The effectiveness of narrative exposure therapy: a review, meta-analysis and meta-regression analysis

Jeannette C.G. Lely, Geert E. Smid, Ruud A. Jongedijk, Jeroen W. Knipscheer & Rolf J. Kleber

To cite this article: Jeannette C.G. Lely, Geert E. Smid, Ruud A. Jongedijk, Jeroen W. Knipscheer & Rolf J. Kleber (2019) The effectiveness of narrative exposure therapy: a review, meta-analysis and meta-regression analysis, European Journal of Psychotraumatology, 10:1, 1550344, DOI: 10.1080/20008198.2018.1550344

To link to this article: https://doi.org/10.1080/20008198.2018.1550344

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

Published online: 25 Mar 2019.

Submit your article to this journal

Article views: 246

View Crossmark data
The effectiveness of narrative exposure therapy: a review, meta-analysis and meta-regression analysis

Jeannette C.G. Lely, Geert E. Smid, Ruud A. Jongedijk, Jeroen W. Knipscheer and Rolf J. Kleber

Foundation Centrum, Diemen, The Netherlands; Arq Psychotrauma Expert Group, Diemen, The Netherlands; Department of Clinical Psychology, Utrecht University, Utrecht, The Netherlands

ABSTRACT

Background: Narrative exposure therapy (NET) is a short-term psychological treatment for post-traumatic stress disorder (PTSD) that has been investigated in various contexts among traumatized refugees and other trauma survivors. Sustained treatment results have been reported, but the methodological quality of the trials needs a more thorough examination.

Objective: To evaluate the effectiveness of NET for survivors of trauma, using a quality assessment, an updated meta-analysis, and a meta-regression analysis.

Method: Following a systematic literature selection, the methodological quality of the included studies was assessed; Non-controlled and controlled effect sizes (Hedges' g) were estimated using a random effects model. Predictor analyses were performed using a model of effect sizes for PTSD and depression included symptom change at post-treatment and follow-up time-points. Controlled effect sizes included post-treatment comparisons of NET with non-active and active comparators: both trauma-focused (TF) and non-trauma-focused (non-TF) interventions.

Results: The selected studies showed high external validity; methodological quality was equivalent to other guideline-supported TF interventions. In 16 randomized controlled trials, involving 947 participants, large non-controlled effect sizes were found for PTSD symptoms, at post-treatment (g = 1.18, 95% confidence interval [0.87; 1.50]) and follow-up (g = 1.37 [0.96; 1.77]). For depression symptoms, medium non-controlled effect sizes were found, at post-treatment (g = 0.47 [0.23; 0.71]) and follow-up (g = 0.60 [0.26; 0.94]). Post-treatment, NET outperformed non-active comparators and non-TF active comparators for PTSD, but not the combined active comparators. For depression, NET only outperformed non-active comparators. Advancing age predicted better treatment results for PTSD and depression symptoms.

Conclusions: The findings of this meta-analysis suggest that patients and providers may expect sustained treatment results from NET. Controlled comparisons with other guideline-supported TF interventions are not yet available.

La efectividad de la terapia de exposición narrativa: una revisión, un metanálisis y un análisis de metarregresión

Antecedentes: la terapia de exposición narrativa (NET en sus siglas en inglés) es un tratamiento psicológico a corto plazo para el trastorno de estrés postraumático (TEPT) que ha sido investigado en diversos contextos entre refugiados traumatisados y otros sobrevivientes de traumas. Se han informado resultados consistentes del tratamiento, pero la calidad metodológica de los ensayos requiere un examen más exhaustivo.

Objetivo: evaluar la efectividad de la NET para sobrevivientes de trauma, mediante una evaluación de la calidad, un metanálisis actualizado y un análisis de metarregresión.

Método: después de una selección sistemática de la literatura, se evaluó la calidad metodológica de los estudios incluidos; se estimaron los tamaños del efecto no controlados y controlados (g de Hedges), utilizando un modelo de efectos aleatorios. Se realizaron análisis predictivos. Los tamaños del efecto no controlados para el TEPT y la depresión incluyeron el cambio de los síntomas en los puntos de tiempo del postratamiento y del seguimiento. Los tamaños de los datos efectos controlados incluyeron las comparaciones post-treatment de la NET con comparadores no activos y activos: ambas intervenciones centradas en el trauma (TF en sus siglas en inglés) y no enfocadas en el trauma (no TF).

Resultados: la calidad metodológica de los estudios incluidos varió de incierto a alto riesgo de sesgo. En los 16 ensayos controlados aleatorios (ECA), con 947 participantes, se encontraron tamaños grandes de efectos no controlados para los síntomas del TEPT, en el postratamiento (g = 1.18, intervalo de confianza del 95% [0.87; 1.50]), y el seguimiento (g = 1.37 [0.96; 1.77]). Para los síntomas de depresión, se encontraron tamaños medios de efectos no controlados: en el postratamiento (g = 0.47 [0.23; 0.71]) y el seguimiento (g = 0.60 [0.26; 0.94]). En los post-tratamientos, la NET superó a los comparadores no activos y los comparadores no TF activos para el TEPT, pero no a los comparadores activos combinados. Para la depresión, la NET solo superó a los comparadores no activos. La edad avanzada predijo mejores resultados.

