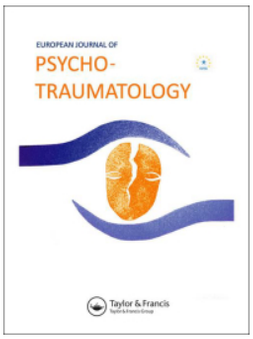

European Journal of Psychotraumatology

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/zept20

# Differences of TF-CBT treatment effects using various outcome measures: a meta-analysis

**Bianca Schreyer, Jonathan Felix Benjamin Thielemann, Barbara Kasparik & Rita Rosner**

**To cite this article:** Bianca Schreyer, Jonathan Felix Benjamin Thielemann, Barbara Kasparik & Rita Rosner (2024) Differences of TF-CBT treatment effects using various outcome measures: a meta-analysis, European Journal of Psychotraumatology, 15:1, 2406136, DOI: 10.1080/20008066.2024.2406136

**To link to this article:** https://doi.org/10.1080/20008066.2024.2406136

EUROPEAN JOURNAL OF PSYCHOTRAUMATOLOGY
2024, VOL. 15, NO. 1, 2406136
https://doi.org/10.1080/20008066.2024.2406136

REVIEW ARTICLE

OPEN ACCESS Check for updates

# Differences of TF-CBT treatment effects using various outcome measures: a meta-analysis

Bianca Schreyer, Jonathan Felix Benjamin Thielemann, Barbara Kasparik and Rita Rosner

Department of Psychology, Catholic University of Eichstätt-Ingolstadt, Eichstätt, Germany

## ABSTRACT

**Background:** Diagnostic criteria of posttraumatic stress disorder in children and adolescents and corresponding instruments have undergone significant changes over time. However, the impact of different outcome measures on treatment effects in the context of posttraumatic stress symptoms (PTSS) has not yet been explored.

**Objective:** TF-CBT is a well-researched first-line treatment for PTSS among children and adolescents and thus, an ideal candidate to examine the potential influence of different outcome measures by meta-analysis.

**Method:** A comprehensive literature search was conducted in December 2023 using seven databases. Studies included RCTs as well as non-controlled studies examining the effects of TF-CBT on pediatric PTSS. We extracted treatment effects and investigated whether there were systematic differences in the effects based on the outcome measures and their underlying DSM version.

**Results:** In total, 76 studies (35 RCTS) met the eligibility criteria. Hedges g effect sizes with 95% confidence intervals (CI) were computed and high-risk of bias studies were excluded. No significant difference was observed between DSM-IV and DSM-5 based instruments. Individual outcome measures were found to be comparable overall, with some appearing somewhat more sensitive to change. Although a small but significant difference in true effect sizes for individual outcome measures was found, this only concerned the UCLA PTSD ($g = 1.06$) and the CPSS ($g = 1.61$) with the effect most likely being due to chance or confounding variables. TF-CBT showed large effect sizes on PTSS in within-study comparison ($g = 1.32$) and medium between-studies effect sizes ($g = .57$).

**Conclusions:** While we could not establish equivalence, there seems to be no difference regarding the measurement of treatment effects based on outcome measure and underlying DSM version. The updated TF-CBT effect size confirmed it as an effective treatment for PTSS and secondary outcomes in children and adolescents.

## Diferencias en los efectos del tratamiento con Terapia Cognitivo Conductual Centrada en el Trauma (TCC-CT) utilizando diversas medidas de resultados: un metaanálisis

**Antecedentes:** Los criterios diagnósticos del trastorno de estrés postraumático en niños y adolescentes y los instrumentos correspondientes han sufrido cambios significativos a través del tiempo. Sin embargo, aún no se ha explorado el impacto de las diferentes medidas de resultado en los efectos del tratamiento en el contexto de los síntomas de estrés postraumático (TEPT).

**Objetivo:** La Terapia Cognitivo Conductual Centrada en el Trauma (TCC-CT) es un tratamiento de primera línea con investigaciones para el TEPT entre niños y adolescentes y, por lo tanto, un candidato ideal para examinar la posible influencia de diferentes medidas de resultado mediante un metaanálisis.

**Método:** Se realizó una búsqueda bibliográfica exhaustiva en diciembre de 2023 utilizando siete bases de datos. Los estudios incluyeron Ensayos Controlados Aleatorizados (ECA) y estudios no controlados que examinaron los efectos de la TCC-CT en el TEPT pediátrico. Se extrajeron los efectos del tratamiento y se investigó si había diferencias sistemáticas en los efectos en función de las medidas de resultado y la versión subyacente del DSM.

**Resultados:** En total, 76 estudios (35 ECA) cumplieron los criterios de elegibilidad. Se calculó la medida del tamaño del efecto con la g de Hedges con un intervalo de confianza del 95% (IC 95%) y se excluyeron los estudios con alto riesgo de sesgo. No se observó ninguna diferencia significativa entre los instrumentos basados en el DSM-IV y el DSM-5. Se encontró que las medidas de resultado individuales eran comparables en general, mientras que algunas parecían más sensibles al cambio. No obstante, se encontró una diferencia pequeña pero significativa en los tamaños de efecto reales para las medidas de resultado

## ARTICLE HISTORY

Received 7 May 2024
Revised 9 September 2024
Accepted 9 September 2024

## KEYWORDS

TF-CBT; meta-analysis; post-traumatic stress disorder; psychotherapy; children; outcome measure; diagnostic criteria

## PALABRAS CLAVE

TCC-CT; metaanálisis; trastorno de estrés postraumático; psicoterapia; niños; medida de resultado; criterios diagnósticos

## HIGHLIGHTS

- No difference between outcome measures for posttraumatic stress symptoms in children and adolescents and their underlying DSM-criteria could be established.
- TF-CBT has again been confirmed TF-CBT as a treatment of first choice for PTSS in children and adolescents.

**CONTACT** Rita Rosner rita.rosner@ku.de Department of Clinical and Biological Psychology, Catholic University of Eichstätt-Ingolstadt, Ostenstraße 25, Eichstätt 85072, Germany

Supplemental data for this article can be accessed online at https://doi.org/10.1080/20008066.2024.2406136

individuales, esto solo afectó al cuestionario UCLA PTSD ($g$ = 1,06) y al cuestionario CPSS ($g$ = 1,61); y el efecto probablemente se debió al azar o a variables de confusión. La TCC-CT mostró grandes tamaños de efecto sobre el TEPT en la comparación dentro del estudio ($g$ = 1,32) y tamaños de efecto medianos entre estudios ($g$ = 0,57).
**Conclusiones:** Si bien no pudimos establecer equivalencia, parece que no existe diferencia con respecto a la medición de los efectos del tratamiento según la medida de resultado y la versión subyacente del DSM. El tamaño del efecto actualizado de la TCC-CT confirmó esta psicoterapia como un tratamiento eficaz para el TEPT y los resultados secundarios en niños y adolescentes.

