Evaluating narrative exposure therapy for post-traumatic stress disorder and depression symptoms: A meta-analysis of the evidence base

Shruti Raghuraman | Nathan Stuttard | Nigel Hunt

Division of Psychiatry and Applied Psychology, School of Medicine, University of Nottingham, Nottingham, UK

Correspondence
Shruti Raghuraman, Division of Psychiatry and Applied Psychology, University of Nottingham, B14, Yang Fujia Building, Wollaton Rd, Nottingham NG8 1BB, UK.
Email: shruti.raghuraman@gmail.com

Narrative exposure therapy (NET) is an intervention for trauma spectrum disorders. Originally developed to treat refugee populations, NET has since been tested for efficacy across different settings. In this review, the NET evidence base is examined through a retrieval, synthesis and appraisal of randomized controlled trials (RCTs) published since 2002. Two independent reviewers (S. R. and N. S.) searched online databases including EMBASE, PsycINFO and PubMed. Twenty-four RCTs were selected for a meta-analysis of three outcomes: post-traumatic stress disorder (PTSD) diagnosis and PTSD and depression symptoms. All outcomes were analysed at short-term (3–4 months), midterm (6–7 months) and long-term (≥12 months) data points. A random-effects model was applied to yield standardized mean differences (SMDs) and odds ratios (ORs) as indicators of NET treatment effect. Subgroup analyses for type of trauma and type of control groups were conducted to examine potential heterogeneity. For the NET group, moderate effect sizes for PTSD symptom severity were observed at midterm and long term and at midterm for depression symptom severity. The number of PTSD diagnoses decreased significantly in the short term for the NET condition, but this was not sustained at the long term. Caution must be exercised when interpreting these results due to high heterogeneity estimates and low quality of evidence across trials. Potential small-study effects further complicate the interpretation of the findings. Recommendations are made for augmenting statistical significance research with qualitative analyses of NET efficacy to better inform clinical practice.

KEYWORDS
narrative exposure therapy, narrative therapy, NET, post-traumatic stress, psychological trauma, PTSD

1 | INTRODUCTION

Narrative exposure therapy (NET; Neuner, Schauer, Roth, & Elbert, 2002) is an evidence-based approach to treating psychological manifestations of trauma in survivors (Hamblen et al., 2019; Robjant...
& Fazel, 2010). On the basis of the principles of cognitive behavioural therapy (CBT), testimony therapy and exposure therapy, NET follows a manualized treatment protocol that is aimed at constructing a consistent, coherent autobiographical representation of the traumatic event(s) within the context of a narrative account of one’s life. This is thought to facilitate emotional processing of trauma memories to bring about improvement in emotional, cognitive and behavioural symptoms of trauma.

NET was developed specifically for use in low-resource settings and victims of organized and family violence. It is highly accessible because it can be delivered by non-mental health professionals following a short training programme using the manual developed by Schauer, Neuner, and Elbert (2005) and Schauer, Elbert, and Neuner (2011). A detailed description of the NET procedure can be found in the treatment manual (Schauer et al., 2005, 2011).

NET has been evaluated using randomized controlled trials (RCTs) primarily with refugee populations and asylum seekers, wherein the trauma experienced may be naturally caused or man-made. The focus in these situations is on the atrocities endured, usually at the hands of single or multiple perpetrators. Over time, NET has been used to treat other traumatized populations with or without a perpetrator, with acute or chronic post-traumatic stress disorder (PTSD) sufferers. This has been documented in RCTs situated in a range of sociocultural-economic settings.

International guidelines for PTSD treatment such as the American Psychological Association (APA), Department of Veterans Affairs and Department of Defense (VA/DoD), International Society for Traumatic Stress Studies (ISTSS) and National Institute for Health and Care Excellence (NICE) have included NET as part of their recommendations for intervention.

NICE recommends NET as a first-line treatment option for PTSD along with CBT, cognitive processing therapy (CPT) and prolonged exposure (PE). On the contrary, APA guidelines published a conditional recommendation for NET, claiming the evidence was insufficient for a strong recommendation (APA, 2017). Strong recommendations were made for CBT, CPT and PE. Similarly, VA/DoD found PE and eye movement desensitization and reprocessing (EMDR) to have the strongest evidence base, whereas NET was found to have ‘sufficient’ evidence for a strong recommendation (VA/DoD, 2017). ISTSS also provided a standard recommendation for NET (Berliner et al., 2019) and a strong recommendation for CPT, EMDR and trauma-focused CBT. As such, the evidence base for CBT, CPT, PE and EMDR appears to be stronger when compared with NET. A systematic evaluation of the NET evidence base is essential to better inform clinical and practice guidelines regarding its efficacy.

### 1.1 | Rationale for review

By publishing data about NET efficacy, one can advocate for its widespread implementation in low-resource settings to alleviate symptoms. Narrative reviews of NET efficacy have found NET favourable to controlled comparisons in reducing traumatic stress in a range of socio-economic and cultural contexts including low dropouts and sustained improvements over time (McPherson, 2012; Robjant & Fazel, 2010). In one meta-analysis of NET efficacy, authors found a medium effect size ($g = 0.63$) for PTSD symptom reduction post-intervention (Gwozdziewycz & Mehl-Madrona, 2013). This estimate was interpreted as evidence of NET efficacy, although it is not clear how the average effect size was calculated in terms of follow-up time points. The control groups used for comparisons with NET are varying and not restricted to ‘bona fide’ or active treatments, that is, treatments that were intended to be ‘therapeutic’ (Wampold & Imel, 2015). For example, RCTs used wait-list groups, no-treatment groups and supportive counselling as controlled comparisons to NET. Importantly, most reviews have not critically appraised the quality of the evidence, which makes these findings inconclusive regarding NET’s true treatment efficacy.

NET trials have been included in large meta-analytic studies of a range of psychological therapies for traumatic stress (Bisson, Roberts, Andrew, Cooper, & Lewis, 2013; Patel, Kellezi, & Williams, 2014). In these reviews, the methodological rigour of the included NET trials was questioned. Further, both reviews included only a handful of NET trials each ($N = 07$, Bisson et al., 2013; $N = 04$, Patel et al., 2014), and since their publication, several recent trials investigating NET efficacy have been published. This suggests a need to conduct a comprehensive, up-to-date quality appraisal of the NET evidence base to inform researchers and practitioners.

Only one other comparable meta-analytic review was identified at the time of this review (Lely, Smid, Jongedijk, Knipscheer, & Kleber, 2019). The authors of this review concluded that despite methodological weakness in the included trials, there is empirical support for NET as a trauma intervention. However, at least seven published RCTs of NET efficacy have not been included in the review, and the authors have used ‘last follow-up’ as the uniform follow-up time...
point. NET trials vary significantly in their measurement of outcomes, especially concerning data points. In the Lely, Smid, et al. (2019) meta-analysis, this ranged from 9–52 weeks. They also do not include PTSD diagnostic status (i.e., meeting the criteria for a diagnosis of PTSD) as a measured outcome, which could potentially be an important consideration when making a recommendation for clinical efficacy.

In this paper, a meta-analysis of NET efficacy is attempted using all available data from NET clinical trials published to date. The term ‘efficacy’ is used unequivocally throughout this paper as NET RCTs have predominantly evaluated the intervention under ideal and controlled circumstances (Roland & Torgerson, 1998). This is differentiated from evaluations of ‘effectiveness’ (a measure of intervention benefit under ‘real-world’ clinical conditions). Of note, some of the included trials as well as a previously published meta-analytic review (Lely, Smid, et al., 2019) have used the term ‘effectiveness’ interchangeable with ‘efficacy’, although the design of the trial is in line with an efficacy evaluation.

For the meta-analysis, the pooled intervention effect of NET will be estimated against controlled comparisons on outcomes of PTSD (symptom severity and diagnostic status) and depression symptom severity. Due consideration will also be given to issues of heterogeneity and methodological quality of the included trials.

2 | OBJECTIVES

The objective was to evaluate the efficacy of the NET evidence base for the treatment of trauma-related psychopathology on outcomes of PTSD and depression across different sociocultural contexts and trauma exposure.

3 | METHODS

Search results that fulfilled the selection criteria were scrutinized at the abstract and full-text stage independently by two reviewers (the first and second authors: S. R. and N. S.). Appropriate data were extracted independently by both reviewers, and disputes, if any, were resolved by the third author, N. H. The risk of bias was assessed by S. R. and N. S. in accordance to version 1.0 of the Cochrane risk of bias tool for randomized trials (Higgins & Green, 2011).

3.1 | Inclusion criteria

3.1.1 | Types of studies

Original RCTs using NET (or an adaptation of NET) were included. Inclusion criteria for the trial population were individuals with a history of exposure to trauma and reporting PTSD outcome measures (diagnostic status and/or symptom severity) following such exposure. No restrictions were placed based on the type of trauma experienced by the population. The search was not limited to RCTs with participants over 18 years of age. Instead, we chose to exclude studies that used the version of NET adapted for children below the age of 18 years: KIDNET (Onyut et al., 2005). From an initial scoping review of the NET literature, it was anticipated that some trials may include a combination of both adults and underage (<18 years) participants, especially in the case of refugees and asylum seekers. Further, KIDNET was developed and tested for efficacy merely 3 years since the first publication of NET (Neuner et al., 2002), which would make it the preferred choice for studies strictly recruiting participants under 18 years of age. By not placing an explicit age restriction and instead using the KIDNET filter, we aimed to include all available data published on NET trials conducted with adult participants.

Studies from any part of the world implementing any control comparisons were included. No restrictions were placed on the number of control comparisons. Outcome measures included PTSD (scales or diagnostic interviews) and depression symptoms.

3.1.2 | Timing of outcome assessment

End point assessments were used. NET does not have a fixed number of sessions or intervention duration, and we anticipated some variations in timings of outcome assessments across studies. As a result, we estimated effects at (a) short term (3–4 months), (b) midterm (6–7 months) and (c) long term (12 months or above).

3.2 | Search strategy

The local databases searched were PubMed, CINHAL, PsycINFO, Medline, Cochrane Library and Embase. The search strategy for PubMed is presented in Table A1 as an example. Additionally, the original authors of NET were contacted and asked to provide a full publication list of NET research. The date last searched is 8 December 2019.

3.3 | Selection and screening

Two authors (S. R. and N. S.) independently screened all titles and abstracts. Full-text articles were reviewed for inclusion, and data were extracted independently by S. R. and N. S. Cohen’s kappa statistic was calculated to indicate inter-rater agreement at the title and abstract screening stages (Cohen, 1960). Kappa values were interpreted using Landis and Koch’s (1977) guidelines. Included publications were extracted for relevant data, and disagreements were resolved by the third author (N. H.).

