age, gender, ethnicity ^{***}) Intrafamilial abuse ^{**} Therapy characteristics (No. sessions ^{**} , No. of months in therapy ^{**} , therapist training)
Macdonald et al. (2006, 2012)	CBT	Depression Anxiety PTSD Behavior problems (Secondary outcomes)	Randomized Quasi-randomized controlled trials compared to treatment as usual, with or without placebo control	Depression: OR: 1.9 (4.0-0.4) ^{***} Anxiety: OR: 0.23 (0.3-0.4) ^{***} PTSD: OR: 0.44 (0.16-0.73) ^{***} Behavior problems: OR: -0.65 (-3.53, 2.24)		Small number of included studies precluded moderator analysis but sensitivity analysis was provided
Peltonen (2014)	CBT	Post-traumatic Stress symptoms	Single group pre- and post-test RCT compared to no treatment, other interventions	Single studies: $d^5=0.66$ Psychosocial Interventions vs. no treatment: $d=-1.26$ (-1.72, -0.79) ^{***} Cognitive behavioral therapy vs. other interventions:		NR

				$d = -0.17 (-0.56, 0.22)^{ns}$		
Reeker, Ensing, & Elliott (1997)	Integrated ⁴ , CBT, drama therapy, play therapy Modality: group, individual	Overall General psychological distress Internalizing Externalizing Sexual behavior Self-esteem Knowledge of sexual abuse/prevention	Single group pre- and post-test	Overall: $d = 0.79$ (SD: 0.44) General psychological distress: $d = 0.73$ (SD: 0.49) Internalizing: $d = 0.64$ (SD: 0.44) Externalizing: $d = 0.56$ (SD: 0.57) Sexual behavior: $d = 0.77$ (SD: 0.15) Self-esteem: $d = 0.88$ (SD: 0.68) Knowledge of sexual abuse/prevention: $d = 0.99$ (SD: 0)	NR	Sex Age Treatment settings* Informants Outcome
Sánchez-Meca et al. (2011)	All types of treatment: CBT, play therapy, supportive therapy, psychodynamic therapy, humanistic therapy, mixture	Overall Anxiety Depression Sexualized behaviors Behavior problems Self-esteem	With and without a control group Unit of analysis the group (not the comparison between treated and control group)	Overall: $d^5 = 0.64$ (0.54, 0.75) ^{***} Anxiety: $d^5 = 0.53$ (0.40, 0.66) ^{***} Depression: $d^5 = 0.41$ (0.32, 0.50) ^{***} Sexualized behaviors: $d^5 = 0.45$ (0.35, 0.56) ^{***} Behavior problems: $d^5 = 0.66$ (0.54, 0.79) ^{***} Self-esteem: $d^5 = 0.61$ (0.37, 0.83) ^{***}		Type of treatment (significant for anxiety outcome, self-esteem) Number of sessions Duration (No. of weeks) Intensity (No. of hours per week) Magnitude (total No. of hours) Age Sex Type of abuse Intra-familial aggressor Study quality N in the post-test

						% attrition in the post-test
Tehrani (2014)	Trauma-Focused CBT Game-Based CBT Client-Centered-Therapy (Combined) Multidimensional therapy Dance/Movement-Therapy Equine-Facilitated-Therapy Psychodynamic Therapy Psychodrama Sexual-Abused-Preschoolers/Youths Sexual-Abuse-Specific Seeking-Safety Support-Based-Therapy Trauma-Focused Modality: group, individual, separate treatment of child and caregiver	Overall PTSD Dissociation Avoidance Re-experiencing Depression Anxiety Anxiety (State) Social (e.g., interpersonal skills, social acceptance) Sexualization Behavior Externalized Behavior Internalized Behavior Self-Perception Aggression Conduct Anger Somatization Attention Trust / Credibility Thought problems Parent practices	Independent measures, Repeated-measures RCTs Non-RCTs	Overall: $d = 0.68$ PTSD: $d = 0.59 (0.47, 0.72)^{NR}$ Dissociation: $d = 0.32 (0.18, 0.46)^{NR}$ Avoidance: $d = 1.50 (1.23, 1.76)^{NR}$ Re-experiencing $d = 1.73 (1.44, 2.02)^{NR}$ Depression: $d = 0.55 (0.47, 0.64)^{NR}$ Anxiety: $d = 0.48 (0.36, 0.60)^{NR}$ Anxiety (State): $d = 0.72 (0.54, 0.90)^{NR}$ Anxiety (Trait): $d = 0.78 (0.59, 0.97)^{NR}$ Social: $d = 0.42 (0.32, 0.52)^{NR}$ Sexualization: $d = 0.39 (0.30, 0.48)^{NR}$ Behavior: $d = 0.63 (0.52, 0.73)^{NR}$ Externalized Behavior: $d = 0.50 (0.41, 0.60)^{NR}$ Internalized Behavior: $d = 0.65 (0.52, 0.77)^{NR}$ Self-Perception: $d = 0.40 (0.26, 0.54)^{NR}$ Aggression: $d = 0.48 (0.18, 0.77)^{NR}$ Conduct: $d = 0.36 (0.10, 0.62)^{NR}$	NR	Type of treatment* Length of treatment Treatment modality ^{†§} Methodology ^{**} Reporters ^{NR} The treatment condition was a primary or secondary condition ^{**}