## 1. Introduction

The diagnosis of posttraumatic stress disorder (PTSD) for children and adolescents is considered to be difficult, as the criteria are often considered as too strict and many cases are misclassified even when symptomatic and impaired (e.g. Scheeringa et al., 2012). Furthermore, the diagnostic criteria have undergone substantial changes over time. The updated DSM-5 PTSD-diagnosis introduced an additional symptom cluster, which made the diagnostic criteria stricter than those of the previous three-cluster model of DSM-IV. In a sample of Danzi and La Greca (2016), the authors reported that only half of the children with a DSM-IV PTSD diagnosis were also identified according to DSM-5 criteria. In accordance with the changes to the diagnostic criteria, screening, and diagnostic instruments have evolved over time. Additionally, some older instruments use a dichotomous scale, while newer ones use a dimensional scale. Consequently, studies and practitioners employ a variety of instruments to assess symptoms, which are frequently based on different criteria in accordance with the corresponding ICD or DSM version. This may produce different evidence regarding the effectiveness of a specific treatment. This may bias clinical decision-making by causing misclassification of patients and influencing treatment choices. Moreover, meta-analyses rely on the assumption that all instruments are equally adequate to measure symptoms, which may be challenged by the changes made over time.

### 1.1. Instruments

The gold standard for assessing PTSD among children and adolescents is the Clinician-Administered PTSD Scale for DSM-5 – Child/Adolescent Version (CAPS-CA-5; Pynoos et al., 2015). It is a clinical interview using Likert-Scale and carefully examines each PTSD symptom according to DSM-5. However, another widely used instrument is the Schedule for Affective Disorders and Schizophrenia for School-Age Children Present and Lifetime Version (K-SADS-PL; Kaufman et al., 1997). This instrument is an interview with the interviewer rating each item categorically (*yes, no, no information*). Other commonly used self-report instruments for assessing posttraumatic stress symptoms (PTSS) in children and adolescents include the Child PTSD Symptom Scale DSM-5 (CPSS-5; Foa et al., 2018), the UCLA Child/Adolescent PTSD Reaction Index for DSM-5 (UCLA PTSD-RI; Pynoos & Steinberg, 2014) and the Child and Adolescent Trauma Screen (CATS; Sachser et al., 2017). Some of these instruments can be used as interview or questionnaire (e.g. CPSS, UCLA), while others are intended to be used only as interview (CAPS-CA, K-SADS) or only as questionnaire (CATS).

### 1.2. Trauma-focused Cognitive Behavioural Therapy

Trauma-focused Cognitive Behavioural Therapy (TF-CBT) is considered to be the first-line treatment for PTSD in children and adolescents in various guidelines (Forbes et al., 2020; National Institute for Health and Care Excellence, 2018; Phoenix Australia Centre for Posttraumatic Mental Health, 2020; Schäfer et al., 2019; overview Steil et al., 2021). Its effectiveness and treatment superiority has been widely studied and supported by various meta-analyses, comprising almost 30 RCTs: in a subgroup analysis of 18 studies, Gutermann et al. (2016) reported large pre–post effect sizes for PTSS. A network meta-analysis by Mavranezouli et al. (2020) comparing 29 studies and 63 study arms suggested TF-CBT according to the manual of Cohen et al. (2006, 2017) to be one of the most effective therapies for reducing PTSS in children and adolescents. Moreover the meta-analyses by Thielemann et al. (2022, 2023) showed a medium controlled effect size for PTSS and small controlled effect sizes for depression, anxiety and grief including waitlist, treatment-as-usual and active conditions. There are several moderators in treatment response of children and adolescents with PTSD symptoms include gender, age, ethnicity, domicile, parent/caregiver involvement, treatment dose and trauma type. However, the findings concerning these moderating factors are highly heterogeneous (Danzi & La Greca, 2021). Other potential moderator may be treatment modality (individual vs. group) or type of measurement (self-report vs. interview). TF-CBT has been rated with the highest certainty of evidence by the German Guidelines (Schäfer et al., 2019). While TF-CBT is used to refer to the specific treatment manual by Cohen et al. (2006, 2017), it also is used as an umbrella term for trauma-focused cognitive behavioural

therapies created by other researchers. For the remainder of this paper, we will use TF-CBT to refer only to the Cohen et al. protocol. TF-CBT according to the manual of Cohen et al. (2006, 2017) consists of nine components, represented by the acronym PRACTICE: Psychoeducation and Parenting skills (P), affective regulation (A), cognitive coping (C), trauma narrative (T), in vivo exposure (I), conjoint parent–child sessions (C) and enhancing future safety and development (E).

Although the effectiveness of TF-CBT is well described, none of the meta-analyses described above examined the diagnostic criteria and instruments used to quantify treatment effects. Since considerable heterogeneity was observed in these meta-analyses, this meta-analysis will examine whether the heterogeneity can be explained by the use of different outcome measures.

### 1.3. Current study

In meta-analyses, the different outcome measures used in individual studies to quantify treatment effects are often combined and considered to be equally strict. This could substantially influence effect sizes, especially when adjustments in construct definitions occur over time, as is the case with the regular updates of the International Classification of Diseases (ICD) and the Diagnostic and Statistical Manual of Mental Disorders (DSM). These revisions have resulted in the refinement of diagnostic criteria and the development of new instruments or revisions of existing ones. As TF-CBT has remained mostly unchanged over time, it can be assumed similarly effective. In addition, a great number of studies on TF-CBT using various outcome measures has been published over the years, making it an ideal candidate to examine their effect. To minimise heterogeneity due to different treatment interventions we limited the analysis to this specific protocol. And to reduce bias due to different control groups, we only used within group effect sizes and included uncontrolled studies to maximise power. Therefore, the aim of this meta-analysis is to evaluate the comparability of TF-CBT treatment effects measured by (1) individual outcome measures as well as (2) outcome measures grouped by their corresponding DSM version. Moreover, this meta-analysis updates the pooled within and between group effect sizes of TF-CBT on PTSS, depression, anxiety and grief of our previous analysis (Thielemann et al., 2022, 2023).

## 2. Method

### 2.1. Search and selection of studies

This study used the datasets from a previous publication (Thielemann et al., 2022), but we conducted follow-up research of studies that were published since then. This previous meta-analysis was pre-registered in the International Prospective Register of Systematic Reviews (PROSPERO), registration number: CRD42020139403.