3.4 | Risk of bias

Two authors (S. R. and N. S.) separately assessed the risk of bias by using criteria according to the Cochrane Collaboration’s risk of bias
tool (Higgins & Green, 2011). At the time of completion of this review, there was no mandate to use the updated version (RoB 2.0; Sterne et al., 2019), and it was still being pilot tested by the Cochrane Review teams (Cochrane Library, 2019). Additionally, assessment issues and therapist qualifications were included as domains of risk. For assessment issues, criteria for low risk of bias included the use of valid and reliable measures or the use of translated measures with sufficient psychometric properties. In the absence of such information for translated versions, the risk was deemed unclear. When trials failed to report whether measures were valid and reliable, or when there was a lack of clarity regarding attempts to translate and back-translate standardized measures, the trials were considered to have a high risk of bias. For therapist qualifications, trials that reported details of the therapist’s NET training and qualifications qualified for a low risk of bias. If such information was found insufficient or lacking, the trials were rated as having an unclear risk of bias. If the intervention was delivered by untrained individuals with a lack of relevant qualifications, or if the study did not report any information at all, it was rated as having a high risk of bias.

Funnel plots were examined for comparisons with 10 or more trials to indicate publication bias in line with the rule of thumb recommended by the Cochrane Handbook for Systematic Reviews of Interventions (Higgins et al., 2019). Disagreements were resolved by the third author, N. H.

3.5 | Data synthesis

A random-effects (RE) model was used to calculate the standardized mean differences (SMDs) and 95% confidence intervals (CIs) for continuous outcomes. The odds ratio (OR) and 95% CIs were calculated by using an RE model for dichotomous variables. When trials reported multiple treatment arms, the non-NET active treatment arm was treated as a control group and was combined with the non-active treatment control group (such as wait-list or no-treatment controls) to generate pooled mean and standard deviation (SD) values. Study authors were contacted for missing data, such as means and SDs. If these data could not be obtained, those trials were excluded from the meta-analysis. The desktop version of Review Manager (RevMan, version 5.3) was used for data analysis (RevMan, 2014).

3.6 | Assessment of heterogeneity

Visual inspection of graphs, a Mantel–Haenszel $\chi^2$ statistic and $I^2$ statistic was used to test for heterogeneity. An $I^2$ estimate $\geq 50\%$ accompanied by a statistically significant $\chi^2$ statistic was interpreted as evidence of substantial levels of heterogeneity.

3.7 | Subgroup analysis

From a clinical perspective, combining studies that varied by the type of trauma experienced by the participants and the type of control conditions used in the analysis could impact the findings. To investigate the influence of such variability, two subgroup analyses were conducted. The first subgroup analysis was based on the type of control, that is, active treatment versus no treatment. Type of traumatic event was also used for a subgroup analysis between trauma with a perpetrator (such as war trauma, combat trauma, abuse and violence and torture) and trauma without a perpetrator (such as natural disasters and occupational trauma). Age is also a potential effect modifier. If enough studies ($N = 10$; Higgins et al., 2019) with participants below 18 years of age are found, a subgroup analysis based on age will be performed.

3.8 | Sensitivity analysis

The effects of excluding trials with a high risk of bias or those that appeared to be outliers upon a visual inspection of forest plots were conducted as part of the meta-analysis. When the removal of these studies did not change the direction or significance of the treatment effect, they were included in the final analysis.

4 | RESULTS

A total of 306 results were retrieved from the database search, and 110 were chosen after screening titles. Inter-rater agreement for screening titles was substantial ($k = 0.676$). Six studies were added from the NET reference database received from NET authors. After discarding duplicates ($N = 76$), the abstracts of 40 records were screened for inclusion. Fifteen studies were excluded due to the following reasons: (a) conference abstracts (02), (b) study protocols (03), (c) dissertation copy of included study (01), (d) KIDNET studies (03), (e) non-RCTs (02), (f) non-experimental study (01), (g) NET was not the intervention in question (02) and (h) PTSD not measured as an outcome (01). Inter-rater agreement for screening abstracts was almost perfect ($k = 0.896$).

Twenty-six full-text studies were retrieved for full-text review. Two studies had to be excluded at this stage; one study did not meet the randomization criteria (Crombach & Siehl, 2018), and the authors of the other study were unable to provide the required data for the meta-analysis (Hinsberger et al., 2017). Twenty-four trials were included in the final meta-analysis.

The PRISMA flow diagram in Figure B1 illustrates the selection of the studies included in the review.

4.1 | Included studies

A total of 1,391 participants were recorded in the trials included, with sample sizes ranging from 18 to 277. Studies were conducted across the world, including countries in Africa, Europe, Asia and North America. Most participants were survivors of war trauma and organized and/or personal violence, and in a few studies,
participants were survivors of natural disasters. The mean age of the participants across trials ranged from 17 to 70 years. NET was administered in the intervention arm in 18 studies. Four trials used an adaptation of NET known as the Narrative Exposure Therapy for Forensic Offender Rehabilitation (FONRET; Hecker, Hermenau, Crombach, & Elbert, 2015) meant for persons with a history of perpetrated violence. In two trials, different versions of brief NET were used. In one trial, a combination of NET and interpersonal psychotherapy (IPT) was used. The most common control condition was wait-list controls (WLC; 10 trials; see Table A2). An average of 7.58 sessions of NET therapy was delivered across all 24 included trials with a range of 14.2. All studies provided details of attrition during treatment except one (Morath, Moreno-Villanueva, et al., 2014). The rate of attrition varied from 0% to 38.64%, with a mean attrition rate of 7.43% during treatment. Studies that reported dropout data highlighted the sensitive nature of refugee and asylum status (such as camp closures, receiving asylum, disappearance and transfers) as common reasons for dropout. Other reasons included a lack of motivation, trust, and psychosocial problems. One study with abuse victims reported emotional suffering over treatment as a reason for dropout (Orang et al., 2018).

Twenty-three trials reported the severity of PTSD symptoms as one of the primary outcomes, and 13 trials reported PTSD diagnostic status as a measured outcome. The Clinician-Administered PTSD Scale (CAPS; American Psychiatric Association, 1994) was used to diagnose and assess the frequency and severity of PTSD symptoms in nine trials. Other measures that were used included the PTSD Symptom Scale Interview (PSS-I; Foa, Riggs, Dancu, & Rothbaum, 1993; Foa & Tolin, 2000), Composite International Diagnostic Interview (CIDI; World Health Organization, 1990), Post-Traumatic Stress Diagnostic Scale (PDS; Foa, Cashman, Jaycox, & Perry, 1997), Scale of Post-traumatic Stress Symptoms (SPTSS), Harvard Trauma Questionnaire (HTQ; Arabic version; Shoeb, Weinstein, & Mollica, 2007) and Impact of Events Scale-Revised (IES-R; Weiss & Marmar, 1997).

Depression symptom severity was most commonly measured in trials along with PTSD symptoms (N = 18). Hamilton Depression Rating Scale (HAM-D: Hamilton, 1960, 1967), Hospital Anxiety and Depression Scale (HADS; Zigmond & Snaith, 1983), Patient Health Questionnaire-9 (PHQ-9; Spitzer, Kroenke, & Williams, 1999), Beck Depression Inventory (BDI; Beck, Steer, & Brown, 1996), Self-Reporting Questionnaire 20 (SRQ-20; Harding et al., 1980), Mini-International Neuropsychiatric Interview (M.I.N.I.; Sheehan et al., 1998) and Hopkins Symptoms Checklist (HSCL; Derogatis, Lipman, Rickels, Uhlenhuth, & Covi, 1974) were all used to measure this outcome in the form of self-report as well as structured interviews.

For a detailed description of the studies, please refer to Table A2.

4.2 | Risk of bias

Figure B2 depicts the risk of bias assessment for the trials included in the meta-analysis.

4.2.1 | Random sequence generation

Recognized randomization procedures were used in 15 out of 24 trials, and these trials were rated as low risk. Nine trials were judged as reporting unclear randomization techniques. Some of these trials did not report the randomization technique, whereas others used techniques that may or may not ensure complete randomization (e.g., randomizing only a section of eligible participants and not the others and using restricted randomization, which could introduce selection bias). In two of these trials, assessment of baseline outcomes was used for the assignment of participants to treatment and control groups, which suggests that the sequence generation may not have been random (Crombach & Elbert, 2015; Hermenau, Hecker, Schaal, Maedl, & Elbert, 2013).

4.2.2 | Allocation concealment

Only two studies were judged to report adequate information about allocation concealment. The remaining 22 trials did not indicate information about allocation concealment. Most of these studies reported that groups did not significantly differ at baseline on outcomes or demographics, which could indicate adequate randomization. These studies were categorized as having unclear risk.

4.2.3 | Blinding of participants and personnel

Only one study reported blinding of participants to decrease the likelihood of further unblinding of outcome assessment. However, there was no mention of blinding personnel, as this is not possible in an RCT using a psychological intervention such as NET. This led to the trial being rated as unclear for performance bias (Jacob, Neuner, Maedl, Schaal, & Elbert, 2014). The remaining 23 trials were rated as having a high risk of performance bias.

4.2.4 | Blinding of outcome assessment

Seven trials were judged to be at high risk for detection bias due to not implementing appropriate blinding while assessing outcomes post-intervention and at follow-up. In four of these trials, patients accidentally revealed their treatment condition to the assessors. In two trials, the lead author, who was also one of the therapists, assessed outcomes, which placed these trials at high risk for detection bias. Four trials did not provide adequate information or suggested plausible, accidental unblinding and were rated as having unclear risk. The remaining studies were rated as having a low risk of bias.

4.2.5 | Incomplete outcome data

Four trials were at high risk for attrition bias due to analysing only treatment completers. In one trial that analysed only treatment
completers, authors found that demographics and study variables did not significantly predict treatment completers or dropouts in a logistic regression analysis (Orang et al., 2018). Further, the type of therapy did not significantly predict dropouts either. In a second trial, missing data from dropouts were replaced by estimation with a restricted maximum likelihood procedure, and no significant differences were found before and after data were replaced (Alghamdi, Hunt, & Thomas, 2015). Thus, the authors analysed only treatment completers. The authors in both trials did not publish these results in the final report. Therefore, these trials have been classified as having unclear risk. The remaining trials were classified as low risk due to either reporting no dropout or using some form of intention-to-treat (ITT) methods to account for missing data.

