Available online at www.sciencedirect.com
ScienceDirect
Clinical Psychology Review 28 (2008) 746–758
CLINICAL PSYCHOLOGY REVIEW

The relative efficacy of bona fide psychotherapies for treating post-traumatic stress disorder: A meta-analysis of direct comparisons

Steven G. Benish *, Zac E. Imel, Bruce E. Wampold

University of Wisconsin — Madison, United States

Received 4 June 2007; received in revised form 8 October 2007; accepted 23 October 2007

Abstract

Psychotherapy has been found to be an effective treatment of post-traumatic stress disorder (PTSD), but meta-analyses have yielded inconsistent results on relative efficacy of psychotherapies in the treatment of PTSD. The present meta-analysis controlled for potential confounds in previous PTSD meta-analyses by including only bona fide psychotherapies, avoiding categorization of psychotherapy treatments, and using direct comparison studies only. The primary analysis revealed that effect sizes were homogenously distributed around zero for measures of PTSD symptomology, and for all measures of psychological functioning, indicating that there were no differences between psychotherapies. Additionally, the upper bound of the true effect size between PTSD psychotherapies was quite small. The results suggest that despite strong evidence of psychotherapy efficaciousness vis-à-vis no treatment or common factor controls, bona fide psychotherapies produce equivalent benefits for patients with PTSD.

Contents

* Corresponding author. Department of Counseling Psychology, University of Wisconsin — Madison, 1000 Bascom Mall, 323 Education Building, Madison, WI 53703, United States. Tel.: +1 608 963 9081.
E-mail address: sgbenish@wisc.edu (S.G. Benish).

0272-7358/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cpr.2007.10.005

Post-traumatic stress disorder (PTSD) is an anxiety disorder that follows a person's exposure to a traumatic stressor and is characterized by three clusters of symptoms: (a) re-experiencing of the traumatic event (e.g., intrusive thoughts, nightmares), (b) avoidance of stimuli associated with or generalized to the trauma, and (c) hyperarousal such as exaggerated startle response or hypervigilance (APA, 1994). Although a minority of persons exposed to trauma develops DSM-IV diagnosable PTSD (Kessler, 2000), it is still one of the most common mental disorders with an estimated lifetime prevalence of approximately 8% (Green & Kaltman, 2002; Kessler, Sonnega, Bromet, Hughes, & Nelson, 1995). Prevalence rates for recent U.S. military veterans are even higher, with rates of up to 20% for combat veterans returning from Iraq and Afghanistan (Hoge et al., 2004). In addition, comorbidity is quite common in patients with PTSD (Engdahl, Dikel, Eberly, & Blank, 1998), with 88% of men and 79% of women with PTSD meeting diagnostic criteria for another mental disorder (Kessler et al., 1995). These statistics highlight the seriousness of PTSD as a public health problem and the necessity of effective treatment.

Over the past few decades, effects of psychotherapy on a wide variety of patients and disorders have been studied to determine: (a) *if* treatments are efficacious, and (b) *which* treatments might be most efficacious. Many controlled trials and numerous meta-analyses of these trials have left little doubt that psychotherapy is a highly efficacious treatment for a broad range of disorders and populations (e.g., Lambert & Ogles, 2004; Wampold, 2001). However, the question of which psychotherapies are most efficacious for specific disorders, including PTSD, is currently the focus of much research.

Substantial evidence supports efficacy of several specific psychological treatments of PTSD, including prolonged exposure (PE), eye movement desensitization reprocessing (EMDR), stress inoculation therapy, trauma management therapy, cognitive therapy, and others (Bisson & Andrew, 2005; Bisson et al., 2007; Bradley, Greene, Russ, Dutra, & Westen, 2005; Van Etten & Taylor, 1998), but consensus does not exist with regard to the relative efficacy of these treatments. To wit, some argue that certain psychotherapies for PTSD have more empirical support than others (e.g., exposure therapy; see e.g., Foa, Keane, & Friedman, 2004; Nemeroff, Bremmer, Foa, Mayberg, North, & Stein, 2006; Robertson, Humphreys, & Ray, 2004), while other researchers assert that no specific psychotherapy has proven to be a gold standard (e.g., Lee, Taylor, & Drummond, 2006; McFarlane & Yehuda, 2000).