**Table 1.** Pre-defined search terms.

| Search categories | Search terms |
|---|---|
| Diagnosis | Trauma* or posttrauma* or post-trauma* or PTSD or PTSS or grief or griev* |
| Trauma-related | Abuse* or assault* or abduct* or accident* or kidnapp* or life-threat* or maltreat* or mistreat* or neglect* or refugee or shooting or terroris* or victim* or violence or war or hurricane or tsunami or earthquake or flood or 'natural disaster' or bereave* or loss |
| Youth | Adolescen* or child* or youth or kid or juvenile or infant or minor or teenager or young* |
| TF-CBT | 'Trauma focused cognitive behavioral treatment' or 'trauma-focused cognitive behavioral treatment' or 'trauma focused cognitive behavioral therapy' or 'trauma-focused cognitive behavioral therapy' or 'trauma focused cognitive behavior*' or 'trauma-focused cognitive behavior*' or 'trauma focused cog*' or 'trauma-focused cog*' or 'trauma focused' or trauma-focused or TF-CBT or grief-focused or 'grief focused' |

Note. Combination for searching the databases: (Diagnosis or Trauma-related) and Youth and TF-CBT.

The same search terms (see Table 1) and databases as in the previous study were used. The databases included PsychInfo, MEDLINE, Cochrane Library, PTSDPubs, PubMed and Web of Science as well as OpenGrey. The existing dataset included studies between Jan 1st, 1990, to Aug 19th, 2021. Thus, the added data included additional studies published between Aug 19th 2021 and Dec 04th 2023. Additionally, we manually searched the references of relevant articles. Two independent raters screened all titles and abstracts using Covidence (Veritas Health Innovation, 2014). Conflicts were resolved by reviewing the abstracts and discussing them with the co-authors. The first or second author reviewed all full texts for the remaining studies after title and abstract screening and assessed whether they met the inclusion or exclusion criteria. Any uncertainties were resolved by contacting the authors of the relevant publications and consulting with co-authors.

### 2.2. Inclusion and exclusion criteria

To coincide with the existing database, we used the same inclusion and exclusion criteria as the previous meta-analyses (Thielemann et al., 2022, 2023). Specifically, these are: (1) patients age between 3 and 21 years, which (2) experienced at least one traumatic event and (3) had at least 8 sessions of TF-CBT according to Cohen et al. manual (2006, 2017) or one of its earlier versions (Cohen & Mannarino, 1993; Deblinger & Heflin, 1996). Furthermore, (4) results had to be on a quantitative PTSS measure applied before and after treatment and based on self-

report or a clinical interview. (5) We only included original research and excluded case reports, reviews, and meta-analyses.

For studies in group settings, there were slightly different criteria. As many group interventions are designed with fewer sessions, there was no requirement for a minimum number of sessions or the implementation of all components of the TF-CBT manual. Instead, only psychoeducation, coping strategies, exposure, cognitive processing/restructuring of trauma-related thoughts and beliefs, and some reference to the manual or one of its earlier versions were mandatory. Studies were excluded if (1) not children themselves were recipients of treatment, (2) studies did not report pre–post PTSS or (3) the data could not be obtained by contacting the authors. There were no restrictions in language.

### 2.3. Outcome and data extraction

To ensure accuracy, all outcome data was extracted by two individuals. Inconsistencies were resolved through discussion. We extracted outcome data on PTSS, depression, anxiety, and grief noting the diagnostic instruments and criteria used. We prioritised data from clinical interviews and if unavailable used those from self-reports. If only subscales of PTSD clusters were reported, we merged them using the formula presented in the Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al., 2022). If studies did not present suitable data, we contacted the authors for missing information and excluded studies without pre – and post-treatment data.

### 2.4. Risk of bias assessment

Two assessors rated the risk of bias of all included studies using the Risk of Bias assessment tool (RoB 2.0; Sterne et al., 2019) and the Risk of Bias In Non-randomized Studies – of Interventions (ROBINS-I; Sterne et al., 2016) assessment tool. The RoB 2.0 tool utilises five domains to rate studies for bias: randomisation process, deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of the reported result. These domains are used to determine whether a study has a 'low risk', 'some concerns', or a 'high risk' of bias by an algorithm. The ROBINS-I tool initially uses seven domains, but since three were already covered by our inclusion criteria (confounding, selection bias, bias in classification of intervention) we only assessed the remaining four domains, which can be matched to the Domains two to five of the RoB tool. To compare both tools, we translated the categories of ROBINS-I in the categories of RoB combining 'severe' and 'critical' as 'high risk'.

### 2.5. Statistical analysis

Comprehensive Meta-Analysis, version 4 (Borenstein et al., 2022) was used for computing effect sizes and subgroup analyses. All eligible studies were analysed using effect size Hedges' *g* and 95% confidence intervals (CI) for PTSS, depression, anxiety, and grief. According to Cohen's (1992) classification for Hedges' *g*, an effect size of 0.20–0.50 indicates a small effect, 0.50–0.80 indicates a medium effect, and ≥0.80 indicates a large effect. Effect sizes were computed for comparisons of pre–post data for within group effects and intervention/control data for between group effects for the outcomes of PTSS, depression, anxiety and grief. Intention-to-treat data were used when available. If the pre–post correlation required for calculating the pre–post effect size was not available, it was estimated using the overall mean of included studies that had available correlations for the respective outcome. A random-effects model was assumed due to heterogeneity in the samples (Hedges & Vevea, 1998), which was verified by Q statistic. In cases where the Q statistic was not statistically significant, a fixed-effects model was stated, and effect sizes were reported accordingly. Heterogeneity was estimated using $I^2$ (Higgins et al., 2003).

To minimise bias due to different control groups we calculated subgroup analyses by instrument and DSM version only in the pre-/post-data. We computed no subgroup analysis by outcome measures in the experimental/control group data, since control groups differ across the studies causing even more heterogeneity (AT, TAU, waitlist). A further subgroup analysis was conducted in the pre-/post-data by study design to assess whether non-controlled and controlled trials differed in effect size. We also calculated a subgroup analysis by control group (treatment as usual (TAU)/active treatment (AT); waitlist) in the experimental/control group data. When multiple control conditions were reported, we always used the stricter control group, with the preference being AT over TAU and TAU over waitlist. If there is a significant difference between effect sizes in subgroups, a sensitivity analysis was performed to identify potential moderators that may contribute to this difference such as age, gender, number of sessions, treatment modality (individual vs. group) and study design (controlled vs. uncontrolled). Subgroup analyses were only performed when at least three TF-CBT conditions or three post-treatment comparisons were available.

For all subgroup analyses, we assumed a common among-study variance component across all subgroups and therefore, pool within-group estimates of tau-squared for all studies.