ARTICLE HISTORY

Received 23 March 2018
Revised 26 October 2018
Accepted 1 November 2018

KEYWORDS

Intervention; meta-analysis; narrative exposure therapy; NET; post-traumatic stress disorder; PTSD; treatment

PALABRAS CLAVES

Intervención; metanalisis; Terapia de exposición narrativa; NET; Trastorno de Estrés Postraumático; TEPT; Tratamiento

CONTACT

Jeannette C.G. Lely a,b

© 2019 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
1. Introduction

Current geopolitical developments and worldwide migration crises stress the urgency of providing effective treatment to trauma-exposed refugees. Narrative exposure therapy (NET) (Schauer, Neuner, & Elbert, 2011), a short-term trauma-focused cognitive behaviour therapy (TFCBT), has been investigated with populations from different cultural and ethnic backgrounds (Palic & Elkiti, 2011). These studies have primarily involved refugees and displaced persons with post-traumatic stress disorder (PTSD) (American Psychiatric Association [APA], 2000, 2013) from non-Western countries.

PTSD is a serious and demoralizing mental health disorder, burdening adaptation and development (McEwen, 2012). Prevalence rates range widely, from 2.7% in population surveys (Hauffa et al., 2011) to 30.6% among refugees worldwide (Steel et al., 2009). PTSD and depression frequently co-occur (Rytwinski, Scru, Feeney, & Youngstrom, 2013; Steel et al., 2009). Treatment guidelines recommend trauma-focused (TF) psychological treatment for PTSD (American Psychological Association [APA], 2017; National Institute for Clinical Excellence [NICE], 2005).

1.1. Narrative exposure therapy

NET is a standardized form of TF psychotherapy, embedding trauma exposure in an autobiographical context. The manual (Schauer et al., 2011) recommends four to 12 sessions of 90 minutes, depending on the number of traumatic events, and treatment focuses on imaginary trauma exposure and on reorganizing memories (Schnyder et al., 2015). Memories of traumatic events are hypothesized to form multiple fear networks dominated by sensory–perceptual information and lacking autobiographical information (Schauer et al., 2011). By connecting these anxiety-provoking implicit memories with episodic context, the autobiographic memory is rebuilt, allowing for reduction of anxiety (Schauer et al., 2011).

In NET, the therapist and the patient create a timeline of the patient’s life, followed by chronologically elaborating this timeline in subsequent sessions. At the end of therapy, the patient receives the written narrative as a documented testimony. Given its focus on the lifespan, NET is particularly suited to populations affected by multiple traumatic experiences (Schauer et al., 2011).

1.2. Research in context

Review findings indicated medium to large non-controlled effect sizes for NET with refugees and displaced persons (Gwozdziewycz & Mehl-Madrona, 2013; Lambert & Alhasoon, 2015; Robjant & Fazel, 2010), even in extremely insecure living conditions. Treatment gains for PTSD symptoms and comorbid disorders were found to be stable at follow-up. Treatment dropout was low (Mørkved et al., 2014), and provision of the intervention by trained counsellors was found to be feasible (Gwozdziewycz & Mehl-Madrona, 2013). Controlled effect sizes have been found to be in line with those of other guideline-supported interventions (Cusack et al., 2016; Watts...
et al., 2013). The methodological quality of the NET trials, however, has been questioned (Bisson, Roberts, Andrew, Cooper, & Lewis, 2013; Mundt, Wünsche, Heinz, & Pross, 2014; Patel, Kellezi, & Williams, 2014). These doubts suggest that the findings need more thorough investigation.

In the review literature on NET, some serious omissions were found. Some reviews were merely descriptive (Robjant & Fazel, 2010), while some were exclusively focused on low-resource settings (Mundt et al., 2014) or survivors of torture (Patel et al., 2014). Some missed important trials (Bisson et al., 2013; Patel et al., 2014) and others lacked information on comorbid depressive symptoms (Gwodzdziwicz & Mehl-Madrona, 2013) or follow-up effects (Lambert & Alhassoon, 2015). As yet, no PTSD guideline-supported comparators have been found (Gerger, Munder, & Barth, 2014; Nosé et al., 2017). Meanwhile, new NET trials have been added to the evidence base of NET, calling for evaluation.

In the current study, these gaps in the literature were addressed and several clinically important moderators were examined as predictors of treatment outcome (e.g. advancing age, gender, migration to Western countries, treatment dose, self-report assessments, language matching between participants and therapists, and the influence of providing NET by trained counsellors). In exposure treatment, participants’ advancing age was found to be associated with larger effects (Rizvi, Vogt, & Resick, 2009). Psychotherapy studies with more women were also found to have larger effects (Stenmark, Guzey, Elbert, & Holen, 2014; Tarrier, Sommerfield, Pilgrim, & Faragher, 2000; Watts et al., 2013). Migration to Western countries, leading to potential post-migration stressors (Li, Liddell, & Nickerson, 2016), may complicate treatment progression. A higher number of sessions was found to predict higher treatment effects (Lambert & Alhassoon, 2015). Treatment dose, capturing session length as well (Mørkved et al., 2014), has not yet been examined as a predictor. Contrast findings existed for the influence of self-reported PTSD symptoms versus diagnostic interviews (Bisson et al., 2013; Lambert & Alhassoon, 2015) and language matching between participants and therapists (Dossa & Hatem, 2012; Lambert & Alhassoon, 2015; Nickerson, Bryant, Silove, & Steel, 2011; Patel et al., 2014). Finally, updating the evidence on the effectiveness of para-professional counsellors (Gwozdziwicz & Mehl-Madrona, 2013) was considered relevant.