				Anger: $d = 0.35 (0.20, 0.49)^{NR}$ Somatization $d = 0.39 (0.16, 0.61)^{NR}$ Attention $d = 0.58 (0.24, 0.91)^{NR}$ Trust / Credibility $d = 0.56 (0.43, 0.68)^{NR}$ Thought problems $d = 0.45 (0.13, 0.76)^{NR}$ Parent practices $d = 0.45 (0.26, 0.64)^{NR}$	
Trask, Walsh, & DiLillo (2011)	CBT, Other	Overall PTSD symptoms, Externalizing problems, Internalizing problems	RCTs, Quasi-experimental, single group pre- and post-test	Pre-and post-test: Overall: $d=0.54 (0.40, 0.69)^{**}$ PTSD symptoms: $d= 0.51 (-0.01, -1.03)^{ns}$ Internalizing: $d= 0.50 (0.39, 0.61)^{**}$ Externalizing: $d= 0.47 (0.31-0.64)^{**}$ Between group: Overall: $d=0.54 (0.33-0.76)^{**}$ PTSD: $d= 0.63 (0.26-1.00)^{**}$ Internalizing: $d= 0.56 (0.33, 0.80)^{**}$ Externalizing: $d= 0.39 (0.60-0.70)^{**}$	Treatment characteristics (modality, duration, and inclusion of caregiver) Participant characteristics (age, gender, and ethnicity) Publication type

NOTE: d = Cohen’s effect size; g = Hedges’ effect size; CCT = child-centered therapy. EMDR = eye movement desensitization reprocessing. IRT =imagery rehearsal therapy; OR = Odd Ration; PTSD = post-traumatic stress disorder; RAP =Recovering from Abuse Program; SC = supportive counseling; SIT = stress inoculation training; SD = Standard deviation; TF-CBT = trauma-focused CBT.

*** $p < .001$., ** $p < .01$., * $p < .05$

¹Adapted from Smith et al. (2011, p. 5).

²Random effect model.

³Other problems included level of academic functioning and risk assessment abilities.

⁴Integrated included: psychoeducation regarding sexual abuse and sexual abuse prevention, exploration of the abuse experience, exploration of feelings, art therapy, play therapy, role plays, problem solving, puppet work, writing exercises, and behavior management.

⁵For treated group.

Table 3: AMSTAR-2 score

AMSTAR-2 Questions	Corcoran & Pillai (2008)	Harvey & Taylor (2010)	Hetzel-Riggin, Brausch, & Montgomery (2007)	Macdonald (2006, 2012)	Peltonen (2014)	Reeker, Ensing, & Elliott (1997)	Sánchez-Meca et al. (2011)	Tehrani (2014)	Trask, Walsh, & DiLillo (2011)	Cohen's Kappa (p value)	Inter-rater reliability (ICC, 95%CI)
<i>1. Did the research questions and inclusion criteria include the components of PICO¹?</i>	Yes	Yes	No	Yes	No	No	Yes	No	Yes	1 (0.003)	1
<i>2*. Did the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?</i>	No	No	No	Yes	No	No	No	No	No	1 (0.01)	1
<i>3. Did the authors explain their selection of the study designs for inclusion in the review?</i>	No	Yes	No	Yes	Yes	No	Yes	No	No	1 (0.01)	1
<i>4*. Did they use a comprehensive literature search strategy?</i>	No	Partial Yes	Partial Yes	Partial Yes	Partial Yes	No	Partial Yes	No	Partial Yes	1 (0.01)	1
<i>5. Did they perform study selection in duplicate?</i>	No	Yes	No	Yes	No	No	No	No	No	1 (.01)	1
<i>6. Did they perform data extraction in duplicate?</i>	No	Yes	Yes	Yes	No	No	Yes	No	Yes	0.64 (SE. 0.212, 0.02)	0.67 (0.18, 0.90)**

<i>7*. Did they provide a list of excluded studies and justify the exclusions?</i>	Partial Yes	Yes	No	Yes	No	No	No	No	No	0.44 SE:0.298 (0.03)	0.58 (0.05, 0.86)*
<i>8. Did they describe the included studies in adequate detail?</i>	Partial Yes	No	No	Yes	Yes	No	No	No	Partial Yes	0.84 SE: 0.151 (.000)	0.68 (0.19, 0.90)**
<i>9a*. Did they use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review? RCT</i>	Partial Yes	No	No	Yes	No	Includes only NRSI	Partial Yes	No	No	0.85 SE:0.145 (.001)	0.65 (0.65, 0.97)***
<i>9b. NRSI</i>	Includes only RCTs	No	No	Includes only RCTs	No	No	No	No	No	1 (0.01)	1
<i>10. Did they report on the sources of funding for the studies included in the review?</i>	No	No	No	No	No	No	No	No	No	1 (0.01)	1
<i>11a*. If meta-analysis was performed did they use appropriate methods for statistical combination of results? RCT</i>	No	Yes	No meta-analysis included	Yes	No	No meta-analysis conducted	No	Yes	Yes	0.58 SE:0.208 (.006)	0.80 (0.43, 0.94)***
<i>11b. NRSI</i>	No meta-analysis conducted	No	No	No meta-analysis conducted	No	No	No	No	No	1 (0.01)	1
<i>12. If meta-analysis was performed, did they assess the potential impact of RoB in individual studies on the results</i>	No	No	No	No	No	No	Yes	No	No	1 (0.01)	1

of the meta-analysis or other evidence synthesis?