We further used the Trim and Fill Method (Duval & Tweedie, 2000) to control for publication bias. If necessary, we imputed studies missing to the left of

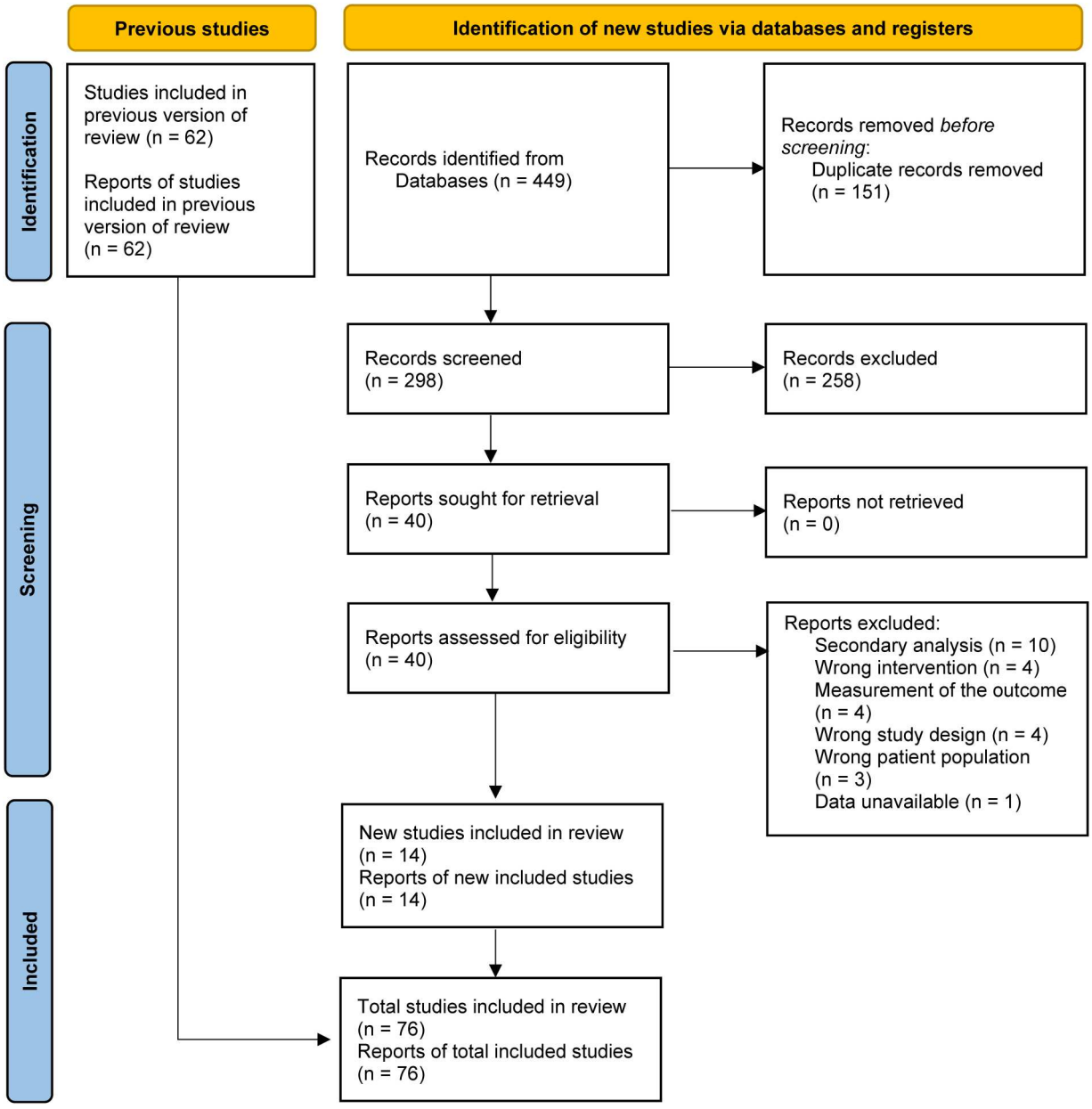


**Figure 1.** PRISMA flow diagram of study selection.

the mean and corrected the effect size estimate. Publication bias was assessed for all outcomes and subgroups if at least 10 independent data points were provided (Sterne et al., 2011).

To provide a tabular overview of the quality of evidence regarding TF-CBT for children and adolescents, we additionally produced a summary of findings table according to the GRADE working group (see Supplement S1) using GRADEpro (*GRADEpro GDT*, 2024; Schünemann et al., 2013).

## 3. Results

### 3.1. Study selection

In total, 76 studies met the eligibility criteria. Of those, 35 were RCT designs. An overview of the study selection is shown in Figure 1. An overview of included studies can be found in Supplement S2. The search revealed no additional studies reporting grief as a secondary outcome. We therefore were not able to update the effect sizes on grief and refer to the previous meta-analysis.

Most studies used an individual treatment ($n = 61$, of which 25 were RCTs) and some were group settings ($n = 15$, of which 10 were RCTs). In seven RCTs, the control condition was waitlist, while 22 RCTs were controlled using a TAU/AT condition. The control conditions of the remaining six RCTs either were a second TF-CBT condition and thus, included as an additional treatment condition or were excluded for other reasons. A variety of outcome measures for PTSS was used: Most commonly the UCLA PTSDI-RI ($n = 34$) and the CPSS ($n = 14$) were used. The K-SADS ($n = 8$), CATS ($n = 6$), CAPS-CA ($n = 5$), TSCC-PTS ($n = 4$; Briere, 1996) were also used

multiple times. NSESS (Kilpatrick et al., 2013), CITES II – PTSD (Wolfe, 2002), CROPS (Greenwald, 1999), ADIS (Silverman & Albano, 1996), CANS-TSS (Lyons, 2004) were each used once.

A wider range of measures was used for the secondary outcomes (see Supplement S2). A total of 15 different outcome measures was used for depression. The most used measures were the Children's Depression Inventory (CDI; $n = 24$) and the Mood and Feelings Questionnaire (MFQ, $n = 4$) or its short version (SMFQ; $n = 8$), while the remaining measures were each used once or twice. Similarly, twelve different outcome measures were used for anxiety, where the most common measures were the Screen for Child Anxiety Related Emotional Disorders (SCARED; $n = 13$) and the State-Trait Anxiety Inventory for Children (STAIC; $n = 4$). The outcome grief was hardly reported and measured with three different outcome measures: The Inventory of Complicated Grief (ICG; $n = 4$), the Extended Grief Inventory-Traumatic Grief subscale (EGI-TG; $n = 3$) and the Grief Screening Scale (GSS; $n = 1$). Accordingly, no subgroup analysis by instrument was conducted for this outcome.

Not all these studies met the quality criteria. We identified five RCTs and 25 uncontrolled studies that were considered to have a high risk of bias and therefore excluded from further analysis. If not stated otherwise, all reported effect sizes below refer to the analyses excluding high risk of bias studies. The risk of bias rating results and decisions for each domain are provided in Supplement S3.

Accordingly, five instruments (CAPS-CA, CATS, CPSS, K-SADS, UCLA PTSD-RI) and two different underlying DSM versions (DSM-IV, DSM-5) were included into subgroup-analysis. No publication biases were found in the overall results or in any subgroups.