4.2.6 Selective reporting

Nine trials were judged as low risk for selective reporting, as all primary and secondary outcomes reported in the protocol were matched with those reported in the final publication. A further six trials with published protocols were judged as having unclear risk due to not reporting secondary outcomes mentioned in the protocol in the final publication. Other trials judged as unclear risk were due to the lack of a protocol available. Two studies were categorized as high risk for selective reporting due to not reporting primary outcomes indicated in the protocol. Only four out of eight comparisons had 10 or more trials contributing to the analysis (PTSD symptoms at short term and midterm and depression symptoms at short term and midterm). A visual examination of the funnel plots (see Figure B3) for these comparisons did not clearly indicate asymmetry.

4.2.7 Other sources of bias

4.2.7.1 Therapist qualification

Seven trials reported that the interventions were delivered by clinical psychologists/counsellors with NET and trauma experience. Nine trials reported that the therapist was a PhD/graduate student or therapist who had explicit NET training per the manual. These studies were classified as low risk. Two trials did not specify the training or manual adherence procedures of their therapists (clinical psychologists and clinical psychology doctoral students). Hence, these trials have been classified as unclear in terms of risk. Three trials used trained 'lay counsellors' to deliver therapy. In these trials, the qualifications of the therapists were unclear. One trial used trained final year undergraduate psychology students, whereas another used clinical psychology undergraduate degree holders who were trained in NET. These trials were classified as having unclear risk because NET is a manualized technique that does not require clinical or medical qualifications for its administration (Schauer et al., 2005). However, the lack of graduate qualification/healthcare training does not allow for these trials to be completely devoid of risk of bias.

4.2.7.2 Assessment issues

Only five studies used valid and reliable measures for the assessment of outcomes. These trials have been categorized as low risk. Eleven trials performed real-time translation and back-translation of outcome measures to capture the symptoms of PTSD and depression, which are primary outcomes of interest to this review. No information was provided about the psychometric properties of the translated measure. Further, translations and back translations achieve linguistic equivalence, which often subsumes the importance of cultural meaningfulness and appropriateness to the context. One trial used structured interviews using interpreters to measure outcomes (Neuner et al., 2010). In these cases, one cannot rule out the risk of the assessor’s influence in achieving the desired outcome. Therefore, these trials were classified as having an unclear risk of bias. A further eight trials did not report information regarding the translation of English language outcomes or the use of valid instruments to measure outcomes, which suggest these trials are at a high risk of bias. Figure B4 depicts the risk of bias summary for the trials included.

4.3 Effects of intervention

The main comparison was NET versus control for the treatment of PTSD and depressive symptoms and PTSD diagnostic status.

4.3.1 Comparison: NET versus any control

Twenty-four full-text studies were included in the meta-analysis. Due to different scales of assessment being used to measure the outcomes discussed, results were combined as SMDs. All 24 included studies contributed to the above comparison. This comparison had nine analyses, and the results have been summarized in Table A3.

4.3.1.1 PTSD diagnostic status

4.3.1.1.1 Short term (3–4 months). Four studies relevant to this outcome were identified (total N = 173; see Figure B5). There was a significant effect of NET intervention on diagnosis compared with control, with an OR of 0.28 (95% CI 0.12 to 0.66, Z = 2.89, p = .004). Heterogeneity was low at 3%.

4.3.1.1.2 Midterm (6–7 months). Four studies contributed to this outcome (total N = 150, see Figure B6). There was evidence of a significant effect of intervention on diagnostic status compared with control (OR 0.26; 95% CI 0.13 to 0.55, Z = 3.59, p = .0003). Heterogeneity was 0%.

4.3.1.1.3 Long term (≥12 months). Three relevant studies with a total of 190 participants were identified (see Figure B7). NET did not significantly outperform controls in this analysis (OR 0.68; 95% CI 0.16 to 2.87, Z = 0.51, p = .60). Further, significantly high levels of heterogeneity were found (χ² = 9.96; df = 2.0; p = .0071; I² = 80%). One study that appeared to be an outlier (Neuner, Schauer, Klaschik, Karunakara,
Seven relevant studies contributed to this outcome, with a total of 813 participants (see Figure B8). Small effect size evidence that NET was significantly different in its effects compared with controlled comparisons was found (SMD = −0.30, 95% CI: −0.49 to −0.11, Z = 3.05, p = .002). Moderate but non-significant heterogeneity was found ($\chi^2 = 21.62, df = 14 [p = .09]; I^2 = 35\%$). A sensitivity analysis was conducted in which an outlier study was removed from the analysis (Adenauer et al., 2011). This study also had a high risk of bias in three domains. This reduced the heterogeneity ($I^2 = 6\%$) and did not alter the significance of treatment effect (SMD = −0.21, 95% CI: −0.37 to −0.06, $Z = 2.78, p = .005$). Therefore, it was considered in the analysis due to not altering the direction/significance of the treatment effect.

### 4.1.1.2 Midterm (6–7 months)

Fourteen trials across 13 studies relevant to this outcome (total $n = 763$; see Figure B9) were identified. NET emerged significantly superior to controlled interventions with a medium effect size (SMD = −0.45, 95% CI: −0.68 to −0.23, $Z = 4.01, p < .0001$). Moderate yet significant levels of heterogeneity were found ($\chi^2 = 23.29, df = 13 [p = .04]; I^2 = 44\%$). A visual inspection of the forest plot suggested that point estimates across trials were not drastically different. The removal of the high risk of bias trials (with three or more domains showing a high risk) slightly increased heterogeneity estimates ($\chi^2 = 18.22, df = 10 [p = .05]; I^2 = 45\%$) but did not alter the direction or significance of treatment effect. As a result, they were considered in the final analysis.

### 4.1.1.3 Long term (≥12 months)

Seven relevant studies contributed to this outcome with a total of 302 participants (see Figure B10). NET emerged superior to control interventions (SMD = −0.49, 95% CI: −0.8 to −0.18, $Z = 3.06, p = .002$) with a medium effect size. Moderate, statistically non-significant heterogeneity was noted among studies ($\chi^2 = 9.19, df = 6 [p = .16]; I^2 = 35\%$). Two outliers were identified using the forest plot (Köbach, Schaal, Hecker, & Elbert, 2017; Neuner et al., 2004), and their removal lowered heterogeneity estimates ($\chi^2 = 1.25, df = 4 [p = .87]; I^2 = 0\%$). The removal of these trials did not alter the significance or direction of the effect (SMD = −0.28, 95% CI: −0.55 to 0.01, $Z = 2.03, p = .04$) and hence will be considered for the final analysis.

### 4.3.2 | Subgroup analysis

#### 4.3.2.1 Types of controls

To account for the clinical heterogeneity in the type of control conditions that NET was compared with across trials, a subgroup analysis of active controls (e.g., other psychological therapies, treatment-as-usual and psychoeducation) versus inactive control (e.g., no treatment and WL) was performed at the three time points for PTSD and depression symptoms.

Subgroup differences were significant only for PTSD symptom severity at the short term. Categorizing studies by type of intervention, NET demonstrated a difference between subgroups ($p = .04$) and reduced heterogeneity in the no intervention group ($N = 407, I^2 = 23\%$). In this group, there was evidence of a medium effect of NET (SMD = −0.44, 95% CI: −0.68 to −0.21, $p = .0002$). In contrast, the active intervention group demonstrated a lack of effect ($N = 454, SMD = −0.07, 95% CI: −0.34 to 0.20, p = .62$). Within the active intervention subgroup, substantial, non-significant heterogeneity estimates were revealed ($p = .12, I^2 = 40\%$).
For depression symptoms at midterm, NET significantly outperformed active interventions ($SMD = -0.50$, $95\% CI -0.84$ to $-0.16$, $p = .004$), whereas a non-significant overall effect was found when compared with the no intervention subgroup. However, the subgroup differences did not reach statistical significance ($p = .27$). In this analysis, the active intervention group was characterized by significant heterogeneity ($\chi^2 = 21.28; df = 9; p = .0; I^2 = 58\%$), and the no intervention subgroup consisted of only three studies.

Subgroup differences for PTSD and depressive symptoms were not significant at any other time point based on the type of control groups (see Table A4).

### 4.3.2.2 Type of trauma

For the purpose of this analysis, types of trauma varied between trauma with a perpetrator and trauma without a perpetrator in trials, and these categories were compared in a subgroup analysis. Trauma without a perpetrator included singular events such as natural disasters or trauma induced by occupational stress. Trauma with a perpetrator consisted of either repetitive or consistent trauma or threat of trauma such as war, combat, torture and abuse.

Subgroup differences for PTSD and depressive symptoms were not significant at the short-term and midterm time points based on the type of trauma. PTSD and depressive symptoms in the long term could not be analysed for subgroup differences based on trauma due to a lack of relevant trials in the simple trauma subgroup (see Table A5).

### 4.3.2.3 Age

There was insufficient data to perform a subgroup analysis based on age of participants.

## 5 | DISCUSSION

### 5.1 | Summary of main results

A meta-analysis of NET treatment effect on PTSD diagnosis and symptoms and depressive symptoms yielded partial evidence of efficacy. The effect size of treatment on PTSD symptoms is small in the short term ($SMD = -0.30$, $95\% CI -0.49$ to $-0.11$) but medium at the midterm ($SMD = -0.45$, $95\% CI -0.68$ to $-0.23$) and long term ($SMD = -0.49$, $95\% CI -0.80$ to $-0.18$), suggesting sustainable treatment gains. These findings are consistent with other narrative analyses of NET efficacy (Mundt, Wünsche, Heinz, & Pross, 2014). Conversely, a statistically significant effect of NET was found on PTSD diagnostic status in the short-term and midterm time points, but these effects were not sustained at long-term follow-up. It is important to note that the outcome at long term was characterized by very high heterogeneity estimates. Further, only a small number of trials contributed to the analyses for PTSD diagnostic status at all time points ($N < 5$), which impacts the findings for this outcome. When RE models are used, including a substantial number of studies is necessary in order to make reliable inferences, especially in the presence of high between-study heterogeneity (Guolo & Varin, 2017; Seide, Röver, & Friede, 2019). This significantly impacts the findings for this outcome.

In terms of depression symptoms, NET emerged statistically superior to controlled comparisons only at midterm with a medium effect size ($SMD = -0.49$, $95\% CI -0.79$ to $-0.20$). Lely, Smid, et al. (2019) found that NET outperformed non-active controls with medium-to-large effect sizes ($g = 0.79$) for depression symptoms. In our study, this was not reflected in the subgroup analysis for types of control for depression symptoms at any time point. However, in line with their overall conclusion, it appears that NET is relatively less effective in treating depressive symptomatology.