Meta-analysis is a method that can be used to systematically examine relative efficacy issues, with advantages over individual studies or narrative reviews. Every individual study necessarily has limitations, so inferences about the superiority of a particular treatment from one study's findings are particularly tenuous (e.g., the superiority of a treatment in one trial can be due to Type I error). Conclusions made from narrative reviews often give primacy to findings in which one particular treatment is superior to another and tend to ignore findings of no differences, a problem that is exacerbated by focusing on one or several outcome variables among many that are found to favor one treatment (Quintana & Minami, 2006). Therefore, conclusions of treatment differences noted in narrative reviews may be spurious. Nevertheless, it is entirely possible that some treatments for PTSD are indeed superior to others. Meta-analysis is able to aggregate studies in such a way as to make quantitative conclusions based on the systematic aggregation of studies and thus produce conclusions based on a preponderance of the evidence rather than on heuristic interpretations.

A number of meta-analyses investigating treatment of PTSD have been conducted using a variety of methods. One method of examining relative efficacy is to compare psychotherapy to no-treatment controls or supportive controls. Using this strategy, the researcher calculates an effect size after locating multiple studies using the same psychotherapy (such as cognitive behavioral therapy) vis-à-vis a control group. The individual studies' effect sizes for each psychotherapy are aggregated (e.g., a mean effect size for exposure is calculated by aggregating studies that compare an exposure condition to a control group). Finally, the mean effect size estimates for each psychotherapy are compared. Using this method, Van Etten and Taylor (1998) reported similar effect sizes for behavioral ($d = 1.12$) and EMDR ($d = .95$) treatments. Bradley, Greene, Russ, Dutra, and Westen (2005) found similar effect sizes for exposure therapy, cognitive behavioral therapy (CBT), exposure plus cognitive therapy, and EMDR, but not for the supportive counseling condition, which was comprised of control treatments not intended to be therapeutic ($d = -0.01$). Similarly, Bisson and Andrew (2005) also reported that trauma focused cognitive behavioral therapies (TFCBT) ($d = 1.36$) and stress management ($d = 1.14$) produced larger effects vis-à-vis control groups than did supportive counseling, which did not differ from wait list controls. In their recent analysis, Bisson et al. (2007) noted that many psychotherapies (viz., TFCBT, EMDR, stress management, and group CBT) were consistently superior to wait list or usual care controls, except for "other therapies" which did not differ from controls. Bisson et al. concluded that "other therapies", which consisted of an assortment of treatments including hypnotherapy, psychodynamic therapies, supportive therapies, and

non-directive counseling, were the least efficacious of the treatment categories and that EMDR and TFCBT were the most efficacious. Importantly, Bisson et al. (2007) urged caution interpreting the conclusion because of the "considerable unexplained heterogeneity observed in these comparisons" (p.13).

Estimation of relative efficacy by comparing treatments vis-à-vis controls in meta-analysis introduces significant threats to validity. As noted by Shadish and Sweeney (1991), studies vary with regard to participants, outcome measures employed, treatment standardization, measure reactivity, blinding procedures, treatment length, severity of disorder, and multiple unmeasured variables. Consequently, any differences between categories of treatments could be attributed to differences among the categories relative to these variables, creating confounds. For example, one class of treatments may employ more reactive measures generally, so that an apparent superiority actually reflects the ease with which variables change. One solution to this problem is to meta-analytically examine potential mediators and moderators of effect size (i.e., statistically control for confounds if differences are found), a strategy employed at the origins of meta-analysis by Smith and Glass (1977). Unfortunately, this strategy introduces a number of additional problems, including omitting important variables, misspecification of models, power, and collinearity of possible confounds. An alternative strategy that eliminates most of the confounds is to include only studies that contain direct comparisons between psychotherapies, as the purported confounding variables are held constant in each study (Shadish and Sweeney, 1991; Wampold et al., 1997). For example, reactivity of measurement is no longer a confound in a direct comparison study because both treatments are assessed with the same measures.

Several PTSD meta-analyses have used the direct comparison method, producing some contradictory findings. The first meta-analysis to include direct comparisons between psychotherapy treatments of PTSD revealed no evidence of differences between EMDR and other exposure-based treatments (Davidson & Parker, 2001), but the number of comparisons was few. Similarly, Bradley et al. (2005), in a thorough examination of PTSD outcomes, found no differences in direct comparison studies but noted that the number of studies was insufficient to draw definitive conclusions. Siedler and Wagner (2006) included direct comparison studies of EMDR and CBT for treatment of PTSD and concluded that neither was superior to the other. Bisson and Andrew (2005) conducted four meta-analyses of direct comparisons between treatment categories, finding TFCBT and stress management equal to each other and superior to "other therapies", as well as no significant differences between group TFCBT and group non-TFCBT. Bisson et al. (2007) analyzed seven meta-analytic direct comparisons between treatment categories, finding TFCBT and EMDR superior to stress management, and all three aforementioned therapies superior to "other therapies". In these meta-analyses, variants of CBT and EMDR appeared to be more efficacious than a category of treatments that contains psychodynamic and hypnotherapy, among others.