1.3. Aims and hypotheses

The first aim of the present study was to assess the efficacy of NET in adults, as both non-controlled and controlled effect sizes. The second aim was to critically assess the methodological quality of the current evidence on NET and to inform clinicians about treatment outcome predictors, facilitating optimization of treatment response. PTSD and depression outcomes have been included, as have post-treatment and follow-up assessments. Comparisons included non-active and active (both non-TF and TF) comparators, followed by a meta-regression analysis of the moderators described above.

Consistent with meta-analytic findings, the methodological quality of the evidence was expected to be modest. NET was expected to show medium to large and sustained treatment effects regarding PTSD and depression outcomes. In addition, NET was expected to outperform non-active and active non-TF comparators for PTSD outcomes (Lambert & Alhassoon, 2015). According to direct comparisons of TF interventions (Brom, Kleber, & Difares, 1989; Nijdam, Gerger, & Munder, 2015; Ford, Pross, & Heinz, 2010; Heide, Mooren, Van de Schoot, De Jongh, & Kleber, 2016), no significant differences were expected comparing NET with all active control conditions (non-TF and TF comparators). Finally, advancing age and female gender were hypothesized to predict larger treatment effects, while having migrated to Western countries was hypothesized to predict lower treatment effects. A higher treatment dose, self-reported PTSD symptoms, counsellors, and language matching were hypothesized to predict larger treatment effects.

2. Method

2.1. Search strategy and data sources

The literature search was conducted independently by the first author and a senior librarian. The search conformed to the Meta-Analysis Reporting Standards (MARS) guidelines (American Psychological Association [APA], 2010) and OvidSP software (2010) was used. The searched internet databases included the ACP Journal Club, Cochrane Central Register of Controlled Trials, Cochrane Database of Systematic Reviews, Cochrane Methodology Register, Database of Abstracts of Reviews of Effects, Health Technology Assessment, National Health Service Economic Evaluation Database, Ovid MEDLINE(R), Ovid MEDLINE(R) in-Process, Ovid MEDLINE(R) Daily Update, Psych INFO, and Published International Literature On Traumatic Stress (PILOTS). Trials published between 1 January 2004 (the first NET trial being published in 2004) and 30 April 2018 were included, and no language restrictions were applied. Search terms in titles and abstracts were: ‘narrative exposure therapy’ OR ‘NET’) AND (‘posttraumatic stress disorder’ OR ‘PTSD’) AND (‘treatment’ OR ‘intervention’). Additional studies were identified by cross-referencing.
2.2. Inclusion criteria and selection of studies

The search strategy targeted independent randomized controlled trials (RCTs) investigating NET outcomes (NET and its brief version, NET-R) compared with TF or non-TF comparators. Studies had to report at least one quantitative measure of PTSD assessed both pre- and post-treatment, and intention-to-treat analyses had to be applied.

2.3. Data extraction and coding

The included interventions were coded as NET or NET-R. Comparators were coded as active or non-active, and TF or non-TF interventions. If a trial included more than one comparison, the most active comparator was selected. PTSD outcome assessments were coded as diagnostic interviews or self-report instruments. Post-treatment and last follow-up effect sizes were computed. Treatment dropout indicated the percentage of participants leaving treatment prematurely. Inter-assessment correlations were obtained from the authors for all studies except for one trial. For that study, the averaged correlations of the other studies were included.

2.4. Quality appraisal

All studies were appraised using the Cochrane Collaboration’s tool for assessing risk of bias (Higgins et al., 2011), focusing on six domains threatening internal validity: random sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective reporting and other bias. The risk of bias of each domain was judged as high, low, or unclear. Consistent with the GRADE framework (Higgins et al., 2011), for each study, the highest risk of bias was considered to determine the quality. Two independent assessors (the first author and a research assistant) agreed on 94% of the resulting scores, and remaining inconsistencies were resolved by a third independent assessor (the fourth author).

2.5. Data synthesis

2.5.1. Heterogeneity

To evaluate heterogeneity in the outcomes, Cochrane’s heterogeneity statistic \(Q\) was computed and tested for significance.

2.5.2. Effect size calculation

First, non-controlled (within-group) effect sizes were computed for the NET groups and control groups (post-treatment and last follow-up, the latter indicated as ‘follow-up’). In addition, the post-treatment effect sizes for NET were grouped according to the participants’ backgrounds. Secondly, controlled (between-groups) effect sizes were computed. All effect sizes were calculated in Hedges’ \(g\). Positive signs indicated symptom reduction or treatment groups outperforming comparators.

2.5.3. Meta-analytic techniques

Effect sizes (random effects model) were computed using the Comprehensive Meta-Analysis program (CMA; version 2.0, 2005). All effect sizes were computed for PTSD outcomes and, if available, for depression outcomes. Publication bias, meaning preferential publication of striking findings, was evaluated by inspecting the funnel plots as implemented in CMA and using Egger’s test of the intercept to evaluate the significance of the bias. The random effects version of Duval and Tweedie’s trim-and-fill procedure (2000), as implemented in CMA, was used to estimate adjustments for the possible effects of publication bias. Finally, in a sensitivity analysis, each study was consecutively removed as a potential outlier to test its influence on the overall effect sizes.