<i>13*. Did they account for RoB in individual studies when interpreting/ discussing the results of the review?</i>	No	No	No	Yes	No	No	Yes	No	No	1 (0.01)	1
<i>14. Did they provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?</i>	Yes	-0.10 se: .070 (0.74)	-0.11 (- 0.75, 0.53) ^{ns}								
<i>15*. If they performed quantitative synthesis, did they carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?</i>	No	Yes	Yes	No	No	No	Yes	No	Yes	1 (0.01)	1
<i>16. Did they report any potential sources of conflict of interest, including any funding they received for conducting the review?</i>	No	No	No	Yes	No	No	No	No	No	1 (0.01)	1

¹ PICO = population, intervention, control group, and outcome.

Table 4: AMSTAR-2 Assessment according to critical domains

AMSTAR-2 Question Number	Yes N (%)	Partial Yes N (%)	No N (%)	Not included/No meta-analysis conducted N (%)
Q2	1 (11%)	0 (0%)	8 (89%)	N/A
Q4	0 (0%)	6 (67%)	3 (33%)	N/A
Q7	2 (22%)	1 (11%)	6 (67%)	N/A
Q9a (RCT)	1 (11%)	2 (22%)	5 (56%)	1 (11%)
Q9b (NRSI)	0 (0%)	0 (0%)	7 (78%)	2 (22%)
Q11a (RCT)	4 (44%)	0 (0%)	3(33%)	2 (22%)
Q11b (NRSI)	0 (0%)	0 (0%)	7 (78%)	2 (22%)
Q13	2 (22%)	N/A	7 (78%)	N/A
Q15	4 (44%)	N/A	5 (57%)	N/A

NOTE: Q11a (RCT) and Q11b (NRSI) do not have the answer “Not included” like Q9a (RCT) and Q9b (NRSI), but “No meta-analysis conducted”

Supplement material

Table 1S: Summary of the AMSTAR-2 domains

AMSTAR-2 Questions	Corcoran & Pillai (2008)	Harvey & Taylor (2010)	Hetzel-Riggin, Brausch, & Montgomery (2007)	Macdonald (2006, 2012)	Peltonen (2014)	Reeker, Ensing, & Elliott (1997)	Sánchez-Meca et al. (2011)	Tehrani (2014)	Trask, Walsh, & DiLillo (2011)
<i>1. Did the research questions and inclusion criteria for the review include the components of PICO?</i>	Yes (all were well described in the inclusion criteria, p.555)	Yes (type of outcome not provided but general mention)	No (outcome not included in inclusion criteria but in computation and analysis of the effect sizes (p.129) or group comparison. It was single- and between-group (pre- and post-test))	Yes (see 3.1 for inclusion criteria for PICO and see 9. Characteristics of studies section for follow up)	No (no comparison group included neither inclusion criteria nor research question. The control group is not reported in the aim of the study)	No (no comparison group included. It was single group (pre-and post-test). Outcome not reported. “Results were based on empirical measures, as opposed to clinical impressions or unstructured interviews with parents” p.672)	Yes (outcome not reported in the inclusion criteria but in the research question)	No (control group and timeframe for follow up were not reported)	Yes (Table 1, p.12 PICO are described, and in the following sections: 1.5 the current meta-analysis (p.9), 2.1 Literature search; 2.2 Development of coding system (p.10))
<i>2. Did the report of the review contain an explicit statement that</i>	No (No registration nor prior (un)published protocol)	No (No registration nor prior (un)published protocol)	No (No registration nor prior (un)published protocol)	Yes (see 3.1.2 “The first protocol said the abuse should have occurred in the	No (No registration nor prior (un)published protocol)	No (No registration nor prior (un)published protocol)	No (No registration nor prior (un)published protocol)	No (No registration nor prior (un)published protocol)	No (No registration nor prior (un)published protocol)

<i>the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?</i>	<p><i>12 months before referral. The review authors subsequently removed this restriction as lacking a sufficiently clear rationale". Differences between the protocol and review are provided in section 8)</i></p>								
<i>3. Did the review authors explain their selection of the study designs for inclusion in the review?</i>	No (although randomization and control were not necessary, single group pre- and post-test designs were excluded but explanation was not provided)	Yes (see 2.1 <i>"The approach taken about inclusion criteria was less stringent than meta-analyses where only studies of the highest methodological quality (i.e., randomized trials or RCTs) are included. Many studies would have been</i>	No (They included pre-post designs but no justification was provided)	Yes (see 3.1.1 <i>"Studies were eligible for the review if the allocation of participants to experimental or control groups was by random allocation or quasi-random allocation (for example, by day of week, case number or alphabetical order). Studies comparing CBT</i>	Yes (<i>"Some investigators suggest that because of the greater difficulties in assessing their methodological quality non-randomized studies should be excluded from reviews. However, in the field of child sexual abuse (as in many other areas of</i>	No (They included studies based on empirical measures without justification)	Yes (at the end of the intro the authors compared what they did and did not do, with previous meta-analyses)	No (they included pre and post studies but justification was not provided)	No (No justification for including between group studies with a control group)