The conclusion that "other therapies" are inferior may be a classification artifact. Many of the treatments in the meta-analyses of direct comparisons were not intended to be therapeutic (i.e., the treatments were not bona fide psychotherapies). One of the therapies frequently included in PTSD trials is labeled *supportive therapy*. Typically supportive therapies are intended to evaluate the specific ingredients of psychological treatment by providing attention of empathic therapists without specific ingredients. Thus, these controls often prohibit therapists from using specific therapeutic techniques, contain no theoretically cogent rationale, are not tailored to individual patient needs, are delivered by therapists who are cognizant that the condition is not intended to be therapeutic, and would not be offered to the public as viable treatments by therapists in practice (Baskin, Tierney, Minami, & Wampold, 2003; Rosenthal & Frank, 1956; Wampold, 2001; Westen, Novotny, & Thompson-Brenner, 2004). An example in a PTSD psychotherapy study is a supportive therapy offered to female survivors of rape that consisted of the following: (a) teaching of a general problem solving technique (b) the therapist responding indirectly and with unconditional support, and (c) immediate refocus by the therapist to everyday problems when the patient tried to talk about the sexual assault (Foa, Rothbaum, Riggs, & Murdock, 1991). Inclusion of supportive counseling in an "other therapies" category might explain some of the conflicting findings and heterogeneity of previous meta-analyses. The Bisson and Andrew (2005), and Bisson et al. (2007) meta-analyses incorporated these supportive therapies into a category of "other therapies", which included a mix of potentially bona fide (e.g., psychodynamic therapy) and non-bona fide therapies (e.g., supportive therapy and psychoeducation). Moreover, Bisson et al. (2007) noted that there existed an unexplained heterogeneity among the effects, particularly for "other therapies," suggesting that there were unexplained confounds.

A second limitation of previous direct comparison meta-analyses for treatment of PTSD has been that categorizing treatments into classes necessitates a large number of pair-wise (i.e., one category versus another, creating $k(k-1)/2$ comparisons for k classes) statistical tests, which creates inflated Type I error rates (see Wampold, 2001). An example is Bisson et al.'s (2007) meta-analysis, which employed 37 statistical tests, approximately six per comparison:

(a) EMDR versus TFCBT, (b) TFCBT versus stress management, (c) TFCBT versus other therapies, (d) EMDR versus stress management, (e) EMDR versus other therapies, (f) stress management versus others therapies, and finally (g) group TFCBT versus group CBT. Of the 15 statistical tests that revealed some clinically important difference between treatments, four indicated the superiority of EMDR and TFCBT over stress management and 11 indicated the superiority of EMDR, TFCBT, and stress management over "other therapies". The four tests indicating inferiority of stress management may likely be spurious given that 37 statistical tests were conducted. These problems are exacerbated by the few studies that compare treatments in two categories (e.g., the mean number of studies used to compare categories was 3.1). Categorization utilizes pair-wise comparisons in lieu of an omnibus test of the hypotheses that outcomes differ between individual psychotherapies and also ignores comparisons within classes of treatments (Shadish & Sweeney, 1991; Wampold et al., 1997). In short, some meta-analyses have attempted to identify a superior class of treatments but have not tested the omnibus hypothesis that all treatments intended to be therapeutic are equally efficacious.

In sum, meta-analyses have consistently found that psychotherapies are more efficacious than waiting list controls. Some have found that one class of treatments is superior to another and all seem to indicate that "other therapies," which includes some treatments that were intended to be therapeutic and some not, are inferior to the more specific classes. Rejection of a null hypothesis that all treatments are equally efficacious for the treatment of PTSD would suggest that the search for one or more particularly efficacious treatments would be profitable. On the other hand, failure to reject the null hypothesis that all treatments intended to be therapeutic are equally efficacious would suggest that further attempts to reveal treatment differences would not be particularly informative. Thus, the omnibus test in important: Are all psychotherapies for the treatment of PTSD that are intended to be therapeutic equally efficacious?