2.6. Meta-regression analyses

Multiple regression analyses were performed to examine the effect of moderators on outcomes. SPSS version 23 (IBM Corp., Armonk, NY, USA), was used, with macros provided by Lipsey and Wilson (2001). For all analyses, the alpha level was set at \(p = .05\). PTSD and depression symptom change at post-treatment were used as dependent variables. Seven potential predictors were specified as independent variables: age (sample mean), female gender (sample percentage), migration (1 = yes, 0 = no), treatment dose (number of sessions by session length, yielding total treatment time in minutes), matching languages (1 = yes, 0 = no), PTSD outcome (1 = diagnostic interview, 0 = self-report) and, finally, therapist’s training being professional or paraprofessional (1 = professional, 0 = paraprofessional). All variables were entered in the model at once. The contribution of all variables to the explained variance was evaluated by calculating \(R^2\).

3. Results

3.1. Study selection

From 182 eligible studies, 16 RCTs (denoted as \(k\)) were selected for the meta-analysis, using several steps. A primary search was conducted in July 2015 and repeated in January 2016; a secondary search was conducted in July 2016 and confirmed in September 2018. The selection process is presented in Figure 1.
3.2. Study characteristics

As one study (Zang, Hunt, & Cox, 2014) included two comparisons (for NET and NET-R), the selection included 16 trials with 17 comparisons. Most NET samples \((k = 13, 76.47\%)\) were small, ranging from five to 29 participants; a minority \((k = 4, 23.52\%)\) ranged from 38 to 111 participants. Characteristics of the trials are listed in Table 1.

3.3. Methodological quality

In most studies, the risk of bias was rated as unclear. In one study, the risk of bias was rated as high. With respect to blinding and allocation concealment, no evidence was reported that violation affected outcomes. Five studies (31%) were conducted by authors who were not involved in the development of NET.

The quality assessment data are included in Figure 2.

3.4. Participants

The selected studies involved 947 individuals in nine countries; 443 (46.78%) participants were offered a NET intervention; 294 (31.05%) participated in an active control intervention, and 210 (22.18%) participated in a non-active control condition. Participants included adults (81.94%) and adolescents (18.06%); 456 participants (48.15%) were female. Mean age ranged from 17.50 to 68.90 years \((M = 37.01, SD = 14.97)\).

Eleven studies (68.75%) involved refugees or displaced persons, identified as refugees, including 783 ethnically diverse participants. Five studies (31.25%), involving 507 participants, were conducted in low-income regions (Ertl, Pfeiffer, Schauer, Elbert, & Neuner, 2011; Jacob, Neuner, Maedl, Schaal, & Elbert, 2014; Neuner et al., 2008; Neuner, Schauer, Klaschik, Karunakara, & Elbert, 2004; Schaal, Elbert, & Neuner, 2009). Six studies (37.50%), with 276 participants, involved migrated refugees or asylum seekers (identified as refugees as well) in Western countries (Hensel-Dittmann et al., 2011; Hijazi et al., 2014; Morath et al., 2014a, 2014b; Neuner et al., 2010; Stenmark, Catani, Neuner, Elbert, & Holen, 2013). All refugee studies included participants surviving multiple traumatic events or torture, the mean number ranging from 6.82 \((SD = 2.09)\) to 19.85 \((SD = 6.4)\). Consequently, those populations may be characterized by high problem complexity (Gerger et al., 2014). Finally, five studies (31.25%), including 164 individuals, involved non-refugee participants (Alghamdi, Hunt, & Thomas, 2015; Al Hadethe, Hunt, Al-Qaysi, & Thomas, 2015; Bichescu, Neuner, Schauer, & Elbert, 2007; Zang, Hunt, & Cox, 2013; Zang et al., 2014). Three of those studies (Alghamdi et al., 2015; Bichescu et al., 2007; Zang et al., 2013) reported multiple traumatic events as well.

3.5. Interventions

The mean number of sessions was 6.47 \((SD = 3.17, range 3–12)\), mean session length was 97.06 minutes \((SD = 17.68, range 75–135)\), and mean treatment dose was 631.18 minutes \((SD = 316.51, range 225–1080)\).
3.6. Comparators

Comparators included non-active \((k = 9; 56.25\%)\) and active interventions \((k = 8; 47.06\%)\). The non-active comparators consisted of waiting lists. The active comparators included one intervention that could be considered TF (Neuner et al., 2008) and one study in which NET was compared with another TF intervention (Al Hadethe et al., 2015). The other active comparators included diverse psychosocial interventions. All comparators are listed in Table 1.

3.7. Outcomes

All studies reported group means; in addition, three studies reported individual change and eight studies reported remission of PTSD diagnosis. All studies
reported reasons for dropout. In five studies, incidents of symptom increase were reported, with no indication of more adverse events among NET participants than in the control group being apparent. In one study, a suicide took place during follow-up in the control condition. None of the adverse events was attributed to treatment stress.

PTSD and depression severity were monitored with well-accepted diagnostic instruments. Time-points of post-treatment measurements varied from 2 to 26 weeks \((M = 9.55, SD = 8.26)\). Time-points of the last follow-up measurements ranged from 9 to 52 weeks \((M = 35.57, SD = 17.38)\). Treatment dropout from NET ranged from 0.00% to 25.49% \((M = 5.85%, SD = 8.51%)\), whereas treatment dropout from the comparators ranged from 0.00% to 26.67% \((M = 7.11%, SD = 9.56%)\).