### 3.2. Outcome analyses

Subgroup analysis by outcome measure is shown in Table 2 and subgroup analysis by underlying DSM criteria is shown in Table 3. In terms of outcome measures for PTSS subgroup analysis, a significant difference in effect sizes ($Q = 10.92$, $p < .05$) was found. The effect sizes of different outcome measures vary between $g = 1.06$ and $g = 1.69$. Among these, the effect sizes measured with CAPS-CA ($g = 1.30$) and K-SADS ($g = 1.32$) were comparable to the overall within-group effect size ($g = 1.32$), while those measured with UCLA ($g = 1.06$) were lower and those with CATS ($g = 1.69$) and CPSS ($g = 1.61$) were higher than the overall effect size. For the most part, the confidence intervals for each subgroup overlap, except for two outcome measures. The confidence intervals of UCLA PTSD-RI ($g = 1.06$, 95%-CI .83–1.29) and CPSS ($g = 1.61$, 95%-CI 1.30–1.93) are close but do not overlap, indicating a difference in the true effect size measured by the two instruments. A sensitivity analysis (provided in Supplement S4) comparing studies using the CPSS with studies using the UCLA PTSD-RI showed no differences in mean age and gender of the participants as well as mean number of sessions, treatment modality (group/individual) and study design (controlled/non-controlled). There was no significant subgroup analysis by outcome measure for other outcomes. The subgroup analysis for diagnostic criteria comparing treatment effects measured with DSM-IV vs. DSM-5 criteria showed no significant difference ($Q = .20$, $p = .65$) indicating a common true effect size of both subgroups.

Regarding the secondary outcomes, the variety of different outcome measures was too heterogeneous to enable comparison of those, respectively.

**Table 2.** Pre-post within-group effect sizes for PTSS, depression, and anxiety by instrument.

| Outcome | Instrument | *N* | *g* | 95% CI | SE | *z* | *Q* | $I^2$ | Duval and Tweedie |
|---|---|---|---|---|---|---|---|---|---|
| PTSS | All instruments | 55 | 1.32 | 1.15–1.49 | .09 | 15.44*** | 617.41*** | 91.25 | 0 |
| | CAPS-CA | 4 | 1.30 [a] | 1.08–1.52 | .11 | 11.39*** | 11.34 | 73.54 | – |
| | CATS | 6 | 1.69 | 1.23–2.14 | .23 | 7.22*** | 58.99*** | 91.52 | – |
| | CPSS | 13 | 1.61 | 1.30–1.93 | .16 | 10.09*** | 78.43*** | 84.70 | 0 |
| | K-SADS | 8 | 1.32 | .91–1.73 | .21 | 6.24*** | 122.58*** | 94.29 | – |
| | UCLA | 23 | 1.06 | .83–1.29 | .12 | 8.92*** | 113.74*** | 80.66 | 0 |
| | Total between | | | | | | 10.92* | | |
| Depression | All instruments | 35 | .77 | .60–.92 | .08 | 9.33*** | 209.20*** | 83.75 | 0 |
| | CDI | 19 | .60 | .41–.78 | .09 | 6.37*** | 81.82*** | 78.00 | 0 |
| | MFQ | 4 | .81 | .41–1.21 | .20 | 3.91*** | 20.54*** | 85.40 | – |
| | SMFQ | 4 | .75 | .35–1.15 | .21 | 3.66*** | 15.80** | 81.01 | – |
| | Total between | | | | | | 1.14 | | |
| Anxiety | All instruments | 23 | .66 | .46–.86 | .10 | 6.421*** | 120.00*** | 81.67 | 0 |
| | SCARED | 11 | .60[a] | .49–.70 | .05 | 11.18*** | 10.66 | 6.22 | 0 |
| | STAIC | 4 | .44 | .22–.65 | .11 | 3.97*** | 8.42* | 64.36 | – |
| | Total between | | | | | | 1.56 | | |

Note. High risk of bias studies are excluded; n: number of included TF-CBT conditions; PTSS: Posttraumatic Stress Symptoms; CAPS-CA: Clinician-Administered PTSD Scale for Children and Adolescents; CPSS: Child PTSD Symptom Scale; K-SADS: Schedule for Affective Disorders and Schizophrenia for School-Age Children; UCLA: UCLA Posttraumatic Stress Disorder-Reaction Index; CDI: Children's Depression Inventory; (S)MFQ: Mood and Feelings Questionnaire, (Short Version); SCARED: Screen for Child Anxiety Related Emotional Disorders; STAIC-T: State-Trait Anxiety Inventory for Children, Trait subscale; *$p < .05$; **$p < .01$; ***$p < .001$.

[a]Fixed model assumed due to non-significant *Q*-value.

**Table 3.** Pre-post within-group effect sizes for PTSS, depression, and anxiety by DSM criteria.

| Outcome | DSM | *n* | *g* | 95% CI | SE | *z* | *Q* | $I^2$ | Duval and Tweedie |
|---|---|---|---|---|---|---|---|---|---|
| PTSS | DSM-IV | 37 | 1.28 | 1.07–1.49 | .11 | 12.11*** | 308.78*** | 88.34 | 0 |
| | DSM-5 | 16 | 1.36 | 1.05–1.68 | .16 | 8.56*** | 281.00*** | 94.66 | 0 |

Note. High risk of bias studies are excluded; *n*: number of included comparisons; PTSS: Posttraumatic Stress Symptoms; *$p < .05$; **$p < .01$; ***$p < .001$.

### 3.3. Effect sizes of TF-CBT on primary and secondary outcomes

Across all included TF-CBT conditions, the mean pre–post within-group effect sizes was large for PTSS ($g$ =1.32, 95%-CI 1.15–1.49) and medium for depression ($g$ = .77, 95%-CI .60–.92) and anxiety ($g$ = .66, 95%-CI .46–.86). The meta-analysis showed similar results for RCTs only (PTSS: $g$ = 1.36, 95%-CI 1.12–1.60; depression: $g$ = .74, 95%-CI .53–.96; anxiety: $g$ = .58, 95%-CI .37–.79). Heterogeneity for all outcomes was high ($I^2$ = 81.67–91.18). An overview of within-group effect sizes and heterogeneity is shown in Table 4.

Comparing TF-CBT conditions to any control condition, effect sizes favouring TF-CBT were medium for PTSS ($g$ = .57, 95%-CI .34–.80) and depression ($g$ = .50, 95%-CI .19–.80) and low for anxiety ($g$ = .22, 95%-CI .03-.41). Compared to TAU or AT conditions, effect sizes are small but significantly favouring TF-CBT as treatment for PTSS ($g$ = .42, 95%-CI .17–.67) and depression ($g$ = .50, 95%-CI .15–.84), but not for anxiety ($g$ = .16, 95%-CI −.03–.34). Heterogeneity for all outcomes but anxiety was high ($I^2$ = 79.54–86.78; anxiety $I^2$ = 42.35). An overview of between-group effect sizes and heterogeneity is shown in Table 5.