1. Purpose and hypothesis of the present meta-analysis

The purpose of the present meta-analysis was to provide a test of the relative efficacy of psychotherapies for PTSD. The current analysis was modeled after the analysis of Wampold et al. (1997) and differs from previous meta-analyses of psychotherapy for PTSD in that we: (a) only included those studies directly comparing two or more psychotherapies for the treatment of PTSD, (b) did not classify treatments into categories, and (c) included only those treatments that were intended to be therapeutic. The null hypothesis is that the true effect size of all PTSD treatment comparisons does not vary significantly from zero, both when using only measures of PTSD symptoms and when using all measures of psychological functioning from each study. If the omnibus null of no differences is rejected, then additional analyses will be employed to explain the effects and identify a treatment that is superior to others or to identify the specific characteristics of the several treatments that are more efficacious than others.

2. Method

2.1. Study selection

To identify head-to-head comparisons of bona fide psychotherapies, we performed a literature search including major databases: Medline, Cinahl Health, Healthsource, PsycInfo, PsychArticles, PubMed, Social Sciences Fulltext, and Web of Knowledge using the keywords "PTSD", "posttraumatic", "post-traumatic", "post traumatic", "psychotherapy" and "treatment". In addition, we searched reference sections of previous meta-analyses of PTSD treatment, controlled studies of psychotherapy outcomes for PTSD, and literature reviews of PTSD treatment.

To be included in the meta-analysis, the study must have been an experimental design in which multiple adult patients were randomly assigned to one of two or more bona fide psychotherapies for treatment of PTSD. Additionally, patients must have met diagnostic criteria for PTSD using DSM-III or DSM-IV criteria. Studies also must have included assessments of PTSD symptoms, contained sufficient statistics to compute effect sizes, and treatments delivering two or more psychotherapy sessions. Prior experience has demonstrated that contacting authors for additional statistics has not proved viable.

To determine if a treatment condition was a bona fide psychotherapy treatment, we used the criteria developed by Wampold et al. (1997). First, the treatment must have been delivered by a trained therapist and include an interaction in which the patient developed a relationship with the therapist and the treatment was tailored to the patient. Consequently, the treatment could not be provided via tape recording or protocols that were not contingent on patient response (e.g., progressive muscle relaxation that was not modifiable). Second, the treatment had to satisfy two of the following four criteria: (a) a citation to an established psychological approach was made (e.g., prolonged exposure), (b) a description of the therapy was provided and based on psychological principles (e.g., extinction), (c) a manual of the treatment was available and used to guide treatment, and (d) active ingredients of treatments

were named and citations for these ingredients were provided in the article. Consequently, any treatments that were designed to control for common factors, such as a supportive counseling, were excluded. Moreover, component, dismantling, or parametric studies that varied the presence or amount of one particular technique, psychotherapy, or intervention were excluded.

2.2. Coding procedure

The primary author first retrieved all studies containing at least two psychotherapy treatments for PTSD. Two volunteer counseling psychology graduate students who were not otherwise involved in the study independently reviewed these studies. The reviewers were trained to rate the bona fide status of psychotherapies using the Wampold criteria noted above and were blind to the results of the study. If both reviewers determined that the study included two bona fide treatments for PTSD, the study was included in the analysis. In case of reviewer disagreement, an advanced doctoral student in counseling psychology (the second author) evaluated the study. If this third rating resulted in the decision that the study included at least two bona fide treatments, the study was retained; if not, the study was excluded. Of the 29 effects included initially for coding, 3 effects were duplicates (or re-analysis of the same data from another included study) and were thus excluded (e.g., Stapleton, Taylor, & Asmundson, 2006). When the remaining 26 effects were then coded for bona fide status, 9 effects from 5 studies were rejected, leaving an inclusion total of 17 effects from 14 studies. The raters disagreed on 5 effects from 3 studies, which after consultation with the third coder, resulted in exclusion of 3 effects (Vaughan et al., 1994) and inclusion of 2 effects (Devilly & Spence, 1999; Tarrier et al., 1999). The studies that met the criteria for inclusion are presented in Table 1.