### 3.8. Treatment effects

#### 3.8.1. Effect sizes

For PTSD, the non-controlled effect sizes were found to be large at post-treatment \((g = 1.18, 95\% \text{ confidence interval } [0.87; 1.50])\) and sustained at follow-up \((g = 1.35 [0.93; 1.77])\). Controlled effect sizes for all active comparators (both TF and non-TF) were non-significant. For active non-TF comparators, controlled effect sizes were medium and significant \((g = 0.43 [0.09; 0.78])\); for non-active comparators, they were large and significant \((g = 1.37 [0.66; 2.07])\).

For depression, the non-controlled effect sizes were found to be medium at post-treatment \((g = 0.47 [0.23; 0.71])\) and sustained at follow-up \((g = 0.60 [0.26; 0.94])\). Controlled effect sizes for all active comparators (both TF and non-TF) were non-significant, and the same applied for active non-TF comparators. For non-active comparators, controlled effect sizes were medium to large and significant \((g = 0.79 [0.48; 1.09])\). The resulting effect sizes are presented in Table 2 and Figure 2.

To take into account the participants’ backgrounds, the mean non-controlled effect sizes at post-test were grouped with respect to refugees in their home regions \((g = 1.03 [0.47; 1.60])\), refugees in Western countries \((g = 1.07 [0.55; 1.60])\), and non-refugees \((g = 1.66 [0.83; 2.48])\). The mean numbers of participants in these subgroups were 41.40, 26.50, and 12.83, respectively. Finally, the non-controlled effect sizes of the comparators are presented in Table 3.

For PTSD, these effect sizes were found to be large at post-treatment \((g = 0.77 [0.41; 1.12])\) and sustained at follow-up \((g = 1.08 [0.55; 1.62])\). For depression, these effect sizes were found to be small at post-treatment \((g = 0.34 [0.08; 0.59])\) and sustained at follow-up \((g = 0.67 [0.26; 1.12])\).

#### 3.8.2. Publication bias

For controlled effect sizes, visual inspection of the funnel plots did not suggest possible publication bias, for both time-points and both outcomes. Duval and Tweedie’s trim-and-fill procedure \((2000)\) indicated unchanged results. For the non-controlled effect sizes of PTSD, however, seven studies were missing; and for the non-controlled effect sizes of depression, two studies were missing at post-treatment and three at follow-up. Imputed effect sizes for PTSD outcomes remained medium post-treatment \((g = 0.71 [0.37; 1.06])\) and at follow-up \((g = 0.79 [0.37; 1.20])\). Imputed effect sizes for depression outcomes were

---

**Table 2. Mean effect sizes of narrative exposure therapy (NET) on post-traumatic stress disorder (PTSD) and depression outcomes, in Hedges’ \(g\).**

<table>
<thead>
<tr>
<th></th>
<th>Effect sizes</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(k)</td>
<td>(\bar{g})</td>
</tr>
<tr>
<td><strong>PTSD</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-controlled effect sizes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre- to post-treatment</td>
<td>17</td>
<td>1.18</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>1.35</td>
</tr>
<tr>
<td>Controlled effect sizes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All active comparators</td>
<td>9</td>
<td>0.24</td>
</tr>
<tr>
<td>Active non-TF comparators</td>
<td>7</td>
<td>0.43</td>
</tr>
<tr>
<td>Non-active comparators</td>
<td>8</td>
<td>1.37</td>
</tr>
<tr>
<td>All comparators</td>
<td>17</td>
<td>0.69</td>
</tr>
<tr>
<td><strong>Depression</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non-controlled effect sizes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre- to post-treatment</td>
<td>13</td>
<td>0.47</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.60</td>
</tr>
<tr>
<td>Controlled effect sizes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All active comparators</td>
<td>7</td>
<td>0.17</td>
</tr>
<tr>
<td>Active non-TF comparators</td>
<td>6</td>
<td>0.33</td>
</tr>
<tr>
<td>Non-active comparators</td>
<td>6</td>
<td>0.79</td>
</tr>
<tr>
<td>All comparators</td>
<td>13</td>
<td>0.49</td>
</tr>
</tbody>
</table>

*Note: CI, confidence interval; \(Q\), Cochrane’s heterogeneity statistic; df, degrees of freedom. All active comparators: Trauma focused comparators (Emotional Freedom Techniques and Trauma Counselling) plus active non-TF comparators; Active non trauma focused (non-TF) comparators, Academic Catch-up Programme; Interpersonal Psychotherapy; Psychoeducation; Stress Inoculation Training; Supportive Counselling, and Treatment as Usual. Non-active comparators, waiting list conditions.
Table 3. Mean non-controlled effect sizes of the comparators in Hedges’ $g$ ($g = 0.02$ small; $g = 0.05$ medium; $g = 0.08$ large) (Cohen, 1992).