## 4. Discussion

This meta-analysis evaluated the differences of TF-CBT treatment effects using various outcome measures, as well as the treatment effects of TF-CBT on PTSS as the primary outcome and depression and anxiety as secondary outcomes. In total, 76 studies were included in this analysis. We found no difference in effect sizes between studies measuring PTSS with instruments based on DSM-IV versus DSM-5 criteria. Regarding individual outcome measures, we also found almost no differences in effect sizes. Yet, there was a small significant difference specifically for CPSS and UCLA PTSD-RI. However, this difference is best explained by chance or confounding factors. No significant differences in outcome measures were found for the other outcomes. In addition, the studies provide strong evidence for the effectiveness of TF-CBT for all outcomes. TF-CBT is superior to both waitlist control conditions and AT/TAU conditions.

To the best of our knowledge, this is the first meta-analysis that has specifically compared outcome measures in a consistent set of treatment conditions using the same manual. As TF-CBT has been extensively researched, we can draw conclusions regarding

**Table 4.** Pre-post within-group effect sizes for PTSS, depression, and anxiety.

| Outcome | Control group | *n* | *g* | 95% CI | SE | *z* | *Q* | $I^2$ | Duval and Tweedie |
|---|---|---|---|---|---|---|---|---|---|
| PTSS | All studies | 55 | 1.32 | 1.15–1.49 | .09 | 15.44*** | 617.41*** | 91.25 | 0 |
| | RCTs only | 38 | 1.36 | 1.12–1.60 | .12 | 11.22*** | 412.80*** | 91.04 | 0 |
| Depression | All studies | 35 | .77 | .60–.92 | .08 | 9.33*** | 209.20*** | 83.75 | 0 |
| | RCTs only | 26 | .74 | .53–.96 | .11 | 6.90*** | 175.02*** | 85.72 | 0 |
| Anxiety | All studies | 23 | .66 | .46–.86 | .10 | 6.421*** | 120.00*** | 81.67 | 0 |
| | RCTs only | 20 | .58 | .37–.79 | .11 | 5.45*** | 102.81*** | 81.52 | 0 |

Note. High risk of bias studies are excluded; n: number of included comparisons; PTSS: Posttraumatic Stress Symptoms; *$p < .05$; **$p < .01$; ***$p < .001$.

**Table 5.** Post-treatment between-group effect sizes for PTSS, depression, and anxiety.

| Outcome | Control Group | *n* | *g* | 95% CI | SE | *z* | *Q* | $I^2$ | Duval and Tweedie |
|---|---|---|---|---|---|---|---|---|---|
| PTSS | Any control | 29 | .57 | .34–.80 | .12 | 4.89*** | 211.86*** | 86.78 | 0 |
| | TAU/AT | 23 | .42 | .17–.67 | .13 | 3.34** | 152.09*** | 85.54 | 0 |
| | WL | 6 | 1.16 | .66–1.66 | .26 | 4.52*** | 41.25*** | 87.88 | – |
| | Total between | | | | | | 6.72* | | |
| Depression | Any control | 20 | .50 | .19–.80 | .16 | 3.18*** | 143.73*** | 86.78 | 0 |
| | TAU/AT | 16 | .50 | .15–.84 | .18 | 2.83** | 142.61*** | 89.48 | 0 |
| | WL | 4 | .51[a] | .29–.74 | .12 | 4.43*** | 1.10 | 0 | – |
| | Total between | | | | | | 0.00 | | |
| Anxiety | Any control | 16 | .22 | .03–.41 | .10 | 2.30* | 26.02* | 42.35 | 0 |
| | TAU/AT | 12 | .16 | -.03–.34 | .09 | 1.69 | 21.73* | 49.38 | 0 |
| | WL | 4 | .38[a] | .15–.60 | .38 | 3.30*** | 2.76 | 0 | – |
| | Total between | | | | | | 1.28 | | |

Note. High risk of bias studies are excluded; *n*: number of included comparisons; PTSS: Posttraumatic Stress Symptoms; TAU/AT: Treatment as usual/active treatment control conditions; WL: Wait-list control conditions. *$p < .05$; **$p < .01$; ***$p < .001$.

[a]Fixed model assumed due to non-significant *Q*-value.

outcome measures. The gold standard measure CAPS-CA (Pynoos et al., 2015) was unfortunately not used as much as the other measures. Nevertheless, treatment measured with CAPS-CA could be considered as being less biased by diagnostic instrument. This meta-analysis identified instruments that are more (CPSS, CATS) and less (UCLA PTSD-RI) sensitive to change than the CAPS-CA. Especially the difference between UCLA and CPSS was notable. However, there is no obvious reason for this difference as both instruments have similar characteristics regarding assessment mode, response format, number of items and reference period of symptoms (one month) and both correspond to the same DSM version. The included samples were also similar regarding other potential influences such as age, gender, number of sessions, treatment modality and research design. Consequently, this finding may be attributed to chance, as the CIs are closely aligned, and the inclusion of a single additional data point could potentially alter the outcome. If there is indeed a true difference, it might be explained by an insufficient concordance between self-report and interview. On the other hand, it is hardly possible to draw a clear dividing line in research with children and young people, as self-report measures are often conducted as a semi-structured interview if the children are too young to answer the questionnaire themselves. Other factors that might be helpful to explain this finding are the number of traumatic events and the treatment dosages in the respective samples. Unfortunately, this data was not sufficiently reported for analysis. Yet another possible explanation might be that some items in the UCLA lack face validity, despite being assigned to specific symptoms (e.g. 'I have thoughts like I am bad', D2). Thus, the lower effect size may also be explained by these rather broadly defined items. More data is needed to examine the different instruments in depth. Although no difference was found between the other outcome measures for depression, and anxiety, drawing definite conclusions is difficult due to the limited number of available studies.

There had been no previous meta-analysis for children and adolescents comparing DSM versions of PTSD. Danzi and La Greca (2016) found that the DSM-5 generally identified lower rates of PTSD than the DSM-IV. Only about half of the children with a DSM-IV PTSD diagnosis were also identified according to DSM-5. However, the children identified with the DSM-5 could, for the most part, also be identified with the criteria of the DSM-IV. This can be explained with the additional symptom cluster (negative alterations in cognition/mood) in DSM-5 compared to the three-cluster model of DSM-IV. Our meta-analysis indicates that there is no significant difference in the treatment effects measured based on DSM-5 versus DSM-IV, which suggests that this issue therefore is not as relevant in meta-analyses comparing treatment effects. However, if the diagnosis is an inclusion criterion, the sample composition may differ because not all participants with a DSM-IV PTSD diagnosis might also be diagnosed according to DSM-5 instruments. This might also be true for ICD-10 and ICD-11 with varying concordance rates with the DSM IV and DSM 5 (Danzi & La Greca, 2016; Eilers et al., 2020). Thus, this issue should be considered when pooling studies using different diagnostic inclusion criteria.