Subgroup analyses for type of controls emerged significant only for PTSD symptoms at the short term, with NET performing better than no intervention controls, when compared with active interventions. Within the active intervention subgroup, a non-significant overall effect of NET was found ($p = .62$). Although this could imply that NET does not outperform active interventions as efficaciously as WLC and no-treatment controls, the wide $95\%$ CIs of the point estimate ($-0.07, 95\% CI -0.34$ to $0.20$) suggest caution in the interpretation of the summary statistic for this subgroup. The relatively high yet non-significant heterogeneity estimates ($I^2 = 40\%, p = .12$) also suggest that within the active intervention group, there is potential for further subgroup analysis based on the type of the controls. Subgroup differences were not statistically significant for either outcome at any other time point. Lely, Smid, et al. (2019) found that NET significantly outperformed only non-active controls for both PTSD and depression symptoms. However, they do not specify the follow-up timings for controlled comparisons in their study, thereby limiting the comparability of these findings. Regarding type of trauma, the lack of subgroup differences suggest that NET is similarly efficacious across a range of trauma populations.

Conclusions about the evidence of treatment effect (especially when applied to clinical practice) must be made with caution due to high heterogeneity estimates when the data were statistically pooled. An exploration of heterogeneity using the visual inspection method allowed us to identify outliers, and the removal of these trials did not reduce heterogeneity in all trials. This is especially true in the case of PTSD symptoms at 6 months (significant effect) and PTSD diagnosis at 9–12 months (non-significant effect).

### 5.2 | Quality of evidence

In a systematic review by Gwozdziewycz and Mehl-Madrona (2013), the authors noted that the trials included in the review were validly designed and executed. However, we found varying degrees of bias across the included trials. Unclear randomization procedures were identified in a handful of trials, thereby potentially compromising the quality of these trials. The risk of allocation bias was mostly unclear across the review, due to a lack of adequate detail provided by study authors to make a definitive judgement. Almost all trials were considered to have a high risk of performance bias. This was due to treatment allegiance to a single psychotherapeutic intervention (NET),
making the blinding of personnel impossible due to the nature of the intervention. However, while blinding of participants could be possible to some extent, the extensive use of WLC and no-treatment control groups made this impossible in most trials. The risk of bias regarding outcome assessments suggested that most studies used self-report or assisted-report psychometric scales, translated and back-translated from English. No information was provided about the psychometric validity and reliability of the translated tool, and one cannot rule out the effect of interpreters or translators in achieving desired outcomes during assessments. As a result, several trials were judged as having an unclear risk of bias.

Studies used varying assessment methods such as scales of assessment, the language of administration, use of interpreters and diagnostic interviews versus self-report versus assisted report. This could potentially explain high heterogeneity among trials. Additionally, many RCTs did not report methodological aspects such as information on blinding, allocation and use of valid and reliable outcome measures clearly, which led to several studies being judged as having an unclear risk of bias.

There were other methodological concerns to be considered. Most of the trials reported WLC in addition to no-treatment conditions. Only a handful of studies used other active, bona fide interventions as controlled comparisons. As in Lely, Smid, et al. (2019), NET performed better against non-active comparators, but the difference was not statistically significant. The use of WLC in anxiety disorders research has been criticized (Patterson, Boyle, Kivlenieks, & Van Ameringen, 2016). Studies have identified ethical and humanitarian issues of delaying treatment to individuals who are undergoing acute distress or may be at risk for self-harm and suicide (Devilly & McFarlane, 2009). Additionally, the use of WLC may be associated with other methodological concerns related to increased risk of bias (Mohr et al., 2009) and larger effect sizes for the psychotherapy group (Furukawa et al., 2014). Specially with regard to PTSD research, a meta-analysis of 20 studies and 418 participants demonstrated small-to-medium effect sizes (g = 0.34) for WLC, whereas bona fide trauma-focused treatments yielded very large effect sizes (g = 1.5; Devilly & McFarlane, 2009). Similar findings were demonstrated in a review of WLC effect sizes in social anxiety disorder research (Steinert, Stadter, Stark, & Leichsenring, 2017). In agreement with a narrative review of NET efficacy (Robjant & Fazel, 2010), it is problematic to draw conclusions regarding NET’s superior efficacy compared with other trauma-focused treatments due to lack of sufficient RCTs using such active treatments as controls.

5.3 Limitations of the current review

The first limitation of this meta-analysis is that a protocol was not registered a priori. Preregistered protocols ensure methodological rigour and commitment, and the lack of a published protocol before the commencement of data collection and analysis is acknowledged as a shortcoming of this paper.

The risk of bias assessments were conducted with the Cochrane Collaboration’s risk of bias tool available at the time of analysis (Higgins & Green, 2011). Since that time, a new version of the tool (RoB 2.0) was released by the Cochrane Collaboration (Sterne et al., 2019). The updated tool is not yet available on the RevMan software and is currently being piloted by the Cochrane Review team. Future studies may employ RoB 2.0 to perform a comparable analysis of internal validity and quality appraisal.

Another limitation is that several trials used adaptations of NET ranging from NET specifically targeted to forensic offenders (FORNET) to brief versions of NET. No subgroup comparisons were planned and therefore conducted between RCTs that used NET and those with adapted NET. As such, targeted analyses of studies using specific adaptations will provide conclusive data of their efficacy, but due to the small number of trials reporting homogenous adaptations, it was not possible to determine whether NET can be successfully adapted to reduce symptomatology.

Other limitations of the paper are related to the included studies themselves. Some RCTs had very small sample sizes (in some cases below 10), with only two notably well-powered studies. Further, only four comparisons had over 10 studies contributing to the analysis. This raises the issue of small-study effects (i.e., the overestimation of intervention effects in trials with small to moderate trial size). Underpowered studies tend to show exaggerated treatment effects when compared with well-powered studies, as well as contributing to higher heterogeneity estimates. (Turner, Bird, & Higgins, 2013). Further, one of the possible causes of small-study effects could be publication bias (Higgins et al., 2019). Only published trials were included in this analysis, and the number of trials included in each comparison was low. It could be argued that with the inclusion of unpublished trials, smaller estimates of treatment effects for NET would emerge. Only four comparisons were eligible for funnel plot analyses, and a visual inspection did not reveal clear evidence of asymmetry. However, it is well known that asymmetry interpretations are subjective and may be caused by other issues such as methodological heterogeneity (Sedgwick, 2013). It must also be noted that tests for publication bias using funnel plots (such as the Egger’s test) are less reliable when small trials dominate the meta-analysis, which is the current case (Egger, Davey Smith, Schneider, & Minder, 1997). Therefore, the issue of small-study effects complicates the reliability of the NET evidence base considerably.

5.4 Strengths and comparisons to other reviews

An attempt was made to access and include all relevant trials through the search strategy. Although it is possible that we may have missed some published or unpublished material, our correspondence with the NET authors suggests that our review is the most comprehensive and complete analysis to date.

In addition to the trials included in Lely, Smid, et al. (2019), we have reviewed a further eight trials as well as analysed PTSD diagnostic status, which makes our findings more comprehensive and more up-to-date. In terms of quality appraisal, we find our ratings
6 | CONCLUSION

6.1 | Implications for practice

Currently, low-quality evidence indicates the efficacy of NET over both active and non-active psychotherapeutic control treatments in primarily decreasing PTSD symptom severity and, to a small extent, depression severity. The data were significantly heterogeneous across most outcomes measured. Further, there was no impact of the type of trauma or type of controls on treatment effect, which leaves heterogeneity estimates unexplained. Low dropout rates from therapy are an indication of NET’s acceptability and feasibility with sensitive populations. NET has been tested with a range of trauma groups, that is, with and without perpetrators. Although NET shows promise with both groups, the evidence base for NET is most strongly suited to the group it was originally intended for, that is, victims of war and organized violence. Although this suggests that NET may be suitable for diverse trauma groups, there is not enough data to draw conclusions regarding NET’s applicability when there is no perpetrator involved. The highly heterogeneous treatment groups (in terms of trauma history, risk of a future threat and sociocultural-economic settings) further confound the pooling of effects.

NET is currently recommended by a range of clinical guidelines for PTSD treatment including APA, NICE, VA/DoD and ISTSS. Regarding clinical efficacy, it is important to consider findings from reviews and meta-analyses. This involves considering the quality of the evidence base as opposed to focusing on data from individual RCTs or using narrative summaries. Although these trials may have found promising effects, the small sample sizes and methodological issues warrant caution when using the findings to inform policies and guidelines. Further, clinical guidelines have questioned the applicability of NET to non-refugee trauma populations. Because a majority of the trials so far have used NET with refugees and asylum seekers, we highlight this as a further limitation of the evidence base. In future research with NET, diverse trauma exposure across a range of sociocultural settings must be considered, specifically with non-refugee populations, to warrant its recommendation as a PTSD treatment approach.

6.2 | Implications for research

The NET evidence base includes RCTs conducted in diverse settings. As previously discussed, Mundt et al. (2014) critiqued the NET evidence base as being inapplicable to low- and middle-income countries (LMIC) as well as being unable to address as aspects of psychosocial well-being. In their review of interventions for torture survivors, Patel et al. (2014) raise similar concerns about the need to consider legal, contextual and psychosocial factors in the delivery of trauma interventions.

From a research perspective, RCTs of short-term interventions (of which NET is an example) have long since been criticized for underestimating or altogether missing out on some of the most crucial aspects of psychotherapy such as its self-correcting nature, addressing multiple, interacting problems and co-morbidities and therapist/practitioner-related moderators that impact outcomes and treatment effects (Fensterheim & Raw, 1996; Persons & Silberschatz, 1998; Seligman, 1995; Shean, 2014). However, as Rasmussen (2014) argues, RCT findings cannot be expected to address all aspects of psychosocial well-being but can augment components of care packages intended to effect recovery at multiple levels of well-being. The larger issue concerns the need for empirical health systems research, in that the scope of NET research needs to be widened to include the socio-economic, political and cultural contexts of the local communities they are intended to be implemented in. Holistic treatment and care systems must be evaluated for efficacy as opposed to a single component, that is, the intervention. This is especially applicable to NET research, a technique whose basis, that is, storytelling, is woven into the unique sociocultural fabric of diverse societies.

Mechanisms of change is an important, under-researched area when it comes to NET inquiry. A few included trials have isolated neurobiological and molecular correlates of recovery when NET is used (Adenauer et al., 2011; Morath, Gola, et al., 2014; Morath, Moreno-Villanueva, et al., 2014). Additionally, psychological mechanisms of change can add rich detail to bettering intervention protocols. A rigorous, in-depth qualitative analysis of the narratives or final testimony could provide new insight into change by identifying and isolating indicators of recovery or deterioration in accordance to treatment protocols in the manual (Schauer et al., 2005, 2011). This evidence adds depth to statistical data, as testimonies are largely personal accounts of change and are bound to integrate the psychosocial, political, economic and cultural context of the treatment setting.