<table>
<thead>
<tr>
<th>Effect sizes</th>
<th>$k$</th>
<th>Mean</th>
<th>$-95%$ CI*</th>
<th>$95%$ CI</th>
<th>$p$-value</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>All comparators</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTSD pre–post</td>
<td>17</td>
<td>0.77</td>
<td>0.41</td>
<td>1.12</td>
<td>.000</td>
<td>162.81</td>
</tr>
<tr>
<td>PTSD pre–FU*</td>
<td>12</td>
<td>1.08</td>
<td>0.55</td>
<td>1.62</td>
<td>.000</td>
<td>95.48</td>
</tr>
<tr>
<td>Depression pre–post</td>
<td>13</td>
<td>0.34</td>
<td>0.08</td>
<td>0.59</td>
<td>.010</td>
<td>43.18</td>
</tr>
<tr>
<td>Depression pre–FU</td>
<td>9</td>
<td>0.67</td>
<td>0.26</td>
<td>1.12</td>
<td>.002</td>
<td>37.07</td>
</tr>
<tr>
<td>Active comparators</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTSD pre–post</td>
<td>8</td>
<td>0.69</td>
<td>0.22</td>
<td>1.16</td>
<td>.004</td>
<td>39.24</td>
</tr>
<tr>
<td>PTSD pre–FU</td>
<td>5</td>
<td>0.57</td>
<td>0.17</td>
<td>1.32</td>
<td>.133</td>
<td>32.45</td>
</tr>
<tr>
<td>Depression pre–post</td>
<td>3</td>
<td>0.31</td>
<td>-0.06</td>
<td>0.67</td>
<td>.100</td>
<td>1.233</td>
</tr>
<tr>
<td>Depression pre–FU</td>
<td>3</td>
<td>0.47</td>
<td>-0.10</td>
<td>1.04</td>
<td>.108</td>
<td>3.79</td>
</tr>
<tr>
<td>Active TF* comparators</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTSD pre–post</td>
<td>2</td>
<td>1.79</td>
<td>1.02</td>
<td>2.56</td>
<td>.000</td>
<td>3.7</td>
</tr>
<tr>
<td>PTSD pre–FU</td>
<td>2</td>
<td>1.50</td>
<td>0.30</td>
<td>2.71</td>
<td>.015</td>
<td>7.20</td>
</tr>
<tr>
<td>Depression pre–post</td>
<td>1</td>
<td>0.97</td>
<td>0.41</td>
<td>1.54</td>
<td>.001</td>
<td>0.00</td>
</tr>
<tr>
<td>Depression pre–FU</td>
<td>1</td>
<td>1.14</td>
<td>0.50</td>
<td>1.79</td>
<td>.001</td>
<td>0.00</td>
</tr>
<tr>
<td>Non-active comparators</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PTSD pre–post</td>
<td>9</td>
<td>0.19</td>
<td>-0.08</td>
<td>0.47</td>
<td>.163</td>
<td>35.10</td>
</tr>
<tr>
<td>PTSD pre–FU</td>
<td>7</td>
<td>1.40</td>
<td>0.54</td>
<td>2.26</td>
<td>.001</td>
<td>71.23</td>
</tr>
<tr>
<td>Depression pre–post</td>
<td>7</td>
<td>0.05</td>
<td>-0.09</td>
<td>0.19</td>
<td>.498</td>
<td>1.32</td>
</tr>
<tr>
<td>Depression pre–FU</td>
<td>6</td>
<td>0.54</td>
<td>-0.01</td>
<td>1.09</td>
<td>.052</td>
<td>32.00</td>
</tr>
</tbody>
</table>

*PTSD, post-traumatic stress disorder; FU, follow-up; TF, trauma-focused; CI, confidence interval; Q, Cochrane’s heterogeneity statistic; df, degrees of freedom.

small post-treatment ($g = 0.39 [0.14; 0.63]$) and at follow-up ($g = 0.36 [0.01; 0.71]$).

3.8.3. Sensitivity analyses

For PTSD and depression outcomes, no study was found to exert a disproportionally large effect on the overall effect sizes, for both time-points, involving non-controlled and controlled effect sizes.

3.9. Predictor analyses

For PTSD, treatment results were predicted by advancing age, but not by the other moderators. The model explained 34% of the inter-study variance. The residual heterogeneity was found to be non-significant (residual $Q = 22.04; df = 9; p = .088$)

For depression, treatment results were predicted by advancing age and migration, but not by the remaining moderators. The model explained 97% of the inter-study variance. Residual heterogeneity was found to be non-significant (residual $Q = 0.25; df = 7; p = .976$). The details of the multiple regression analyses are presented in Table 4.

4. Discussion

4.1. Main findings

This meta-analysis shows that NET is effective in the reduction of PTSD and depression symptoms across diverse, predominantly war-affected refugee populations. Treatment results are better for older adults. NET was found to be superior to non-active comparators and active non-TF comparators. As yet, no controlled comparisons with other guideline-supported TF interventions were not available. The application of paraprofessional therapists and very low treatment doses in trials can be seen as strengthening external validity. Using the GRADE framework, the risk of bias in the primary studies was found to range from unclear to high.

4.2. Predictors

Advancing age predicted better treatment outcomes for PTSD and depression symptoms. These findings may be explained by the lifespan perspective of NET. Trauma exposure with several retellings may foster cognitive coherence, self-reflection, and restored meaning in life. Thus, it may appeal to the needs of older adults (Rizvi et al., 2009). Unexpectedly, a higher proportion of women did not show significant associations with treatment results. Such deviations from earlier findings may be explained by differences in sample size and sample specificity. The sample of Watts et al. (2013) contained nearly 10 times more participants than the present sample. In the study by Stenmark et al. (2014), male gender may have been influenced by violent offences. In the study by Tarrier et al. (2000), the female gender may have been influenced by psychoticism or treatment motivation. Migration to Western countries predicted smaller treatment results for depression, but not for PTSD. This finding implies that treatment response for PTSD is not influenced by post-migration stress. For the treatment of depression, however, NET appears to be less satisfying. Higher treatment dose showed no significant association with treatment results. This finding may encourage more research on NET with brief treatment formats.
Outcome assessment did not predict treatment outcomes; neither did language matching. The latter finding supports that of Lambert and Alhassoon (2015). Although this finding may not take away methodologists’ concerns regarding assessments, it implies that the assistance of interpreters does not influence treatment outcomes. Similar conclusions were found for ethnic matching in providing mental healthcare to Mediterranean migrants in the Netherlands (Knipscheer & Kleber, 2004). Unexpectedly, treatment progression was not predicted by the therapist’s training. This finding supports the contribution of trained paraprofessionals as NET therapists, being highly relevant in low-resource settings.