		<i>excluded had such criteria been applied, and the results would therefore only relate to a relatively small area of research with limited generalizability”)</i>		<i>versus treatment as usual (referred to in the protocol as 'another intervention'), with or without placebo control, were eligible, as were studies comparing one intervention versus control)”</i>	<i>health care) few randomized controlled intervention trials exist p.13)”</i>				
<i>4. Did the review authors use a comprehensive literature search strategy?</i>	No (publication restrictions were not provided)	Partial Yes (restriction for English language not provided but for unpublished studies. Completion of the review was not provided)	Partial Yes (list of terms provided but not fully to replicate the search. No search for trials or studies registries. No grey literature considered or time for completion for the search)	Partial Yes (although all these were included (see 3.2.) list of terms to replicate the search provided in Appendix. Search conducted within 24 months. They did not support justification for language restriction)	Partial Yes (search restricted to “ <i>international scientific journals in English</i> ” (p.13) but no justification for this restriction. Trial registers, consulted experts, grey literature were not searched. The search period was not reported)	No (only one data base and small number of keywords, and no justification for publication restrictions)	Partial Yes (only 2 databases, conducted search within 24 months was not provided)	No (language restrictions were not justified)	Partial Yes (time of completion of the search not provided or additional search)
<i>5. Did the review authors</i>	No (the selection in duplicate	Yes (the abstracts identified in	No (the selection in	Yes (see 3.3.1: “ <i>at least) two</i>	No (the selection in duplicate	No (the selection in duplicate	No (“ <i>The search procedure</i>	No (the selection in duplicate	No (the selection in duplicate

<i>perform study selection in duplicate?</i>	was not mentioned but it was done for coding)	the search were reviewed by both authors and appropriate reports identified. A consensus process was used (p.520))	duplicate was not reported)	<i>authors independently selected studies for inclusion in the review. Disagreements were resolved in conference and a third author”</i> Flow graph is provided (Figure 1)	was not reported. There was only one author)	was not reported)	<i>enabled us to select 33 papers that fulfilled the selection criteria producing a total of 51 groups of children”</i> (p.73-4). No specification of how search was done)	was not reported. There was only one author)	was not reported)
<i>6. Did the review authors perform data extraction in duplicate?</i>	No (principal investigator and 2 assistants coded the studies and discrepancies were discussed to consensus (p.455) but the kappa score was not reported)	Yes (inter-rater agreement was provided, p.521)	Yes (“ <i>The studies were coded separately by each of the three authors”</i> (p.129). Inter-rater reliabilities were reported)	Yes (see 3.3.2: “ <i>Two review authors independently extracted data. We resolved any disagreements or uncertainties by discussion. All decisions were documented and where necessary, we contacted authors of studies to assist in resolving problems</i>)	No (the extraction in duplicate was not reported. There was only one author)	No (the extraction in duplicate was not reported)	Yes (“ <i>two researchers independently coded a random sample of the meta-analyzed studies (20%) by applying the norms detailed in a previously produced codebook”</i> p.75)	No (the extraction in duplicate was not reported. There was only one author)	Yes (“ <i>all studies were completed by the first author. Further, the second author coded a randomly selected 25%”</i> p.10)
<i>7. Did the review authors</i>	Partial Yes (they provided	Yes (list of excluded	No (No flow diagram was	Yes (see 14.1 Figure 1)	No (a flow chart is provided	No (authors justified the	No (no mention on this topic, just	No (No PRISMA diagram, nor	No (No PRISMA

<i>provide a list of excluded studies and justify the exclusions?</i>	common reasons for exclusion but “the exact number of titles and abstracts screened was not tracked” (p. 455). No PRISMA diagram or list of complete excluded studies and the justification for exclusion were provided	studies and reasons for each study, p.520)	reported nor list of excluded studies)		with justifications for excluding the studies but there is not a complete list with justification for the exclusion of each study)	exclusion (p. 672) but they did not provide a list and justification for each study. No PRISMA diagram)	exclusion of single studies in inclusion criteria. No PRISMA diagram)	list of excluded studies with justification were reported)	flow nor list of excluded studies with justification were reported)
<i>8. Did the review authors describe the included studies in adequate detail?</i>	Partial Yes (Table 1 described all PICOS)	No (comparators not provided but the rest, including the YES are provided. See Tables 1, 2 and 3)	No (outcomes, comparators, research design not described (Table 2)	Yes (see section 9)	Yes (see Tables 1a and 1b)	No (comparators and research design not described)	No (Despite Appendix 1 having several characteristics of participants and studies they did not provide information about the outcome nor the control)	No (There is no Table nor description of included studies)	Partial Yes (Doses for intervention and comparator, setting were not described in detail for each study even though treatment duration was a moderator)
<i>9a. Did the review</i>	Partial Yes	No	No	Yes	No (although	Includes only NRSI	Partial Yes (they	No	No

<i>authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?</i>	(quality of studies was assessed: “type of randomization method, the extent to which participant and assessment blinding occurred, the extent of attrition, and whether intention-to-treat analyses were conducted”, p.455 but was not included)	(RoB not reported)	(RoB not reported)	(see 3.3.3: “We used the Cochrane Collaboration’s tool for assessing the risk of bias of included studies (Higgins 2008). Two review authors independently assessed the risk of bias within each included study based on the following six domains, with review authors’ judgements of low risk of bias, high risk of bias and unclear (uncertain) risk of bias”	authors reviewed the quality of the studies, this tool did not include domains of bias reported by AMSTAR-2)	developed a risk of bias assessment random assignment, sample size, the presence of attrition, the use of blind evaluators, the comparison between dropouts and completers, the use of intention-to-treat analysis, and the use of one or several clinicians, p.74-5)	(RoB not reported)	(RoB not reported)
<i>9b. NRSI</i>	Includes only RCTs	No (RoB not reported)	No (RoB not reported)	Includes only RCTs	No (although authors reviewed the quality of the studies, this tool did not include domains of bias reported	No (they applied the same methodological questions to RCT and non RCTs. That is, they did not include items related to RoB for confounding	No (RoB not reported)	No (RoB not reported)