Homogenous, well-powered trials using uniform methodological design as well as comparable outcome assessments (e.g., measures and follow-up time) are needed to substantiate NET’s spot among active trauma interventions. When complemented by deeper, qualitative analyses of mechanisms of change, NET’s evidence base can be strengthened to reveal its true effect across trauma populations.

ACKNOWLEDGEMENT

The authors wish to extend their gratitude to Dr Farhad Shokraneh for his direction and guidance to author S. R. in carrying out the meta-analysis, as well as for making the time to clarify doubts at every step.
of the way. His expertise in conducting systematic reviews and meta-analyses was immensely valuable in undertaking this study. The authors received no specific funding for this work.

CONFLICT OF INTEREST
The authors have declared that no conflicting interests exist.

ORCID
Shruti Raghuraman  https://orcid.org/0000-0002-7366-1147

REFERENCES
References marked with an asterisk indicate studies included in the meta-analysis. The in-text citations to studies selected for meta-analysis are not preceded by an asterisk.

Adenauer, H., Catani, C., Gola, H., Keil, J., Ruf, M., Schauer, M., & Neuner, F. (2011). Narrative exposure therapy for PTSD increases top-down processing of aversive stimuli—Evidence from a randomized controlled treatment trial. BMC Neuroscience, 12(1), 127. https://doi.org/10.1186/1471-2202-12-127

Alghamdi, M., Hunt, N., & Thomas, S. (2015). The effectiveness of narrative exposure therapy with traumatised firefighters in Saudi Arabia: A randomized controlled study. Behaviour Research and Therapy, 66, 64–71. https://doi.org/10.1016/j.brat.2015.01.008

All-Hadethe, A., Hunt, N., Al-Qaysi, G., & Thomas, S. (2015). Randomised controlled study comparing two psychological therapies for post-traumatic stress disorder (psdt): Emotional freedom techniques (eft) vs. narrative exposure therapy (net). Journal of Traumatic Stress Disorders & Treatment, 04(04). https://doi.org/10.4172/2324-8947.1000145

American Psychiatric Association. (1994). Diagnostic and statistical manual of mental disorders, 4th edition (4th ed.). American Psychiatric Association. https://doi.org/10.1097/00000485-199406000-00001

American Psychological Association (APA). (2017). Clinical practice guidelines for the treatment of posttraumatic stress disorder (PTSD). https://www.apa.org/pi/ptsd/guidelines/ptsd

Beck, A., Steer, R., & Brown, G. (1996). Manual for the Beck Depression Inventory-II. https://scholar.google.com/scholar_lookup?q=+&+publication_year=1996&author=A.+T.+Beck&author=R.+A.+Steer&author=G.+K.++Brown&title=Manual+for+the+Beck+Depression+Inventory-II

Berliner, L., Bisson, J., Cloitre, M., Forbes, D., Golbeck, L., Jensen, T., ... Shapiro, F. (2019). Posttraumatic stress disorder prevention and treatment guidelines methodology and recommendations. International Society for Traumatic Stress Studies (ISTSS). http://www.istss.org/getattachment/Treating-Trauma/New-ISTSS-Prevention-and-Treatment-Guidelines/ISTSS_PreventionTreatmentGuidelines_FNL-March-19-2019.pdf.aspx

Bichescu, D., Neuner, F., Schauer, M., & Elbert, T. (2007). Narrative exposure therapy for political imprisonment-related chronic posttraumatic stress disorder and depression. Behaviour Research and Therapy, 45(9), 2212–2220. https://doi.org/10.1016/j.brat.2006.12.006

Bisson, J. I., Roberts, N. P., Andrew, M., Cooper, R., & Lewis, C. (2013). Psychological therapies for chronic post-traumatic stress disorder (PTSD) in adults (review). Cochrane Database of Systematic Reviews, 13(2). https://doi.org/10.1002/14651858.CD003388.pub4. www.cochranelibrary.com

Cochrane Library. (2019). Risk of bias 2 (RoB 2) tool|Cochrane methods. https://methods.cochrane.org/risk-bias-2

Cohen, J. (1960). A coefficient of agreement for nominal scales. Educational and Psychological Measurement, 20(1), 37–46. https://doi.org/10.1177/001316460002000104

Crombach, A., & Elbert, T. (2015). Controlling obsessive behavior using narrative exposure therapy. Clinical Psychological Science, 3(2), 270–282. https://doi.org/10.1177/2167706214534239

Crombach, A., & Siehl, S. (2018). Impact and cultural acceptance of the narrative exposure therapy in the aftermath of a natural disaster in Burundi. BMC Psychiatry, 18(1), 233. https://doi.org/10.1186/s12888-018-1799-3

Derogatis, L. R., Lipman, R. S., Rickels, K., Uhlenhuth, E. H., & Covi, L. (1974). The Hopkins Symptom Checklist (HSCl): A self-report symptom inventory. Behavioral Science, 19(1), 1–15. https://doi.org/10.1002/bs.3830190102

Devilly, G. J., & McFarlane, A. C. (2009). When wait lists are not feasible, nothing is a thing that does not need to be done. Journal of Consulting and Clinical Psychology, 77(6), 1159–1168. https://doi.org/10.1037/a0016878

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. BMJ (Clinical Research Ed.), 315(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629

‘Erti, V., Pfeiffer, A., Schauer, E., Elbert, T., & Neuner, F. (2011). Community-implemented trauma therapy for former child soldiers in Northern Uganda: A randomized controlled trial. JAMA: The Journal of the American Medical Association, 306(5), 503–512. https://doi.org/10.1001/jama.2011.1060

Fensterheim, H., & Raw, S. D. (1996). Psychotherapy research is not psychotherapy practice. Clinical Psychology: Science and Practice, 3(2), 168–171. https://doi.org/10.1111/1468-2850.1996.tb00067.x

Foa, E. B., Cashman, L., Jaycox, L. & Perry, K. (1997). The validation of a self-report measure of posttraumatic stress disorder: The Post-traumatic Diagnostic Scale. Psychological Assessment, 9(4), 445–451. https://doi.org/10.1037/1040-3590.9.4.445

Foa, E. B., Riggs, D. S., Dancu, C. V., & Rothbaum, B. O. (1993). Reliability and validity of a brief instrument for assessing post-traumatic stress disorder. Journal of Traumatic Stress, 6(4), 459–473. https://doi.org/10.1002/jts.2490060405

Foa, E. B., & Tolin, D. F. (2000). Comparison of the PTSD symptom scale-interview version and the clinician-administered PTSD scale. Journal of Traumatic Stress, 13(2), 181–191. https://doi.org/10.1023/A:1007781909213

Furukawa, T. A., Noma, H., Caldwell, D. M., Honayashi, M., Shinohara, K., Imai, H., ... Churchill, R. (2014). Waiting list may be a nocebo condition in psychotherapy trials: A contribution from network meta-analysis. Acta Psychiatrica Scandinavica, 130(3), 181–192. https://doi.org/10.1111/acps.12275

Guolo, A., & Varin, C. (2017). Random-effects meta-analysis: The number of studies matters. Statistical Methods in Medical Research, 26(3), 1500–1518. https://doi.org/10.1177/0962280215583568

Gwozdziewycz, N., & Mehl-Madrona, L. (2013). Meta-analysis of the use of narrative exposure therapy for political imprisonment-related chronic posttraumatic stress disorder in adults: An update. Psychiatry, Neurosurgery, and Psychiatry, 111(11), 26. doi.org/papers3/publication/-uuid/D2921EAD-7C0D-4A1A-B313-7456F6D4A945. https://doi.org/10.1111/acps.12275

Hamblen, J. L., Norman, S. B., Sonis, J. H., Phelps, A. J., Bisson, J. I., Nunes, V. D., ... Schnurr, P. P. (2019). A guide to guidelines for the treatment of posttraumatic stress disorder in adults: An update. Psychotherapy, 56(3), 359–373. https://doi.org/10.1037/pst0000231

Hamilton, M. (1960). A rating scale for depression. Journal of Neurology, Neurosurgery, and Psychiatry, 23, 56–62. doi.org/papers3/publication/-uuid/D9291EAD-7C0D-4A1A-B313-7456F6D4A945. https://doi.org/10.1136/jnpn.23.1.56

Hamilton, M. (1967). Development of a rating scale for primary depressive illness. The British Journal of Social and Clinical Psychology, 6(4), 278–296. https://doi.org/10.1111/j.2044-8260.1967.tb00530.x

Harding, T. W., de Arango, M. V., Baltazar, J., Climent, C. E., Ibrahim, H. H., Ladrino-Ignacio, L., Murthy, R. S., & Wig, N. N. (1980). Mental...
disorders in primary health care: A study of their frequency and diagnosis in four developing countries. *Psychological Medicine*, 10(2), 231–241. http://www.ncbi.nlm.nih.gov/pubmed/7384326.

Hecker, T., Hermenau, K., Crombach, A., & Elbert, T. (2015). Treating traumatized offenders and veterans by means of narrative exposure therapy. *Frontiers in Psychiatry*, 6(JUN), 80. https://doi.org/10.3389/fpsyt.2015.00080

Hensel-Dittmann, D., Schauer, M., Ruf, M., Catani, C., Odenwald, M., Elbert, T., & Neuner, F. (2011). Treatment of traumatized victims of war and torture: A randomized controlled comparison of narrative exposure therapy and stress inoculation training. *Psychotherapy and Psychosomatics*, 80(6), 345–352. https://doi.org/10.1159/000327253

Hermenau, K., Hecker, T., Schaal, S., Maedl, A., & Elbert, T. (2013). Addressing post-traumatic stress and aggression by means of narrative exposure: A randomized controlled trial with ex-combatants in the eastern DRC. *Journal of Aggression, Maltreatment & Trauma*, 22(8), 916–934. https://doi.org/10.1080/10926771.2013.824057

Higgins, J. P. T., & Green, S. (2011 & (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions*. Version 5.1.0. The Cochrane Collaboration. http://handbook-5-1.cochrane.org

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (Eds.). (2019). *Cochrane handbook for systematic reviews of interventions*, version 6.0 (updated July 2019), Cochrane. (updated July 2019). Handbook. Available from http://www.training.cochrane.org/handbook

Hijazi, A. M., Lumley, M. A., Ziadni, M. S., Haddad, L., Rapport, L. J., & Arnetz, B. B. (2014). Brief narrative exposure therapy for post-traumatic stress in Iraqi refugees: A preliminary randomized clinical trial. *Journal of Traumatic Stress*, 27, 314–322. https://doi.org/10.1002/jts.21922