2.3. Analytic strategy

As noted by Wampold et al. (1997), avoiding categorization of treatments results in the sign (+/−) of any particular effect size becoming arbitrary, as the effect size for any comparison of treatments is formed by the difference of the means (i.e., Tx A − Tx B or Tx B − Tx A). Consequently, the primary strategy is to, randomly assign positive and negative signs to each effect (i.e., half of the effects were thus labeled as negative, half as positive). The assignment of signs was accomplished with the EXCEL random number generator. As this strategy necessarily results in an aggregate effect of approximately zero (i.e., half the effects are positive and half are negative), the appropriate test of treatment differences is the test of homogeneity of effects sizes about zero, a test commonly used in meta-analyses (Cramer, 1946; Hedges & Olkin, 1985). If there are no true differences among treatments (i.e., the null hypothesis that all treatments are equally efficacious is true), then the obtained effect sizes will be clustered close to zero with deviations from zero due to sampling error. The homogeneity statistic gauges the degree to which the effects deviate from zero; the null hypothesis of homogeneity is rejected when the effects deviate from zero to a greater extent than would be expected by chance.

A secondary strategy involves aggregating the absolute value of each effect size. Under the null hypothesis of no differences, deviations from zero will be positive (i.e., the absolute value of the effect) and the aggregate will thus necessarily be greater than zero. Consequently, the aggregation of the absolute values of the effect sizes provides an upper bound of the true differences among treatments.

Means and standard deviations of post-treatment symptoms were used to calculate effect sizes. Like Bradley et al. (2005), we were unable to calculate long-term follow-up data because too few studies reported this data. In addition, follow-up assessments may attenuate differences between psychotherapies due to confounds introduced in the typically naturalistic period following treatment (Crits-Christoph, 1997). We calculated two effect sizes for each comparison, PTSD symptoms only and all post-treatment measures of psychological functioning. Typically, PTSD symptoms in the primary studies were assessed with the Impact of Events Scale (46%), Clinician Administered PTSD Scale (46%), State-Trait Anxiety Inventory (46%), or the PTSD Symptom Scale (40%). Given the exceptionally high rates of comorbidity in PTSD (Brown, Campbell, Lehman, Grisham, & Mancil, 2001) the analysis of all outcome measures provided important information regarding aspects of psychological functioning considered by the studies' authors to be important in patients with PTSD. In studies where subscales and full-scale scores were reported, the full-scale measure scores were used. The effect size for dependent variables was computed with the assumption that the correlation of assessment variables was 0.50, an assumption justified by Wampold et al. (1997) and which reduces the standard error of estimate, thus resulting in a more accurate *d* estimate. In cases where only subscales of a full-scale measure were reported, subscale scores were aggregated using published subscale correlation(s) (e.g., 0.67 to 0.70 for CAPS subscales; Foa & Tolin, 2000). Finally, to avoid additional confounds due to dropout rates, intent-to-treat final scores were used when available (80% of studies reported intent-to-treat scores).

2.4. Statistical analysis

The unit of analysis was a direct comparison of post-treatment symptoms of two bona fide psychotherapies. We calculated the effect size, Cohen's *d*, and a variance for each outcome measure within each comparison using standard meta-analytic procedures (Hedges & Olkin, 1985), and then aggregated across all outcome variables within each treatment (see Wampold et al., 1997). In studies where three bona fide psychotherapies were compared, separate effect sizes were computed, A to B, A to C, B to C.

Table 1
Direct comparison studies of bona fide psychotherapies for treatment of post-traumatic stress disorder