					by AMSTAR-2)		and from selection bias)		
<i>10. Did the review authors report on the sources of funding for the studies included in the review?</i>	No (funding sources not reported)	No (funding sources not reported)	No (funding sources not reported)	No (funding sources not reported)	No (funding sources not reported)	No (funding sources not reported)	No (funding sources not reported)	No (funding sources not reported)	No (funding sources not reported)
<i>11a. If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results? RCT</i>	No (they provided a fixed and random explanation but outcomes were analyzed separately so that they should have used a fixed effect because a priori this analysis reduce the heterogeneity . However, no explanation was provided for using	Yes (see 2.6. They considered random effects due to the heterogeneous nature of the studies included (different designs, treatments, and outcomes measures))	No (not included)	Yes (see 3. “We based primary analyses on available data from all included studies relevant to the comparison and outcome of interest. We performed random effects meta-analyses, and illustrated these within the review using standard forest plots. In studies with more than one experimental group, we combined these arms to form a	No (although they ran separate analyses for RCT and NRSI, they did not explain what model was used (fixed or random effect model). They did not find heterogeneity, so that causes of heterogeneity was not examined)	No meta-analysis conducted	No ((No justification for combining the data) ethical reasons. Random and mixed models were applied because they are more realistic than fixed effects. See note 2 p.73)	Yes (see Data Analysis. “When results from an initial analysis find a group of studies to be homogeneous , the more common fixed effects model is used (Lipsey & Wilson, 2001). However, in the current study, many of the analyses resulted in heterogeneous	Yes (for quasi-experimental and randomized controlled designs, a standardized mean difference effect size was calculated, which examines differences between groups on mean values p.10)

	both models p.456)			<i>single group</i> ". Heterogeneity in section 3.3.7)				<i>s effect sizes, ...In such instances, effect sizes from a random effects model were reported</i> ". Heterogeneity was assessed with ANOVA and meta- regression p.39)	
<i>11b. NRSI</i>	No meta- analysis conducted	No (they provided information for all criteria except for the combination of raw or adjusting data)	No (Authors pooled estimates together without confounding)	No meta- analysis conducted	No (authors did not combine NRSI with RCTs, but statistical adjustments for confounders was not performed)	No (they do not mention what analyses they used (fixed/random)	No (They pooled RCT and NRSI and did not adjust for confounding)	No (NRSI were not adjusted for confounding)	No (they justified why they conducted separate analyses for RCT and non RCT but no cofounding information was provided; p.10)
<i>12. If meta- analysis was performed, did the review authors</i>	No (moderator analysis for RoB was not carried out	No (but power analysis for each individual	No (RoB analyses were not performed)	No (they did not perform regression analysis or	No (RoB analysis was not performed)	No (RoB analysis was not performed)	Yes (see Table 3. They run moderator analyses for	No (RoB analysis was not performed)	No (RoB analysis was not performed)

<i>assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?</i>	but for others)	study was provided)		estimated a pooled effect with low RoB studies because all reported high risk bias)				quality of studies)		
13. Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review?	No (RoB analysis was not performed, so it was not discussed)	No (RoB analysis was not performed, so it was not discussed)	No (RoB analysis was not performed, so it was not discussed)	Yes (see 5.3, 5.4, 5.5: “ <i>this review emphasizes the inherent methodological weaknesses of the available studies</i> ” and 6.2: “ <i>The single most important implication for research in this area is for researchers to better document and report study design and execution</i> ”)	No (RoB analysis was not performed, so it was not discussed)	No (RoB analysis was not performed, so it was not discussed)	Yes (see “implications for research”. Although 2 moderators were significant. Authors recognized the methodological issues and suggested what future studies should include)	No (RoB analysis was not performed, so it was not discussed)	No (RoB analysis was not performed, so it was not discussed)	
14. Did the review authors provide a satisfactory explanation for, and	Yes (they stated that there was not heterogeneity and they ran moderator	Yes (see Discussion. Heterogeneity analysis was performed. They discussed	Yes (moderators variability)	Yes (heterogeneity analysis was performed but not substantial heterogeneity in the results.	Yes	Yes (see Discussion. “ <i>This meta-analysis contained a heterogeneous</i>	Yes (“ <i>The heterogeneity analysis showed that the 44 treated groups</i>	Yes (The discussion (pp.74-82) is focused on sources of heterogeneity	Yes (In the discussion the section 4.1 Moderators of treatment	

<i>discussion of, any heterogeneity observed in the results of the review?</i>	analyses and the outcomes were analyzed separately)	the effects of moderators. “These differing effects were reflected among the heterogeneity among the studies leading to a focus in the present meta-analysis on investigating potential factors that could moderate the effectiveness of therapy in addition to outcomes measured...” p.529)		Heterogeneity for subgroups was not possible because of the small number of studies included. See 9 and 10.2)		<i>group of studies. The studies varied greatly in terms of the age range of participants, the type of treatment offered, the length of the treatment, the gender composition of the group, and other variables”.</i> They stated that conclusions that can be drawn from the analyses are limited)	<i>exhibited a great variability in their effect estimates and, as a consequence, it is necessary to search for which treatment, participant, and methodological characteristics of the studies can explain at least part of the variability”</i> (p. 78))	such as publication bias, outliers, study design, secondary treatment conditions, treatment modality)	effectiveness describe the effect of each moderator including study’s and participants’ characteristics (p.16). Heterogeneity is also considered in the section 4.3 Future research directions (p.17))
<i>15. If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias)</i>	No (performed but not discussed)	Yes (see 2.6.3, 2.6.4. They performed the Forest plot, funnel plot, statistical tests (Begg and Mazumdar's rank correlation test, Egger's	Yes (performed plot and statistical test and mentioned in limitations)	No (no statistical test for publication bias was reported but with the search strategy used, publication bias will be not expected)	No (there was not significant heterogeneity)	No (mentioned in limitations but no graphical or statistical test performed)	Yes (the Egger test and the Fail-safe N were applied for each outcome)	No (Author stated that publication bias was performed but neither the type of analysis, nor the result were reported)	Yes (In section 4.2. Limitations of this meta-analysis, authors mentioned the impact of the publication bias (p.16).