Hinsberger, M., Holtzhausen, L., Sommer, J., Kaminer, D., Elbert, T., Seedat, S., ... Weinerstall, R. (2019). Long-term effects of psychotherapy in a context of continuous community and gang violence: Changes in aggressive attitude in high-risk South African adolescents. *Behavioural and Cognitive Psychotherapy*, 48, 1–13. https://doi.org/10.1017/S0135246519000365

Hinsberger, M., Holtzhausen, L., Sommer, J., Kaminer, D., Elbert, T., Seedat, S., ... Weinerstall, R. (2017). Feasibility and effectiveness of narrative exposure therapy and cognitive behavioral therapy in a context of ongoing violence in South Africa. *Psychological Trauma Theory Research Practice and Policy*, 9(3), 282–291. https://doi.org/10.1037/ptp0000197

Jacob, N., Neuner, F., Maedl, A., Schaal, S., & Elbert, T. (2014). Dissemination of psychotherapy for trauma spectrum disorders in postconflict settings: A randomized controlled trial in Rwanda. *Psychotherapy and Psychosomatics*, 83(6), 354–363. https://doi.org/10.1159/000365114

Köbach, A., Schaal, S., Hecker, T., & Elbert, T. (2017). Psychotherapeutic intervention in the mobilization process: Addressing combat-related mental injuries with narrative exposure in a first and second dissemination stage. *Clinical Psychology & Psychotherapy*, 24(4), 807–825. https://doi.org/10.1002/cpp.1986

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. In *Source: Biometrics*, 33(1), 159–174. https://doi.org/10.2307/2529310

Lely, J. C. G., Knipscheer, J. W., Moerbeek, M., Ter Heide, F. J. J., Van Den Bout, J., & Kleber, R. J. (2019). Randomised controlled trial comparing narrative exposure therapy with present-centred therapy for older patients with post-traumatic stress disorder. *British Journal of Psychiatry*, 214(6), 369–377. https://doi.org/10.1192/bjp.2019.59

Lely, J. C. G., Smid, G. E., Jongedijk, R. A., Knipscheer, J. W., & Kleber, R. J. (2019). The effectiveness of narrative exposure therapy: A review, meta-analysis and meta-regression analysis. *European Journal of Psychotraumatology*, 10(1), 1550344. https://doi.org/10.1080/20008198.2018

Management of Posttraumatic Stress Disorder and Acute Stress Disorder Working Group. (2017). *VA/DoD Clinical Practice Guideline for the Management of Posttraumatic Stress Disorder and Acute Stress Disorder Version 3.0*. Department of Veterans Affairs, Department of Defense. https://doi.org/10.1016/j.addbeh.2017.07.010

McPherson, J. (2012). Does narrative exposure therapy reduce PTSD in survivors of mass violence? *Research on Social Work Practice*, 22(1), 29–42. https://doi.org/10.1177/1049731511414147

Mohr, D. C., Spring, B., Freedland, K. E., Beckner, V., Arean, P., Hollow, S. D., ... Kaplan, R. (2009). The selection and design of control conditions for randomized controlled trials of psychological interventions. *Psychotherapy and Psychosomatics*, 78, 275–284. https://doi.org/10.1159/000228248

Morath, J., Gola, H., Sommershof, A., Hamuni, G., Kolassa, C., Catani, C., ... Kolassa, I.-T. (2014). The effect of trauma-focused therapy on the altered T cell distribution in individuals with PTSD: Evidence from a randomized controlled trial. *Journal of Psychiatric Research*, 54, 1–10. https://doi.org/10.1016/j.jpyrsch.2014.03.016

Morath, J., Morath, J., Moreno-Villanueva, M., Hamuni, G., Kolassa, S., Ruf-Leuschner, M., Schauer, M., ... Kolassa, I.-T. (2013). Effects of psychotherapy on DNA strand break accumulation originating from traumatic stress. *Psychotherapy and Psychosomatics*, 82(5), 289–297. https://doi.org/10.1159/000362739

Mundt, A. P., Wünsche, P., Heinz, A., & Pross, C. (2014). Evaluating interventions for posttraumatic stress disorder in low and middle income countries: Narrative exposure therapy. *Intervention: Journal of Mental Health and Psychosocial Support in Conflict Affected Areas*, 12(2), 250–266. https://doi.org/10.1097/WTF.0000000000000001

Neuner, F., Kurreck, S., Ruf, M., Odenwald, M., Elbert, T., & Schauer, M. (2010). Can asylum-seekers with posttraumatic stress disorder be successfully treated? A randomized controlled pilot study. *Cognitive Behaviour Therapy*, 39(4), 1–11. https://doi.org/10.1080/16506070903121042

Neuner, F., Onyut, P. L., Ertl, V., Odenwald, M., Schauer, E., & Elbert, T. (2008). Treatment of posttraumatic stress disorder by trained lay counselors in an African refugee settlement: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, 76(4), 686–694. https://doi.org/10.1037/0022-006X.76.4.686

Neuner, F., Schauer, M., Klaschik, C., Karunakara, U., & Elbert, T. (2004). A comparison of narrative exposure therapy, supportive counseling, and psychoeducation for treating posttraumatic stress disorder in an African refugee settlement. *Journal of Consulting and Clinical Psychology*, 72(4), 579–587. https://doi.org/10.1037/0022-006X.72.4.579

Neuner, F., Schauer, M., Roth, W. T., & Elbert, T. (2002). A narrative exposure treatment as intervention in a Macedonia’s refugee camp: A case report. *Behavioural and Cognitive Psychotherapy*, 30(2), 205–209. https://doi.org/10.1017/S0144918201342072

Onyut, L. P., Neuner, F., Schauer, E., Ertl, V., Odenwald, M., Schauer, M., ... Elbert, T. (2005). Narrative exposure therapy as a treatment for child war survivors with posttraumatic stress disorder: Two case reports and a pilot study in an African refugee settlement. *BMC Psychiatry*, 5, 7. https://doi.org/10.1186/1471-244X-5-7

Orang, T., Ayoughi, S., Moran, J. K., Ghaffari, H., Mostafavi, S., Rasoulian, M., & Elbert, T. (2018). The efficacy of narrative exposure therapy in a sample of Iranian women exposed to ongoing intimate partner violence—A randomized controlled trial. *Clinical Psychology & Psychotherapy*, 25(6), 827–841. https://doi.org/10.1002/cpp.2318

Pabst, A., Schauer, M., Bernhardt, K., Ruf-Leuschner, M., Goder, R., Elbert, T., ... Seeck-Hirschner, M. (2014). Evaluation of narrative exposure therapy (NET) for borderline personality disorder with comorbid posttraumatic stress disorder. *Clinical Neuropsychiatry: Journal of Treatment Evaluation*, 11(3), 108–117. https://doi.org/10.3390/ bs40x000x

Patel, N., Kellezi, B., & Williams, A. (2014). Psychological, social and welfare interventions for psychological health and well-being of torture
survivors. The Cochrane Library, 11. https://doi.org/10.1002/14651858.CD009317.pub2.wwcochranelibrary.com

Patterson, B., Boyle, M. H., Kivlenieks, M., & Van Ameringen, M. (2016). The use of waitlists as control conditions in anxiety disorders research. Journal of Psychiatric Research, 83, 112–120. https://doi.org/10.1016/j.jpsychires.2016.08.015

Persons, J. B., & Silberschatz, G. (1998). Are results of randomized controlled trials useful to psychotherapists? Journal of Consulting and Clinical Psychology, 66(1), 126–135. https://doi.org/10.1037/0022-006X.66.1.126

Rasmussen, A. (2014). Don’t fault RCTs for not testing systems of care. Intervention, 12(2), 289–291. https://doi.org/10.1097/WTF.000000000000035

Review Manager (RevMan). [Computer Program] Version 5.3. Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration. 2014. https://community.cochrane.org/help/tools-and-software/revman-5

Robjant, K., & Fazel, M. (2010). The emerging evidence for narrative exposure therapy: A review. Clinical Psychology Review, 30(8), 1030–1039. https://doi.org/10.1016/j.cpr.2010.07.004

Roland, M., & Torgerson, D. J. (1998). Understanding controlled trials. What are pragmatic trials? In British Medical Journal. British Medical Journal Publishing Group, 316(7127), 285–289. https://doi.org/10.1136/bmj.316.7127.285.

Schaal, S., Elbert, T., & Neuner, F. (2009). Narrative exposure therapy versus interpersonal psychotherapy. Psychotherapy and Psychosomatics, 78(5), 298–306. https://doi.org/10.1159/000229768

Schauer, M., Elbert, T., & Neuner, F. (2011). Narrative exposure therapy: A short-term treatment for traumatic stress disorders (2nd ed.). Hogrefe. https://us.hogrefe.com/shop/narrative-exposure-therapy-67506.html

Schauer, M., Neuner, F., & Elbert, T. (2005). Narrative exposure therapy. Hogrefe Publishing. https://doi.org/10.1016/B978-0-88939-120-5.0000000000000035

Sedgwick, P. (2013). Meta-analyses: How to read a funnel plot. BMJ, 346(3), f1342. https://doi.org/10.1136/bmj.f1342

Seide, S. E., Röver, C., & Friede, T. (2019). Likelihood-based random-effects meta-analysis with few studies: Empirical and simulation study. British Medical Journal, 368, k1863. https://doi.org/10.1136/bmj.k1863

Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The consumer reports study. American Psychologist, 50(12), 965–974. https://doi.org/10.1037/0003-066X.50.12.965

Shean, G. (2014). Limitations of randomized controlled design in psychotherapy research. Advances in Psychiatry, 2014, 1–5. https://doi.org/10.1155/2014/561452

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., ... Dunbar, G. C. (1998). The Mini-International Neuropsychiatric Interview (M.I.N.I.); The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. The Journal of Clinical Psychiatry, 59(2), 22–33. http://www.ncbi.nlm.nih.gov/pubmed/9981538

Shoeb, M., Weinstein, H., & Mollica, R. (2007). The Harvard Trauma Questionnaire: Adapting a cross-cultural instrument for measuring torture, trauma and posttraumatic stress disorder in Iraqi refugees. International Journal of Social Psychiatry, 53(5), 447–463. https://doi.org/10.1177/0020764007078362

Spitzer, R. L., Kroenke, K., & Williams, J. B. (1999). Validation and utility of a self-report version of PRIME-MD: The PHQ primary care study. JAMA, 282(18), 1737–1744. http://www.ncbi.nlm.nih.gov/pubmed/10568646, https://doi.org/10.1001/jama.282.18.1737