Lead author	*N*	Psychotherapies	*d*		Author quote on intent-to-treat post-test relative efficacy for (1) PTSD measures (2) all measures
			PTSD	All outcomes	
Brom, Kleber, and Defares (1989)	60	Trauma desensitization Hypnotherapy	0.30	0.07	(1) "the differences between the therapies are small" (p.610) (2) "Similarity of…treatment(s)…based on quite diverging theoretical considerations" (p.610)
Brom et al. (1989)	60	Trauma desensitization Psychodynamic	0.14	0.09	(1) "the differences between the therapies are small" (p.610) (2) "Similarity of…treatment(s)…based on quite diverging theoretical considerations" (p.610)
Brom et al. (1989)	60	Hypnotherapy Psychodynamic	0.18	0.19	(1) "the differences between the therapies are small" (p.610) (2) "Similarity of…treatment(s)…based on quite diverging theoretical considerations" (p.610)
Devilly and Spence (1999)	23	Trauma treatment EMDR	0.34	0.09	(1) "While there was no effect for Condition,…there was an effect for…Time and Condition," (p.143) (2) "TTP proved to be more effective although to a lesser extent than with the PTSD symptomatology" (p.153)
Foa, Rothbaum, Riggs, and Murdock (1991)	31	Stress inoculation Exposure	0.43	0.33	(1) "SIT produced superior immediate symptom reduction" (p.722) (2) "On other measures of psychopathology, no significant group differences emerged" (p.721)
Foa, Dancu, Hembree, Jaycox, Meadows, & Street (1999)	51	Exposure Stress inoculation	0.14	0.16	(1) "active treatments…did not differ significantly from each other" (p.194) (2) "active treatments…did not differ significantly from each other" (p.194)
Ironson, Freund, Strauss, and Williams (2002)	25	EMDR Prolonged exposure	0.12	0.49	(1) "our results suggest that both EMDR and PE are equally effective at reducing symptoms of PTSD" (p.123) (2) "EMDR and PE are equally effective at reducing symptoms of…depression" (p.123)
Lee, Gavriel, Drummond, Richards, and Greenwald (2002)	24	EMDR Stress inoculation	0.52	0.51	(1) "On global PTSD measures, there were no significant differences between treatments" (p.1084) (2) "There were no significant difference between groups" (p.1079)
Marks et al. (1998)	42	Exposure Cognitive	0.06	0.14	(1) "E (exposure) did no better than C (cognitive)" (p.323) (2) "almost complete overlap of the confidence intervals" (p.320)
McDonagh et al. (2005)	51	CBT Present centered	0.23	0.19	(1) "treatments did not differ significantly at any assessment time point on any measure" (p.519) (2) "no significant differences" (p.519)
Paunovic and Ost (2001)	20	Exposure CBT	0.20	0.14	(1) "There was no difference between E and CBT on any measure." (p.1183) (2) "No significant difference between them on any of the 14 measures." (p.1193)
Power et al. (2002)	76	EMDR Exposure+ cognitive	0.52	0.51	(1) "no significant difference between experimental groups" (p.314) (2) "No difference between EMDR and E+CR" (p.309)
Resick, Nishith, Weaver, Astin, and Feuer (2002)	90	Cognitive processing Prolonged exposure	0.23	0.10	(1) "the treatment groups…did not differ from each other" (p.872) (2) "CPT was superior to PE in remediating guilt cognitions" (p.877)
Rothbaum, Astin, and Marsteller (2005)	48	Prolonged exposure EMDR	0.31	0.38	(1) "PE and EMDR did not differ significantly…for any quantitative scale" (p. 607) (2) "did not differ" (p.613)

(continued on next page)

Table 1 (*continued*)

Lead author	*N*	Psychotherapies	*d*		Author quote on intent-to-treat post-test relative efficacy for (1) PTSD measures (2) all measures
			PTSD	All outcomes	
Schnurr et al. (2003)	324	Trauma focused group Present centered group	0.05	0.06	(1) "no overall differences between therapy groups on any outcome." (p.481) (2) "no overall differences between therapy groups on any outcome." (p.481)
Tarrier et al. (1999)	62	Imaginal exposure Cognitive therapy	0.20	0.12	(1) "There were no significant differences between the 2 treatments at any assessment." (p.13) (2) "There were no differences between treatment groups." (p.16)
Taylor, Thordarson, Maxfield, Fedoroff, Lovell, and Ogrodniczuk (2003)	41	EMDR Exposure	0.11	0.31	(1) " the three treatments did not differ on any of the outcome measures." (p.336) (2) "These results indicate that the treatments did not significantly differ..." (p.335)

Table 2
Effect size estimates and tests of homogeneity for unconditional methods

PTSD measures/random signs

Fixed effect	*Coefficient*	*Standard error*	*t ratio*	*P*	
Grand mean λ_o	−0.01	0.06	−0.18	0.859	
Random effect	*Variance component*	*df*	X^2	*P*	I^2
True *E.S.* δ_j	0.00	16	15.18	>0.500	0.00

All measures/random signs

Fixed effect	*Coefficient*	*Standard error*	*t ratio*	*P*	
Intercept λ_o	−0.02	0.05	−0.29	0.777	
Random effect	*Variance component*	*df*	X^2	*P*	I^2
True *E.S.* δ_j	0.00	16	17.66	0.344	9.39

PTSD measures/upper bound

Fixed effect	*Coefficient*	*Standard error*	*t ratio*	*P*	
Intercept λ_o	0.19	0.06	3.25	0.005	
Random effect	*Variance component*	*df*	X^2	*P*	I^2
True *E.S.* δ_j	0.00	16	7.74	>.500	0.00