<i>and discuss its likely impact on the results of the review?</i>		regression, failsafe N, Orwin's failsafe N, Duval and Tweedie's trim and fill), and the cumulative Forest plot)								In section 2.4 Calculating effect sizes, publication bias test was performed in CMA(p.10))
<i>16. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?</i>	No (Conflict of interest and funding sources were not reported)	No (Despite the authors reporting “not official endorsement to the funding agency” they did not report the ties with the agency, nor conflict of interests)	No (Conflict of interest and funding sources were not reported)	Yes (support/funding : Northern Ireland Research & Development, UK Nordic Campbell Center, Denmark Potential conflicts of interest: Julian PT Higgins - received a payment from the Nordic Campbell Centre for his work on the first version of this review in 2006. All other authors - none known)	No (Conflict of interest and funding sources were not reported)	No (Conflict of interest and funding sources were not reported)	No (Conflict of interest and funding sources were not reported)	No (Conflict of interest and funding sources were not reported)	No (Funding sources were reported but not the conflict of interest)	

Total Score	Critically Low	Critically Low	Critically Low	Moderate		Critically Low	Critically Low	Critically Low	Critically Low
--------------------	----------------	----------------	----------------	----------	--	----------------	----------------	----------------	----------------

Table 2S: Excluded studies and reasons for exclusion

No meta-analytical review (n = 23)
Avinger, K.A., & Jones, R.A. (2007). Group treatment of sexually abused adolescent girls: A review of outcome studies. <i>The American Journal of Family Therapy</i> , 35(4), 315-326. doi: 10.1080/01926180600969702
Beutler, L.E., Williams, R.E., & Zetzer, H.A. (2009). <i>Efficacy of treatment for victims of child sexual abuse</i> . The Future of Children Center for the Future of Children the David and Lucile Packard Foundation 4(2), 156–175. Bowlby, J.
Capella C., & Gutiérrez C. (2014). Psychotherapy with sexually abused children and adolescents: Reparation, reconstruction of meaning and healing. <i>Psicoperspectivas: Individuo y Sociedad</i> , 13(2), 93-105.
Choudhary, V., Satapathy, S., & Sagar, R. (2016). Review of randomized controlled trials on psychological interventions in child sexual abuse: Current status and emerging needs in the Indian context. <i>Indian Journal of Psychological Medicine</i> , 38(4), 279–284. doi: 10.4103/0253-7176.185954
Cummings, M., Berkowitz, S.J., Scribano, P.V. (2012). Treatment of childhood sexual abuse: An updated review. <i>Current Psychiatry Reports</i> , 14(6), 599-607.
Herrera C.R., & Parra A.F. (2011). Child sexual abuse: A review based on empirical evidence [Abuso sexual infantil: Una revisión con base en pruebas empíricas]. <i>Behavioral Psychology/ Psicología Conductual</i> 19(1), 7-39.
Hunter, S.V. (2006). Understanding the complexity of child sexual abuse: a review of the literature with implications for family counseling. <i>The Family Journal</i> 14(4), 349-358.
James, S., & Mennen, F. (2001). Treatment outcome research: How effective are treatments for abused children? <i>Child and Adolescent Social Work Journal</i> , 18(2) 73–95.
Kim, S., Noh, D., & Kim, H. (2016) A summary of selective experimental research on psychosocial interventions for sexually abused children, <i>Journal of Child Sexual Abuse</i> , 25(5), 597-617. doi: 10.1080/10538712.2016.1181692
Lalor, K. & McElvaney R. (2010). Child sexual abuse, links to later sexual exploitation/high-risk sexual behavior, and prevention/treatment programs. <i>Trauma Violence Abuse</i> , 11(4), 159-77. doi: 10.1177/1524838010378299.
Lev-Wiesel, R. (2008). Child sexual abuse: A critical review of intervention and treatment modalities. <i>Children and Youth Services Review</i> , 30(6), 665-673. doi: 10.1016/j.chilyouth.2008.01.008.
Luther, S.L., & Price, J.H. (1980). Child sexual abuse: A review. <i>The Journal of School of Health</i> , 50(3),161 -5.
Miffitt, L.A. (2014). State of the science: Group therapy interventions for sexually Abused Children. <i>Archives of Psychiatric Nursing</i> , 28, 174–179.
Parker, B., & Turner, W. (2014). Psychoanalytic/psychodynamic psychotherapy for sexually abused children and adolescents: A systematic review. <i>Research on Social Work Practice</i> , 24(4), 389-399. doi: 10.1177/1049731514525477.
Passarella, C.D.M., Mendes, D.D., & De Jesús, Mari, J. (2010). A systematic review to study the efficacy of cognitive behavioral therapy for sexually abused children and adolescents with posttraumatic stress disorder. <i>Archives of Clinical Psychiatry (São Paulo)</i> , 37(2), 60-65.
Putnam, F.W. (2003). Ten-year research update review: Child sexual abuse. <i>Journal of the American Academy of Child and Adolescent Psychiatry</i> , 42(3), 269-278.
Ramchandani, P., & Jones, D. (2003). Treating psychological symptoms in sexually abused children: From research findings to service provision. <i>British Journal of Psychiatry</i> , 183(6), 484-490. doi:10.1192/03-99.
Ross, G., & O'Carroll, P. (2004). Cognitive behavioural psychotherapy intervention in childhood sexual abuse: Identifying new directions from the literature. <i>Child Abuse Review</i> , 13(1), 51-64.