### 4.3. Findings in context

The included NET trials have been conducted in various settings and, consequently, in different healthcare systems. NET for non-refugees was found to be highly efficacious, albeit involving notably smaller treatment groups than the other studies. Remarkably, NET trials with migrated refugees (treated in healthcare systems in Western countries) did not show higher mean effect sizes than those with refugees in their home regions. Apparently, refugees in Western countries did not benefit more from NET in the settings of those countries. This finding may be explained by the influence of time on persisting PTSD symptoms. In current network analyses of PTSD, a common factor (trauma) is assumed to cause the onset of PTSD (influenced by vulnerability and protecting variables), whereas the persistence of the disorder seems to be governed by a larger network of variables (Fried & Cramer, 2017). Although the scope of the present study does not allow for a direct comparison, it can be hypothesized that refugees resettled in Western countries on average suffer from chronic symptoms, potentially complicated by post-migration stressors, especially language barriers and the necessity to adjust to Western cultures (Li et al., 2016). Addressing those symptoms with the help of NET – focusing on the past – can be less efficacious than might be expected in high-quality healthcare systems. This finding may call for research investigating whether migrated refugees benefit as much from NET as from other guideline-supported TF interventions. To effectively treat resettled refugees in Western healthcare systems, individualized decisions are needed, based on comparative evidence of different TFCBT interventions or eye movement desensitization and reprocessing (EMDR). Some meta-analytic findings show effect sizes of TFCBT to be superior to those of the current study (Bradley, Greene, Russ, Dutra, & Westen, 2005; Watts et al., 2013). The findings of the current study are, however, in accordance with recent meta-analytic findings, involving participants with complex problems (Gerger et al., 2014) or refugees resettled in Western countries (Nosè et al., 2017). The results were also consistent with those involving survivors of childhood abuse (Ehrling et al., 2014).

The comparators varied from waiting-list conditions to active psychosocial interventions. The active interventions suggested contextual face validity and credibility. By using the strict criteria of the GRADE framework for the evaluation of the trials, the assessment focused on the internal validity of the studies, providing rather conservative ratings. Factors of external validity included intended face validity and credibility of the comparators, the complexity of the difficulties within the target population, and the local circumstances.

The framework of GRADE reduces the overall methodological quality to the weakest qualification,

---

Table 4. Meta-regression analyses relating mean age, being female, migration, treatment dose, matching languages, outcome, and therapist’s training to non-controlled effect sizes.

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Moderator</th>
<th>R²</th>
<th>Q² (df = 7)</th>
<th>B*</th>
<th>SE*</th>
<th>B*</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre- to post treatment</td>
<td>Mean Age</td>
<td>.34</td>
<td>22.04 (9)</td>
<td>.05</td>
<td>0.02</td>
<td>.75</td>
<td>.088</td>
</tr>
<tr>
<td></td>
<td>Female Gender</td>
<td></td>
<td></td>
<td>-.01</td>
<td>0.01</td>
<td>-.39</td>
<td>.145</td>
</tr>
<tr>
<td></td>
<td>Migration</td>
<td></td>
<td></td>
<td>-.70</td>
<td>0.41</td>
<td>-.44</td>
<td>.084</td>
</tr>
<tr>
<td></td>
<td>Treatment Dose</td>
<td></td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>0.41</td>
<td>.325</td>
</tr>
<tr>
<td></td>
<td>Matching Languages</td>
<td></td>
<td></td>
<td>-.48</td>
<td>0.60</td>
<td>-.30</td>
<td>.421</td>
</tr>
<tr>
<td></td>
<td>Outcome</td>
<td></td>
<td></td>
<td>-.15</td>
<td>0.45</td>
<td>-.09</td>
<td>.744</td>
</tr>
<tr>
<td></td>
<td>Therapist’s Training</td>
<td></td>
<td></td>
<td>-.33</td>
<td>0.72</td>
<td>-1.18</td>
<td>.643</td>
</tr>
<tr>
<td>Depression</td>
<td>Mean Age</td>
<td>.97</td>
<td>.82 (5)</td>
<td>.33</td>
<td>0.14</td>
<td>.99</td>
<td>.039</td>
</tr>
<tr>
<td></td>
<td>Female Gender</td>
<td></td>
<td></td>
<td>0.00</td>
<td>0.01</td>
<td>-.04</td>
<td>.993</td>
</tr>
<tr>
<td></td>
<td>Migration</td>
<td></td>
<td></td>
<td>-.46</td>
<td>0.19</td>
<td>-.60</td>
<td>.018</td>
</tr>
<tr>
<td></td>
<td>Treatment Dose</td>
<td></td>
<td></td>
<td>0.00</td>
<td>0.00</td>
<td>-.22</td>
<td>.793</td>
</tr>
<tr>
<td></td>
<td>Matching Languages</td>
<td></td>
<td></td>
<td>-.96</td>
<td>0.98</td>
<td>-.11</td>
<td>.328</td>
</tr>
<tr>
<td></td>
<td>Outcome</td>
<td></td>
<td></td>
<td>-.22</td>
<td>0.36</td>
<td>-.26</td>
<td>.548</td>
</tr>
<tr>
<td></td>
<td>Therapist’s Training</td>
<td></td>
<td></td>
<td>0.17</td>
<td>1.18</td>
<td>0.10</td>
<td>.887</td>
</tr>
</tbody>
</table>