Steinert, C., Stadter, K., Stark, R., & Leichsenring, F. (2017). The effects of waiting for treatment: A meta-analysis of waitlist control groups in randomized controlled trials for social anxiety disorder. Clinical Psychology & Psychotherapy, 24(3), 649–660. https://doi.org/10.1002/cpp.2032

Stubmark, H., Catani, C., Neuner, F., Elbert, T., & Holen, A. (2013). Treating PTSD in refugees and asylum seekers within the general health care system: A randomized controlled multicentre study. Behaviour Research and Therapy, 51(10), 641–647. https://doi.org/10.1016/j.brat.2013.07.002

Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., ... Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. The BMJ, 366, I4898. https://doi.org/10.1136/bmj.i4898

Turner, R. M., Bird, S. M., & Higgins, J. P. T. (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

Wampold, B. E., & Imel, Z. E. (2015). Counseling and psychotherapy: The great psychotherapy debate: The evidence for what makes psychotherapy work (2nd ed.). Routledge/Taylor & Francis Group. https://doi.org/10.4324/9780203582015

Weiss, D. S., & Marmar, C. R. (1997). The impact of events scale—Revised. In J. P. Wilson & Keane. (Eds.), Assessing psychological trauma and PTSD (pp. 339–441). Guilford Press. https://books.google.com/books?id=BID3WWwAJ5EC&pgis=1

World Health Organization. (1990). Composite International Diagnostic Interview (CIDI) (1st ed.). https://www.who.int/

Zang, Y., Cox, T., & Hunt, N. (2013). A randomised controlled pilot study—The effectiveness of narrative exposure therapy with adult survivors of the Sichuan earthquake. BMC Psychiatry, 13, 41. https://doi.org/10.1186/1471-244X-13-41

Zang, Y., Hunt, N., & Cox, T. (2014). Adapting narrative exposure therapy for Chinese earthquake survivors: A pilot randomised controlled feasibility study. BMC Psychiatry, 14(1), 262. https://doi.org/10.1186/s12888-014-0252-3

Zigmond, A. S., & Snaith, R. P. (1983). The hospital anxiety and depression scale. Acta Psychiatrica Scandinavica, 67(6), 361–370. https://doi.org/10.1111/j.1600-0447.1983.tb09716.x

How to cite this article: Raghuraman S, Stuttard N, Hunt N. Evaluating narrative exposure therapy for post-traumatic stress disorder and depression symptoms: A meta-analysis of the evidence base. Clin Psychol Psychother. 2020;1–23. https://doi.org/10.1002/cpp.2486
### TABLE A1  Search strategy example: PubMed

1. narrative exposure* [Title/Abstract] OR 'narrative exposure therapy' [Title/Abstract] OR 'NET' [Title/Abstract] OR narrative therapy [Text Word]
   - Articles retrieved: 106,999

2. 'ptsd' [Text Word] OR posttraumatic stress* [Text Word] OR post-traumatic stress* [Text Word] OR 'psychological trauma' [Text Word]
   - Articles retrieved: 48,966

3. randomi?ed controlled trial [Publication Type] OR 'RCT' [Title/Abstract] OR controlled clinical trial [Publication Type] OR randomi?ed [Title/Abstract] OR randomly [Title/Abstract] OR trial [Title/Abstract]
   - Articles retrieved: 1,098,239

4. 1 AND 2 AND 3
   - Articles retrieved: 64

### TABLE A2  Characteristics of included studies

| Study                          | Population                        | Location    | Intervention                  | Controls                  | Psychological outcomes |
|-------------------------------|-----------------------------------|-------------|-------------------------------|---------------------------|------------------------|
| Adenauer et al. (2011)        | Refugees and asylum seekers        | Germany     | NET (N = 16)                  | WLC (N = 18)              | Trauma exposure        |
|                               |                                   |             |                               |                           | PTSD symptoms          |
|                               |                                   |             |                               |                           | Co-morbidities         |
| Alghamdi, et al. (2015)       | Firefighters                      | Saudi Arabia| NET (N = 17)                  | WLC (N = 17)              | PTSD symptoms          |
|                               |                                   |             |                               |                           | Depression symptoms    |
|                               |                                   |             |                               |                           | Anxiety symptoms       |
|                               |                                   |             |                               |                           | Coping skills          |
|                               |                                   |             |                               |                           | Social support         |
| AlHadethe, Hunt, Al-Qaysi, and Thomas (2015) | Young refugees | Iraq      | NET (N = 20)                  | Emotional freedom technique (N = 20), no intervention (N = 20) | Traumatic events |
|                               |                                   |             |                               |                           | PTSD symptoms          |
|                               |                                   |             |                               |                           | Anxiety symptoms       |
|                               |                                   |             |                               |                           | Depression symptoms    |
|                               |                                   |             |                               |                           | Coping strategies      |
|                               |                                   |             |                               |                           | Religious coping       |
|                               |                                   |             |                               |                           | Social support         |
| Bichescu et al. (2007)        | Former Romanian political detainees | Romania     | NET (N = 9)                   | Psychoeducation (N = 9)   | PTSD diagnosis         |
|                               |                                   |             |                               |                           | PTSD symptoms          |
|                               |                                   |             |                               |                           | Depression symptoms    |
| Crombach and Elbert (2015)    | Former street children            | Burundi     | FORNET (N = 15)               | Treatment-as-usual (TAU) (N = 25) | Exposure to traumatic events |
|                               |                                   |             |                               |                            | PTSD diagnosis         |
|                               |                                   |             |                               |                            | PTSD symptoms          |
|                               |                                   |             |                               |                            | Recent offences        |
|                               |                                   |             |                               |                            | Appetitive aggression  |
| Study                                      | Population                          | Location                | Intervention | Controls                  | Psychological outcomes                                                                 |
|-------------------------------------------|-------------------------------------|-------------------------|--------------|---------------------------|----------------------------------------------------------------------------------------|
| Ertl, Pfeiffer, Schauer, Elbert, and Neuner (2011) | Former child soldiers              | Uganda                  | NET (N = 29) | Academic catch-up (N = 28), WLC (N = 28) | Potentially traumatic events, PTSD diagnosis, PTSD symptoms, Major depression symptoms, Suicide risk and ideations, Trauma-related guilt, Survivor guilt, Perceived stigmatization |
| Hensel-Dittman et al. (2011)              | Refugees and asylum seekers          | Germany                 | NET (N = 15) | Stress inoculation training (N = 13) | Experiences of organized violence, PTSD symptoms, Depression symptoms, Co-morbidities |
| Hermenau et al. (2013)                    | Former child soldiers and ex-combatants | Congo                   | FORNET (N = 19) | No intervention (N = 19) | PTSD symptoms, Appetitive aggression, Integration |
| Hijazi et al. (2014)                      | Iraqi refugees                      | United States           | Brief NET (N = 41) | WLC (N = 22) | Post-traumatic growth, Well-being, Post-traumatic stress, Depression symptoms, Somatic symptoms |
| Hinsberger et al. (2019)                  | Ex-prisoners and at-risk youth      | South Africa            | FORNET (N = 20) | CBT (N = 20), WLC (N = 48) | PTSD symptoms, Appetitive aggression, Perpetrated violence |
| Jacob et al. (2014)                       | Genocide victims (widows and orphans) | Rwanda                  | NET + IPT (N = 38) | WLC (N = 38) | PTSD severity and frequency, PTSD diagnosis, Functional impairment |
| Kobach et al. (2015)                      | Former members of military groups   | Congo                   | FORNET (N = 29) | TAU (N = 23) | Exposure to violence, PTSD diagnosis, PTSD symptoms, Appetitive aggression, Depression diagnosis, Depression symptoms |

(Continues)
| Study                          | Population                                      | Location          | Intervention                      | Controls          | Psychological outcomes                              |
|-------------------------------|--------------------------------------------------|-------------------|-----------------------------------|-------------------|-----------------------------------------------------|
| Lely, Knipscheer, et al. (2019) | Older adults with trauma exposure (political, sexual and childhood trauma) | The Netherlands NET (N = 18) | Present-centred therapy (N = 15) | PTSD diagnosis | Drug dependence                                      |
|                               |                                                  |                   |                                   |                   | Reintegration success                                |
| Morath, Gola, et al. (2014)   | African and Middle Eastern refugees              | Germany           | NET (N = 17)                      | WLC (N = 17)      | PTSD symptoms                                       |
|                               |                                                  |                   |                                   |                   | Depression symptoms                                  |
|                               |                                                  |                   |                                   |                   | Co-morbidities                                      |
|                               |                                                  |                   |                                   |                   | Somatic complaints                                  |
| Morath, Moreno-Villanueva, et al. (2014) | African and Middle Eastern refugees          | Germany           | NET (N = 19)                      | WLC (N = 19)      | PTSD diagnosis                                       |
|                               |                                                  |                   |                                   |                   | PTSD symptoms                                       |
| Neuner et al. (2004)          | Sudanese refugees                               | Uganda            | NET (N = 17)                      | Supportive counselling (N = 14), Psychoeducation (N = 12) | PTSD diagnosis |
|                               |                                                  |                   |                                   |                   | PTSD symptoms                                       |
|                               |                                                  |                   |                                   |                   | Traumatic experiences, health                       |
|                               |                                                  |                   |                                   |                   | Depression symptoms                                  |
|                               |                                                  |                   |                                   |                   | Anxiety symptoms                                     |
|                               |                                                  |                   |                                   |                   | Psychological functioning                            |
| Neuner et al. (2008)          | Rwandan and Somali refugees                      | Uganda            | NET (N = 111)                     | Trauma counselling (N = 111), no treatment            | PTSD diagnosis |
|                               |                                                  |                   |                                   |                                                               | PTSD symptoms  |
|                               |                                                  |                   |                                   |                                                               | Physical health|
| Neuner et al. (2010)          | Asylum seekers                                   | Germany           | NET (N = 16)                      | TAU (N = 16)      | Traumatic experiences                               |
|                               |                                                  |                   |                                   |                                                               | PTSD diagnosis |
|                               |                                                  |                   |                                   |                                                               | PTSD symptoms  |
|                               |                                                  |                   |                                   |                                                               | Pain symptoms  |
|                               |                                                  |                   |                                   |                                                               | Depression symptoms |
| Orang et al. (2018)           | Intimate partner violence survivors              | Iran              | NET (N = 24)                      | TAU (N = 21)      | PTSD symptoms                                       |
|                               |                                                  |                   |                                   |                                                               | Depression symptoms |
|                               |                                                  |                   |                                   |                                                               | Perceived stress |
|                               |                                                  |                   |                                   |                                                               | Abusive behaviours |
|                               |                                                  |                   |                                   |                                                               | Traumatic events |
|                               |                                                  |                   |                                   |                                                               | Childhood traumatic events |
|                               |                                                  |                   |                                   |                                                               | Disability |
|                               |                                                  |                   |                                   |                                                               | Borderline symptoms |
TABLE A2  (Continued)