Saywitz, K. J., Mannarino, A. P., Berliner, L., & Cohen, J. A. (2000). Treatment of sexually abused children and adolescents. *American Psychologist*, 55(9), 1040-1049. doi: 10.1037/0003-066X.55.9.1040.

Swenson C.C., & Chaffin M. (2000). Beyond psychotherapy: Treating abused children by changing their social ecology. *Aggression and Violent Behavior* 11(2), 120-137.

Tavkar, P., & Hansen, D. J. (2011). Interventions for families victimized by child sexual abuse: Clinical issues and approaches for child advocacy center-based services. *Aggression and Violent Behavior*, 16(3), 188-199.

Tourigny, M., Hébert, M., Daigneault, I., & Simoneau, A.C. (2005). Efficacy of a group therapy for sexually abused adolescent girls. *Journal of Child Sexual Abuse*, 14(4), 71-93. doi: 10.1300/J070v14n04_04.

Van Toledo A., & Seymour F. (2013). Interventions for caregivers of children who disclose sexual abuse: A review. *Clinical Psychology Review*, 33(6), 772-781.

Therapy but not for children/adolescent sexually abused (n = 4)

St Amand, A., Bard, D.E., & Silovsky, J.F. (2008). Meta-analysis of treatment for child sexual behavior problems: Practice elements and outcomes. *Child Maltreatment*, 13(2), 145-66. doi: 10.1177/1077559508315353.

Rodenburg, R., Benjamin, A., de Roos, C., Meijer, A.M., & Stams, G.J. (2009). Efficacy of EMDR in children: A meta-analysis. *Clinical Psychology Review*, 29(7), 599-606. doi: 10.1016/j.cpr.2009.06.008.

Thomas, R., Abell, B., Webb, H.J., Avdagic, E., & Zimmer-Gembeck, M.J. (2017). Parent-child interaction therapy: A meta-analysis. *Pediatrics*, 140(3). pii: e20170352. doi: 10.1542/peds.2017-0352.

Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcome studies. *Psychological Bulletin*, 117(3), 450-468. doi: 10.1037/0033-2909.117.3.450

Trauma (n=4)

Cary, C.E., & McMillen, J.C. (2012). The data behind the dissemination: A systematic review of trauma-focused cognitive behavioral therapy for use with children and youth. *Children and Youth Services Review*, 34(4), 748-757. doi: 10.1016/j.childyouth.2012.01.003.

Chen, R., Gillespie, A., Zhao, Y., Xi, Y., Ren, Y., & McLean, L. (2018). The efficacy of eye movement desensitization and reprocessing in children and adults who have experienced complex childhood trauma: A systematic review of randomized controlled trials. *Frontiers in Psychology*, 9(534), 1-11. doi:10.3389/fpsyg.2018.00534

Purgato, M., Gross, A. L., Jordans, M. J., de Jong, J. T., Barbui, C., & Tol, W. (2014). Psychosocial interventions for children exposed to traumatic events in low- and middle-income countries: Study protocol of an individual patient data meta-analysis. *Systematic Reviews*, 3(34), 1-7. doi:10.1186/2046-4053-3-34

Wethington, H.R., Hahn, R.A., Fuqua-Whitley, D.S., Sipe, T.A., Crosby, A.E., Johnson, R.L., Liberman, A.M., Mościcki, E., Price, L.N., Tuma, F.K., Kalra, G., & Chattopadhyay, S.K. (2008). The Effectiveness of interventions to reduce psychological harm from traumatic events among children and adolescents: A systematic review. *American Journal of Preventive Medicine*, 35(3), 287-313.

Maltreatment (n= 3)

Chen, M., & Chan, K.L. (2016). Effects of parenting programs on child maltreatment prevention: A meta-analysis. *Trauma Violence Abuse*, 17(1), 88-104. doi: 10.1177/1524838014566718.

Fry, D., McCoy, A., & Swales, D. (2012). The consequences of maltreatment on children's lives: A systematic review of data from the East Asia and Pacific Region. *Trauma, Violence, and Abuse*, 13(4), 209-233. doi: 10.1177/1524838012455873.

Skowron, E., & Reinemann, D. H. S. (2005). Effectiveness of psychological interventions for child maltreatment: A meta-analysis. *Psychotherapy: Theory, Research, Practice, Training*, 42(1), 52-71. doi: 10.1037/0033-3204.42.1.52

No treatment (n=2)

Collin-Vézina, D., Daigneault, I., & Hébert, M. (2013). Lessons learned from child sexual abuse research: prevalence, outcomes, and preventive strategies. *Child and Adolescent Psychiatry and Mental Health*, 7(1), 22. doi:10.1186/1753-2000-7-22.

Mohammadi, M.R., Zarafshan, H., & Khaleghi, A. (2014). Child abuse in Iran: A systematic review and meta-analysis. *Iranian Journal of Psychiatry* 9(3), 118-124.