Regarding TF-CBT treatment effects on PTSS, the newly added studies did not substantially change effect sizes found in our previous analysis (Thielemann et al., 2022), and effects are in line with other meta-analysis showing very large effect sizes for TF-CBT compared to waitlist conditions and medium effect size for AT (Gutermann et al., 2016; Mavranezouli et al., 2020; Morina et al., 2016). Regarding depression and anxiety, TF-CBT was shown to be superior to control conditions, consistent with previous findings reporting medium effect sizes (Gutermann et al., 2016). This underlines TF-CBT as a first-line treatment for PTSS among children and adolescents. An in-depth discussion of treatment effects can be found in our previous meta-analysis.

### 4.1. Limitations

While the assumption of TF-CBT being similarly effective in all studies is a prerequisite for the presented analyses on outcome measures, it needs to be reviewed with some caution. We minimised bias by excluding high risk of bias studies and were able to include enough studies allowing conclusions on outcome measures. However, there still was substantial heterogeneity within all subgroups and therefore, the assumption that therapy would be similar effective is somewhat vulnerable as other confounders may be influencing the effect sizes. The measures for example were used in interview or self-report formats. Self-perception and interviewer ratings may not coincide causing more heterogeneity. The number of traumatic events, treatment dosage, therapist training and therapists' supervision frequency might also influence treatment effects but were oftentimes not reported. Not accounting for these variables that might influence the effects of the intervention could bias the subgroup analyses. Another limiting factor is the simple search strategy used. As only keywords and not MeSH terms were used, it is possible that the strategy is lacking synonyms. Therefore, some studies may have been overlooked. MeSH terms could also narrow down the search string and potentially make it easier to replicate the procedure. One statistical limitation of the analysis was that it did not allow for the establishment of equivalence of effects, but only for the identification of differences. More research is needed to determine

if the outcome measures are indeed equivalent. Furthermore, since we only looked at it in the specific context of TF-CBT, generalizability is limited. Conclusions on outcome measures of secondary outcomes were not possible due to limited number of studies. For the limitations concerning TF-CBT effectiveness, the reader is referred to our previous analysis (Thielemann et al., 2022, 2023).

### 4.2. Conclusion and Implications

This meta-analysis found no differences in measuring treatment effects on PTSS according to DSM-IV and DSM-5 based instruments. Although the changes from DSM-IV to DSM-5 and their comparability were critically discussed, the effect sizes for outcomes measures based on the different version were very similar. Thus, it seems unproblematic to combine data from both in further meta-analyses. On the level of outcome measures, there may be some variance in the sensitivity to change of diagnostic and screening instruments, but the overall results are satisfactory for combining them in meta-analysis. The significant difference between the UCLA and CPSS is probably best explained by chance or potential unknown moderators. More data is needed to draw definite conclusions. From a clinical perspective, all the presented outcome measures are good and reliable options to assess PTSS. The choice of outcome measure may be based on the sensitivity needed in a given setting. In terms of diagnosis, the CAPS-CA is known as gold standard for pediatric PTSD and should be used accordingly. In line with the literature and our previous meta-analysis, this meta-analysis also confirms TF-CBT as a treatment of first choice for PTSS in children and adolescents as it is superior in reducing PTSD, depression, and anxiety compared to control conditions.

Future research should be aware of the issue of different outcome measures and control for different measures used, as they may be able to explain some variance in data. Moreover, in future publications, it would be beneficial to report information such as the number of traumatic events, the treatment dose, and other relevant variables more often as these variables may serve as moderators. Furthermore, the results of this meta-analysis should be replicated with other therapy formats and outcomes. A similar analysis in adult PTSD would also be desirable. Finally, it should be emphasised that there is still a gap in research regarding the effects of TF-CBT on grief symptoms, as no additional studies have been conducted in the last two years.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – project number 512640851.

## Data availability statement

Data are available on reasonable request.

## ORCID

*Bianca Schreyer* http://orcid.org/0009-0009-8485-8699
*Jonathan Felix Benjamin Thielemann* http://orcid.org/0000-0002-8719-5167
*Barbara Kasparik* http://orcid.org/0000-0003-1814-6879
*Rita Rosner* http://orcid.org/0000-0002-7960-8398

## References

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2022). *Comprehensive meta-analysis* (Version 4) [Computer software].

Briere, J. (1996). *Trauma symptom checklist for children: Professional manual.* Psychological Assessment Resources Inc.

Cohen, J. (1992). Statistical power analysis. *Current Directions in Psychological Science*, *1*(3), 98–101. https://doi.org/10.1111/1467-8721.ep10768783

Cohen, J. A., & Mannarino, A. P. (1993). A treatment model for sexually abused preschoolers. *Journal of Interpersonal Violence*, *8*(1), 115–131. https://doi.org/10.1177/088626093008001009

Cohen, J. A., Mannarino, A. P., & Deblinger, E. (2006). *Treating trauma and traumatic grief in children and adolescents*. The Guilford Press.

Cohen, J. A., Mannarino, A. P., & Deblinger, E. (2017). *Treating trauma and traumatic grief in children and adolescents* (2nd ed.). Guilford Publications.

Danzi, B. A., & La Greca, A. M. (2016). DSM-IV, DSM-5, and ICD-11: Identifying children with posttraumatic stress disorder after disasters. *Journal of Child Psychology and Psychiatry*, *57*(12), 1444–1452. https://doi.org/10.1111/jcpp.12631

Danzi, B. A., & La Greca, A. M. (2021). Treating children and adolescents with posttraumatic stress disorder: Moderators of treatment response. *Journal of Clinical Child & Adolescent Psychology*, *50*(4), 510–516. https://doi.org/10.1080/15374416.2020.1823849

Deblinger, E., & Heflin, A. H. (1996). *Treating sexually abused children and their nonoffending parents: A cognitive behavioral approach. Interpersonal violence*. Sage Publications.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Eilers, R., Rimane, E., Vogel, A., Renneberg, B., Steil, R., & Rosner, R. (2020). The impact of the new ICD-11 criteria on abused young people: 30% less PTSD and CPTSD diagnoses compared to DSM-IV. *Psychotherapy and Psychosomatics*, *89*(1), 59–61. https://doi.org/10.1159/000503794

Foa, E. B., Asnaani, A., Zang, Y., Capaldi, S., & Yeh, R. (2018). Psychometrics of the child PTSD symptom scale

for DSM-5 for trauma-exposed children and adolescents. *Journal of Clinical Child & Adolescent Psychology*, *47*(1), 38–46. https://doi.org/10.1080/15374416.2017.1350962

Forbes, D., Bisson, J. I., Monson, C. M., & Berliner, L. (Eds.). (2020). *Effective treatments for PTSD: Practice guidelines from the international society for traumatic stress studies* (3rd ed.). The Guilford Press.