| Study                              | Population                        | Location   | Intervention   | Controls               | Psychological outcomes                      |
|-----------------------------------|-----------------------------------|------------|----------------|------------------------|--------------------------------------------|
| Pabst et al. (2014)               | Physical/sexual abuse/assault survivors | Germany   | NET (N = 11)   | Treatment by experts (N = 11) | PTSD diagnosis  
PTSD symptoms  
Borderline personality disorder symptoms  
Depression  
Dissociative symptoms |
| Schaal et al. (2009)              | Genocide orphans                  | Rwanda     | NET (N = 12)   | IPT (N = 14)            | PTSD diagnosis  
PTSD symptoms  
Depressive symptoms  
Depression symptoms  
Guilt |
| Stenmark, Catani, Neuner, Elbert, and Holen (2013) | Refugees                          | Norway     | NET (N = 51)   | TAU (N = 30)            | PTSD diagnosis  
PTSD symptom  
Depression diagnosis  
Depression symptoms |
| Zang, Cox, and Hunt (2013)        | Earthquake survivors              | China      | NET (N = 11)   | WLC (N = 11)            | PTSD symptoms  
Depression and anxiety  
General mental health  
Positive and negative post-traumatic changes  
Social support  
Coping style |
| Zang, Hunt, and Cox (2014)        | Earthquake survivors              | China      | NET and Brief NET (N = 20) | WLC (N = 10)            | PTSD symptoms  
Depression and anxiety  
General mental health  
Positive and negative post-traumatic changes  
Social support  
Coping strategies |

Abbreviations: CBT, cognitive behavioural therapy; FORNET, narrative exposure therapy for forensic offender rehabilitation; IPT, interpersonal psychotherapy; N, number of participants randomized; NET, narrative exposure therapy; PTSD, post-traumatic stress disorder; WLC, wait-list controls.
### TABLE A3  Summary of findings

| Outcomes          | Treatment effect | Heterogeneity |
|------------------|------------------|---------------|
|                  | SMD/OR 95% CI    | Z  | p   | N  | χ² | df | p | I² (%) |
| PTSD Diagnosis (ST) | 0.28 (OR) 0.12 to 0.66 | 2.89 | .004 | 173 | 3.09 | 3 | .38 | 3 |
| PTSD Diagnosis (MT) | 0.26 (OR) 0.13 to 0.55 | 3.59 | .0003 | 150 | 1.75 | 3 | .63 | 0 |
| PTSD Diagnosis (LT) | 0.68 (OR) 0.16 to 2.87 | 0.52 | .60 | 190 | 9.96 | 2 | .007 | 80 |
| Symptoms (ST) | −0.30 (SMD) −0.49 to −0.11 | 3.05 | .002 | 813 | 21.62 | 14 | .09 | 35 |
| Symptoms (MT) | −0.45 (SMD) −0.68 to −0.23 | 4.01 | <.0001 | 763 | 23.29 | 13 | .04 | 44 |
| Symptoms (LT) | −0.49 (SMD) −0.80 to −0.18 | 3.06 | .002 | 302 | 9.19 | 5 | .16 | 35 |
| Depression Symptoms (ST) | −0.30 (SMD) −0.49 to −0.11 | 1.88 | .06 | 425 | 12.25 | 9 | .20 | 27 |
| Depression Symptoms (MT) | −0.49 (SMD) −0.79 to −0.20 | 3.25 | .001 | 424 | 20.64 | 10 | .02 | 52 |
| Depression Symptoms (LT) | −0.19 (SMD) −0.46 to 0.09 | 1.34 | .18 | 224 | 1.66 | 4 | .80 | 0 |

Abbreviations: CI, confidence interval; LT, long term; MT, midterm; OR, odds ratio; PTSD, post-traumatic stress disorder; SMD, standardized mean difference; ST, short term; Z, test of overall effect.

### TABLE A4  Subgroup analysis: Active controls versus no intervention

| Outcomes            | Treatment effect Active intervention | Treatment effect No intervention | Test for subgroup differences |
|---------------------|--------------------------------------|----------------------------------|--------------------------------|
|                     | N         | SMD 95% CI | Z  | p   | N         | SMD 95% CI | Z   | p         | χ² | df | p | I² (%) |
| PTSD symptoms Short term | 454 | −0.07 | −0.34 to 0.20 | 0.49 | .62 | 407 | −0.44 | −0.68 to −0.21 | 3.68 | .0002 | 4.12 | 1 | .04 | 75.7 |
| PTSD symptoms Midterm | 608 | −0.41 | −0.74 to −0.08 | 2.43 | .02 | 320 | −0.46 | −0.69 to −0.23 | 3.91 | <.0001 | 0.05 | 1 | .82 | 0 |
| PTSD symptoms Long term | 241 | −0.42 | −0.88 to 0.04 | 1.8 | .07 | 380 | −0.44 | −0.78 to −0.10 | 2.56 | .01 | 0.00 | 1 | .95 | 0 |
| Depression symptoms Short term | 199 | −0.15 | −0.51 to 0.20 | 0.85 | .40 | 293 | −0.34 | −0.73 to −0.05 | 1.71 | .09 | 0.48 | 1 | .49 | 0 |
| Depression symptoms Midterm | 354 | −0.50 | −0.84 to −0.16 | 2.90 | .004 | 127 | −0.23 | −0.58 to 0.12 | 1.28 | .20 | 1.21 | 1 | .27 | 17.1 |
| Depression symptoms Long term | 213 | −0.19 | −0.57 to 0.19 | 0.98 | .33 | 96 | −0.35 | −1.12 to 0.42 | 0.88 | .38 | 0.13 | 1 | .72 | 0 |

Abbreviations: CI, confidence interval; long term, ≥12 months; midterm, 6–7 months; PTSD, post-traumatic stress disorder; short term, 3–4 months; SMD, standardized mean difference; Z, test of overall effect.
### Table A5
**Subgroup analysis: Trauma with a perpetrator versus trauma without a perpetrator**

| Outcomes                  | Treatment effect | Test for subgroup differences |  |
|---------------------------|------------------|-------------------------------|---|
|                           | With a perpetrator |                               |   |
|                           | N    | SMD  | 95% CI  | Z   | p    | N    | SMD  | 95% CI  | Z   | p    | χ² | df | p  | I² (%) |
| PTSD symptoms             |       |      |         |     |      |       |      |         |     |      |    |    |    |        |
| Short term                | 732   | -3.39| -5.92 to -0.86 | 2.63 | .009 | 81   | -1.84| -5.82 to 2.13 | 0.91 | .36  | 0.42 | 1  | .52 | 0  |
| Midterm                   | 741   | -0.47| -0.70 to -0.23 | 3.91 | <.0001 | 28  | -0.31| -1.06 to 0.43 | 0.83 | .41  | 0.15 | 1  | .70 | 0  |
| Long term                 | 315   | -0.49| -0.79 to -0.19 | 3.18 | .001 | NA  |      |         |     |      |    |    |    |        |
| Depression symptoms       |       |      |         |     |      |       |      |         |     |      |    |    |    |        |
| Short term                | 363   | -0.39| -0.73 to -0.06 | 2.29 | .02  | 81   | -0.09| -0.79 to 0.62 | 0.24 | .81  | 0.59 | 1  | .44 | 0  |
| Midterm                   | 402   | -0.53| -0.85 to -0.21 | 3.17 | .002 | 28  | -0.13| -0.88 to 0.61 | 0.35 | .72  | 0.93 | 1  | .33 | 0  |
| Long term                 | 289   | -0.28| -0.53 to -0.03 | 2.23 | .03  | NA  |      |         |     |      |    |    |    |        |

Abbreviations: CI, confidence interval; long term, ≥12 months; midterm, 6–7 months; NA, not applicable; PTSD, post-traumatic stress disorder; short term, 3–4 months; SMD, standardized mean difference; Z, test of overall effect.

---

### Figure B1
PRISMA study flow diagram

Records retrieved through database search (n = 306)

Records identified after screening titles (n = 110)

Abstracts screened after duplicates removed (n = 40)

Full-text articles chosen for review (n = 28)

Studies included in meta-analysis (n = 24)

Additional records identified through other sources (n = 06)

Records that did not meet inclusion criteria (n = 14)

Full-text articles excluded due to ineligibility/lack of data (n = 2)
**FIGURE B2** Risk of bias graph [Colour figure can be viewed at wileyonlinelibrary.com]

**FIGURE B3** Funnel plots of comparison: narrative exposure therapy versus controls (all): (a) post-traumatic stress disorder (PTSD) symptoms (short term), (b) PTSD symptoms (midterm), (c) depression symptoms (short term) and (d) depression symptoms (midterm). Abbreviation: SMD, standardized mean difference [Colour figure can be viewed at wileyonlinelibrary.com]
FIGURE B4  Risk of bias summary [Colour figure can be viewed at wileyonlinelibrary.com]
FIGURE B5  Forest plot: Post-traumatic stress disorder diagnosis (short term). Abbreviations: CI, confidence interval; NET, narrative exposure therapy [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE B6  Forest plot: Post-traumatic stress disorder diagnosis (midterm). Abbreviations: CI, confidence interval; NET, narrative exposure therapy [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE B7  Forest plot: Post-traumatic stress disorder diagnosis (long term). Abbreviations: CI, confidence interval; NET, narrative exposure therapy [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE B8  Forest plot: Post-traumatic stress disorder symptoms (short term). Abbreviations: CI, confidence interval; NET, narrative exposure therapy [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE B9  Forest plot: Post-traumatic stress disorder symptoms (midterm). Abbreviations: CI, confidence interval; NET, narrative exposure therapy [Colour figure can be viewed at wileyonlinelibrary.com]
FIGURE B10  Forest plot: Post-traumatic stress disorder symptoms (long term). Abbreviations: CI, confidence interval; NET, narrative exposure therapy [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE B11  Forest plot: Depression symptoms (short term). Abbreviations: CI, confidence interval; NET, narrative exposure therapy [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE B12  Forest plot: Depression symptoms (midterm). Abbreviations: CI, confidence interval; NET, narrative exposure therapy [Colour figure can be viewed at wileyonlinelibrary.com]

FIGURE B13  Forest plot: Depression symptoms (long term). Abbreviations: CI, confidence interval; NET, narrative exposure therapy [Colour figure can be viewed at wileyonlinelibrary.com]