PTSD, post-traumatic stress disorder; R², explained variance; Q, Cochrane’s heterogeneity statistic; df, degrees of freedom; B, intercept; SE, standard error; β, regression coefficient.
which presents serious dilemmas in psychotherapy research. Whereas underrating the risks of bias may lead to accepting disappointing treatments, overrating such risks may withhold effective interventions from patients in need (Patel et al., 2014). As for NET, numerous studies, qualitatively equivalent to other psychotherapy research, conducted with high external validity, have supported evidence on the safety and efficacy of NET across various populations. Therefore, NET has been described as being of proven effectiveness (Jong, Knipscheer, Ford, & Kleber, 2014) and can be considered the psychotherapeutic treatment of choice in post-conflict settings.

4.4. Strengths and limitations

The present study is the first meta-analysis summarizing the current evidence base of NET. Its strengths include a systematic literature selection, a domain-based methodological quality appraisal, a differentiation between TF and non-TF control conditions, and, finally, the exploration of heterogeneity in the main outcomes by meta-regression analyses.

Simultaneously, this study has a number of limitations. First, restricted study quality limited the interpretation of results. This is not unexpected, since the majority of TFCBT interventions shares this qualification (Bisson et al., 2013; Patel et al., 2014). Secondly, the current absence of guideline-supported TF comparators presents a persisting research gap. Thirdly, the short mean length of follow-up intervals may be considered a limitation. In a long follow-up interval, the protection of the intervention against new traumatic exposure can be assessed. Fourthly, the findings regarding follow-up outcomes were based on fewer studies, limiting the conclusions about stability of follow-up results. Fifthly, the heterogeneity of the PTSD outcomes was only partly explained by the selected predictors. Furthermore, for the non-controlled effect sizes, the possibility of publication bias was suggested. This finding calls for extended research. Finally, the high frequency of a limited group of authors may present a risk of confounding on grounds of therapist effects and allegiance bias (Wampold et al., 2010).

4.5. Clinical implications

Notwithstanding shared methodological weaknesses, in recent guidelines TFCBT and EMDR are recommended as the most effective treatments for PTSD (Foa, 2009; Hamblen, Schnurr, Rosenberg, & Eftekhari, 2009). NET has been considered qualitatively comparable to those evidence-based trauma therapies (Hoge, 2011), and is currently included in the suggested interventions for treating PTSD in adults (APA, 2017).

Despite the need for caution in interpretation, this study’s findings provide evidence for empirically informed decision making in clinical practice and research policy. Providing NET to adults with high problem complexity may be expected to be safe and effective. Symptoms of depression require specific attention when treating refugees in Western countries. In clinical care, the findings imply that no reluctance is needed with regard to the provision of NET with interpreters. The same accounts for trained paraprofessionals providing NET in low-resource settings.

4.6. Future research

Although rigorous research methodology and the needs of clinical care in naturalistic settings may be hard to align, some research options can be outlined. To enhance the methodological quality of future research, larger samples should be used, and individual treatment results, including adverse developments, should be systematically reported. These recommendations also apply to the generalization of the results of non-refugee trials. To explain the residual heterogeneity for PTSD outcomes, more research is recommended on trauma-related variables such as traumatic load, which may influence treatment response (Lonergan, 2014), and clinical variables such as the influence of comorbid depression symptoms (Haagen, Ter Heide, Mooren, Knipscheer, & Kleber, 2017). In addition, to assess long-term treatment results, future research should preferably follow up patients for 2 years after the end of treatment (Bradley et al., 2005). Next, strengthening comparative evidence of NET relative to TF comparators would seem to be supportive. Direct comparisons, however, hardly yield clear differences (Tran & Gregor, 2016). More can be expected from studies with high-frequency measuring, which may reveal relevant information on response patterns or treatment strategies (Nijdam et al., 2012). Furthermore, the development of future NET research by independent research groups would strengthen the evidence for NET. Finally, applications of NET beyond refugee populations are awaiting qualitatively rigorous research.

5. Conclusions

Although the evidence regarding the treatment effects of NET must be interpreted with caution, this meta-analysis provides empirical support for the effectiveness of NET for traumatized and highly burdened populations, facilitating optimization of treatment response in clinical care.
Acknowledgments

Foundation Centrum ‘45, partner in Arq Psychotrauma Expert Group, funded this study. The authors wish to thank the following persons for their contributions to this meta-analysis: Jonna Lind and Wouter de Nooij of Arq library; research assistants Hilde van de Beek, Heleen van Dinhé, and Zeynep Karaca; and the authors replying to the requests for information.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Jeanette C.G. Lely http://orcid.org/0000-0003-1880-419X

References