PosTraumatic Stress Disorder (n=2)

Gillies D., Taylor F., Gray C., O'Brien L., & D'Abrew N. (2012). Psychological therapies for the treatment of post-traumatic stress disorder in children and adolescents. *Cochrane Database Systematic Reviews.*, 12(12):CD006726. doi: 10.1002/14651858.CD006726.pub2.

Hetrick, S.E., Purcell, R., Garner, B., & Parslow, R. (2010). Combined pharmacotherapy and psychological therapies for post-traumatic stress disorder (PTSD). *Cochrane Database Systematic Reviews*, 7(7):CD007316. doi: 10.1002/14651858.CD007316.pub2.

Violence exposure (n = 1)

Lindstrom Johnson, S., Elam, K., Rogers, A.A., & Hilley, C. (2018). A meta-analysis of parenting practices and child psychosocial outcomes in trauma-informed parenting interventions after violence exposure. *Prevention Science*, 19, 927–938. doi:10.1007/s11121-018-0943-0

An overview of meta-analyses (n = 1)

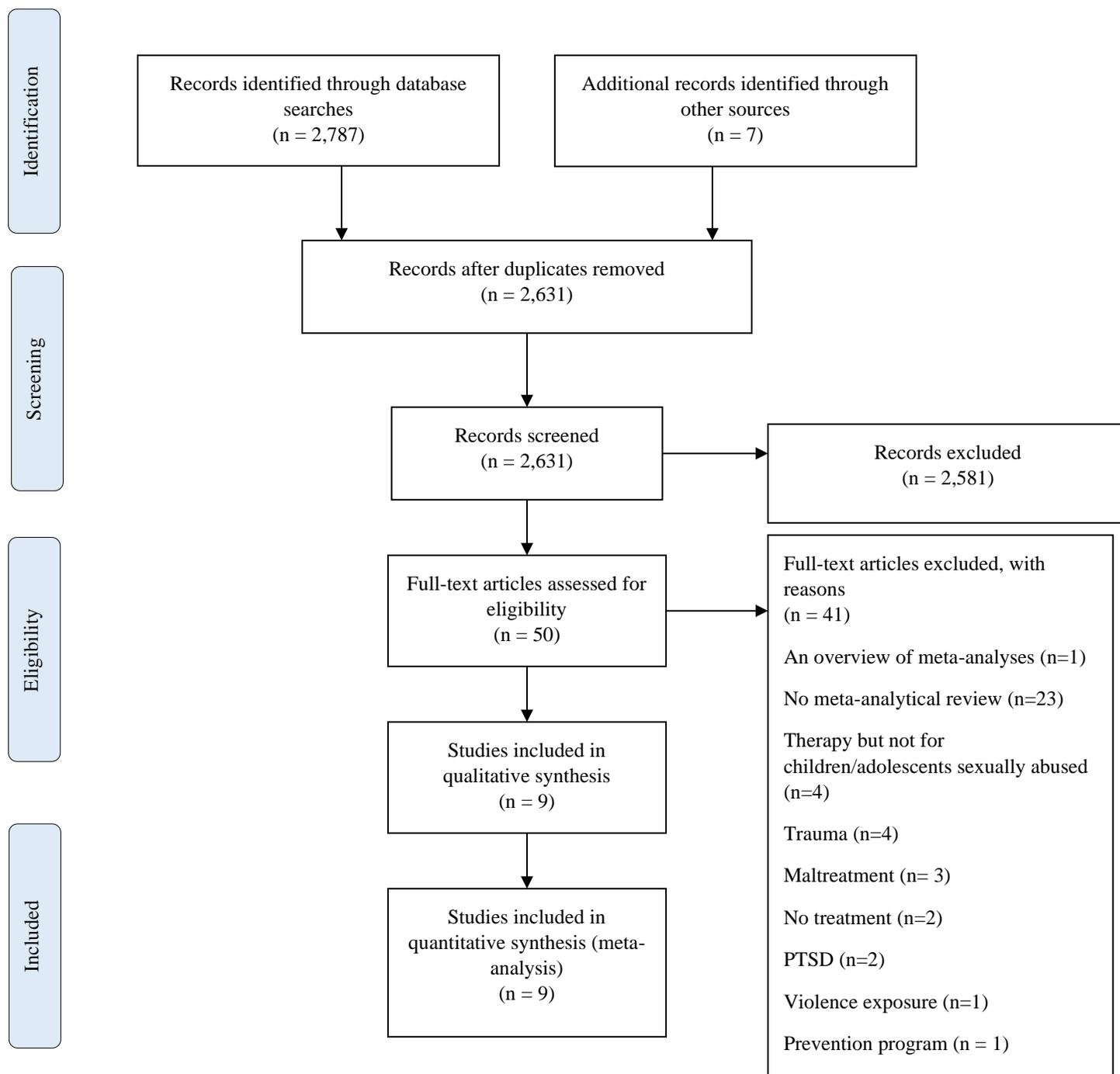
Benuto, L.T., & O'Donohue, W. (2015). Treatment of the sexually abused child: Review and synthesis of recent meta-analyses. *Children and Youth Services Review*, 56, 52-60. doi: 10.1016/j.childyouth.2015.06.009.

Prevention Program (n = 1)

Walsh, K., Zwi, K., Woolfenden, S., & Shlonsky, A. (2015). School-based education programmes for the prevention of child sexual abuse. *Cochrane Database of Systematic Reviews*, 4: CD004380. doi: 10.1002/14651858.CD004380.pub3



Figure 1. PRISMA 2009 Flow Diagram



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097