All measures/upper bound

Fixed effect	*Coefficient*	*Standard error*	*t ratio*	*P*	
Intercept λ_o	0.16	0.05	3.20	0.006	
Random effect	*Variance component*	*df*	X^2	*P*	I^2
True *E.S.* δ_j	0.00	16	8.10	>0.500	0.00

Note: Fixed effects, Random effects and significance tests for four statistical models. In the first two models, we randomly assigned positive and negative signs to effect sizes calculated for PTSD measures alone and all measures. The random assignment results in the aggregate effect (λ_o, or the intercept) necessarily close to zero. Consequently, the estimate of treatment differences is provided by variance component, or the variability in the true effect sizes, δ_j. In the final two models, we calculated the absolute value of effect sizes obtained from PTSD measures alone and all measures. In these models, the intercept, λ_o, provides an upper bound of treatment differences. The I^2 statistic provided an estimate of the extent of heterogeneity in each model.

We considered studies to be sampled from a larger population of studies, consistent with a random effects model for our analysis (Hedges & Olkin, 1985). The analysis was performed using a multilevel model where variances are known (Raudenbush & Bryk, 2002, Chapter 7), using HLM6 (Bryk, Raudenbush, & Congdon, 1996). The models tested were unconditional models (i.e., not conditioned on study level variable, patient characteristics, allegiance of researchers e.g.). At level 1,

$$d_j = \delta_j + e_j$$

where d_j provides the estimate of the differences between treatments for study j, δ_j is the true effect for study j, and the variance of the errors e_j is known. At level 2,

$$\delta_j = \gamma_o + u_j$$

where γ_o is the grand mean of the effects (i.e., the aggregate effect) and u_j is the level 2 error. Note that when random signs are assigned to effect sizes, the grand mean γ_o will be close to zero. If there are no true differences between treatments, var(u_j) will be due entirely to sampling error (i.e., the effects will be distributed as expected under the null hypothesis of no difference and the variance will not be due to true differences among treatments); that is, the effects will be homogenously distributed around zero. Homogeneity is tested with the statistic H that provides an estimate of extent to which sampled effects deviate from the grand mean, weighted by the inverse of the variance (Hedges & Olkin, 1985; Raudenbush & Bryk, 2002). H has an approximate chi-square distribution with $k-1$ degrees of freedom, where k is the number of studies aggregated.

To gauge the extent of heterogeneity, we also calculated an I^2 index (Huedo-Medina, Sanchez-Meca, Marin-Martinez, & Botella, 2006). The I^2 index, which is calculated by subtracting the degrees of freedom ($k-1$) from the H statistic, dividing by the H statistic, and then multiplying by 100, is interpreted as the proportion of variability in effect sizes due to true differences among treatments. Higgins and Thompson (2002) proposed that percentages of 25%, 50%, and 75% be interpreted as low, medium, and high heterogeneity, respectively. For example, an I^2 value of 50 indicates that half of the aggregate variability among effect sizes between studies is caused by true heterogeneity and not sampling error (Huedo-Medina et al., 2006).

3. Results

Fifteen studies with 17 comparisons between treatments met inclusion criteria (from a pool of 22 studies with 26 comparisons). A total of 958 participants were included from these studies, with the per-study mean and median numbers of participants at 56 and 42, respectively. The results for these analyses are presented in Table 2.

As discussed above, the primary test of the null hypothesis that there are no differences among treatments involves using the allocation of random signs to the effects and examining the extent to which effects are heterogeneously distributed about the grand mean of zero. With respect to the first hypothesis (for PTSD measures only), and the second hypothesis (for all measures included in each primary study), the grand mean was near zero, as determined by randomly assigning the sign of the effect. For the PTSD symptoms, the estimate of the variance of the true effect sizes was 0.00; the H statistic was 15.18 which, when compared with a chi-square distribution with 16 degrees of freedom, was not sufficiently large to reject the null hypothesis that the effects were homogenously distributed around zero ($p>0.500$). For all symptom measures, the estimate of the variance of the true effect sizes was 0.00; the H statistic was 17.66 which, when compared with a chi-square distribution with 16 degrees of freedom, was not sufficiently large to reject the null hypothesis that the effects were homogenously distributed around zero ($p=0.344$). In all cases the I^2 index was very small, zero for PTSD measures, and 9.39% for all measures. These analyses indicate that the effects were homogenously distributed about zero; that is, deviations from zero were as expected under the null hypothesis of no differences between treatments. Any effects in particular studies were thus likely due to chance and not to true differences among treatments.