GRADEpro GDT: GRADEpro Guideline Development Tool [Software]. (2024). McMaster University and Evidence Prime. gradepro.org

Greenwald, R. (1999). Assessment of posttraumatic symptoms in children: Development and preliminary validation of parent and child scales. *Research on Social Work Practice*, *9*(1), 61–75. https://doi.org/10.1177/104973159900900105

Gutermann, J., Schreiber, F., Matulis, S., Schwartzkopff, L., Deppe, J., & Steil, R. (2016). Psychological treatments for symptoms of posttraumatic stress disorder in children, adolescents, and young adults: A meta-analysis. *Clinical Child and Family Psychology Review*, *19*(2), 77–93. https://doi.org/10.1007/s10567-016-0202-5

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods*, *3*(4), 486–504. https://doi.org/10.1037/1082-989X.3.4.486

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ (Clinical Research Ed.)*, *327*(7414), 557–560. https://doi.org/10.1136/bmj.327.7414.557

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2022). Cochrane handbook for systematic reviews of interventions. *Cochrane,* 2022. www.training.cochrane.org/handbook

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., Williamson, D., & Ryan, N. (1997). Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(7), 980–988. https://doi.org/10.1097/00004583-199707000-00021

Kilpatrick, D. G., Resnick, H. S., & Friedman, M. J. (2013). *Severity of posttraumatic stress symptoms – child age 11–17 (National Stressful Events Survey PTSD Short Scale [NSESSS]*. https://www.psychiatry.org/File%20Library/Psychiatrists/Practice/DSM/APA_DSM5_Severity-of-Post-Traumatic-Stress-Symptoms-Child-Age-11-to-17.pdf

Lyons, J. S. (2004). *Redressing the emperor: Improving our children's public mental health system*. Praeger.

Mavranezouli, I., Megnin-Viggars, O., Daly, C., Dias, S., Stockton, S., Meiser-Stedman, R., Trickey, D., & Pilling, S. (2020). Research review: Psychological and psychosocial treatments for children and young people with post-traumatic stress disorder: A network meta-analysis. *Journal of Child Psychology and Psychiatry*, *61*(1), 18–29. https://doi.org/10.1111/jcpp.13094

Morina, N., Koerssen, R., & Pollet, T. V. (2016). Interventions for children and adolescents with posttraumatic stress disorder: A meta-analysis of comparative outcome studies. *Clinical Psychology Review*, *47*, 41–54. https://doi.org/10.1016/j.cpr.2016.05.006

National Institute for Health and Care Excellence. (2018). *Post-traumatic stress disorder: NICE guideline*. https://www.nice.org.uk/guidance/ng116

Phoenix Australia Centre for Posttraumatic Mental Health. (2020). *Australian guidelines for the prevention and treatment of acute stress disorder, posttraumatic stress disorder and complex PTSD*. https://www.phoenixaustralia.org/australian-guidelines-for-ptsd/

Pynoos, R. S., & Steinberg, A. M. (2014). *The UCLA PTSD reaction index for children and adolescents - DSM-5*. Behavioral Health Innovations.

Pynoos, R. S., Weathers, F. W., Steinberg, A. M., Marx, B. P., Layne, C. M., Kaloupek, D. G., & Kriegler, J. A. (2015). *Clinician-administered PTSD scale for DSM-5 child/adolescent version*. www.ptsd.va.gov

Sachser, C., Berliner, L., Holt, T., Jensen, T. K., Jungbluth, N., Risch, E., Rosner, R., & Goldbeck, L. (2017). International development and psychometric properties of the Child and Adolescent Trauma Screen (CATS). *Journal of Affective Disorders*, *210*, 189–195. https://doi.org/10.1016/j.jad.2016.12.040

Schäfer, I., Gast, U., Hofmann, A., Knaevelsrud, C., Lampe, A., Liebermann, P., Lotzin, A., Maercker, A., Rosner, R., & Wöller, W. (2019). *S3-Leitlinie Posttraumatische Belastungsstörung*. Springer.

Scheeringa, M. S., Myers, L., Putnam, F. W., & Zeanah, C. H. (2012). Diagnosing PTSD in early childhood: An empirical assessment of four approaches. *Journal of Traumatic Stress*, *25*(4), 359–367. https://doi.org/10.1002/jts.21723

Schünemann, H., Brożek, J., Guyatt, G., & Oxman, A. (2013). *GRADE handbook for grading quality of evidence and strength of recommendations. Updated October* 2013. guidelinedevelopment.org/handbook

Silverman, W. K., & Albano, A. M. (1996). *Anxiety disorders interview schedule for DSM-IV, child and parent versions*. Psychological Corporation.

Steil, R., Fischer, A., & Rosner, R. (2021). Internationale und deutsche Leitlinien zur Behandlung der PTBS bei Kindern und Jugendlichen. *Kindheit Und Entwicklung*, *30*(3), 154–163. https://doi.org/10.1026/0942-5403/a000343

Sterne, J. A. C., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A.-W., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., Pigott, T. D., … Higgins, J. P. (2016). Robins-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ (Clinical Research Ed.)*, *355*, i4919. https://doi.org/10.1136/bmj.i4919

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H.-Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., … Higgins, J. P. T. (2019). Rob 2: A revised tool for assessing risk of bias in randomised trials. *BMJ (Clinical Research Ed.)*, *366*, l4898. https://doi.org/10.1136/bmj.l4898

Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P. A., Terrin, N., Jones, D. R., Lau, J., Carpenter, J., Rücker, G., Harbord, R. M., Schmid, C. H., Tetzlaff, J., Deeks, J. J., Peters, J., Macaskill, P., Schwarzer, G., Duval, S., Altman, D. G., Moher, D., & Higgins, J. P. T. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ (Clinical Research Ed.)*, *343*(jul22 1), d4002. https://doi.org/10.1136/bmj.d4002

Thielemann, J. F. B., Kasparik, B., König, J., Unterhitzenberger, J., & Rosner, R. (2022). A systematic review and meta-analysis of trauma-focused cognitive

behavioral therapy for children and adolescents. *Child Abuse & Neglect*, *134*, 105899. https://doi.org/10.1016/j.chiabu.2022.105899

Thielemann, J. F. B., Kasparik, B., König, J., Unterhitzenberger, J., & Rosner, R. (2023). Stability of treatment effects and caregiver-reported outcomes: A meta-analysis of trauma-focused cognitive behavioral therapy for children and adolescents. *Child Maltreatment*, *29*(2), 375–387. https://doi.org/10.1177/10775595231167383

Veritas Health Innovation. (2014). *Veritas health innovation Covidence systematic review software* [Computer software]. www.covidence.org

Wolfe, V. V. (2002). *The children's impact of traumatic events scale II (CITES-II)* [Unpublished assessment instrument]. London Health Sciences.