In the models in which the absolute values of effect sizes were aggregated to provide an upper bound of the differences between treatments, the aggregated effect sizes were quite small ($d=0.19$ for PTSD measures and 0.16 for all measures) and were homogeneously distributed (see Table 2). It should be noted that this upper bound is necessarily an overestimation of the differences between treatments and should not be misunderstood as the true difference. A d of 0.16 translates into a needed number to treat (NNT) of 12, which means that even if 0.16 were the difference between psychotherapies and not the upper bound, 12 patients would need to be treated with the superior treatment in order to have one more success as compared to the less effective treatment (Kraemer & Kupfer, 2006). This NNT is quite large, particularly in light of the fact that attrition rates are a problem (e.g., approximately 20% of patients in exposure therapy drop out of treatment (Hembree et al., 2003).

As all effects were homogeneous, we did not model moderators or mediators of the effect sizes.

4. Discussion

The purpose of the present meta-analysis was to test the relative efficacy of bona fide psychotherapies for adults with post-traumatic stress disorder. Our primary analysis revealed that effect sizes were homogenously distributed around zero. The analysis provided no evidence to suggest outcome differences between bona fide psychotherapies in the treatment of PTSD, for either PTSD symptoms or all outcome measures. Moreover, the upper bound of the true effect size between

PTSD psychotherapies, necessarily an over-estimate of treatment differences, was quite small, d=0.16–0.19, and marginally smaller than found for adult treatments in general (cf, Wampold et al., 1997).

In contrast to previous meta-analyses of PTSD treatment, we (a) included only bona fide psychotherapies, (b) incorporated only studies with direct comparison of two or more psychotherapies, and (c) avoided categorization of treatments, yielding a focused test of hypothesis that all treatments are equally efficacious. Comparisons of bona fide psychotherapies yielded effect sizes homogenously distributed about zero, thus it appears that the "considerable unexplainable heterogeneity" (Bisson et al., 2007, p. 13) in differences between treatments is likely accounted for by these different methodological decisions. An important aspect in this meta-analysis is that the "other therapies" were disaggregated and those that were intended to be therapeutic were included in the present meta-analysis and those that were not intended to be therapeutic were excluded. The conclusion that "other therapies" are inferior (Bisson et al., 2007; Bradley et al., 2005) is not supported in the present meta-analysis.

Our findings are consistent with previous PTSD treatment meta-analyses that found little evidence of differences between treatments (e.g., Bradley et al., 2005; Siedler & Wagner, 2006). The paucity of differences is illustrated by the conclusions reached in many of the primary studies (see Table 1), where the modal statement was most often a variation of the following: "Both psychotherapy treatments were superior to the control group, but neither treatment was superior to the other."

The present study also addresses a primary criticism of earlier psychotherapy meta-analyses on relative efficacy. Specifically, Wampold et al.'s (1997) meta-analysis, on which the current study is modeled, found no evidence of differences between treatments but included studies heterogeneous in terms of patient population, disorders and patient severity. Critics hypothesized that Wampold's finding of no differences between treatments could mask important differences in some specific problem or disorder (Crits-Christoph, 1997; Howard, Krause, Saunders, & Kopta, 1997). Pointedly, DeRubeis, Brotman, and Gibbons (2005) stated that examining relative efficacy blind to the type of disorder "is akin to asking whether insulin or an antibiotic is better, without knowing the condition for which these treatments are to be given.... Alternatively, researchers should begin with a problem and ask how treatments compare in their effectiveness for that problem" (p. 175). The present study addressed this limitation by estimating treatment differences in a clinical sample with the same disorder, specifically PTSD.

Had there been evidence for differences among treatments for PTSD, further analyses could have been conducted to determine the source of the heterogeneity of effect sizes. In the presence of such heterogeneity, these analyses could have illuminated the specific ingredients that were present in the more efficacious treatments. Lack of treatment differences suggests that specific ingredients may not be critical for the treatment of PTSD, a conclusion that is supported by dismantling studies that have failed to identify the validity of specific ingredients. Adding cognitive restructuring to exposure does not augment cognitive changes over exposure therapy alone (Bryant, Moulds, Guthrie, Dang, & Nixon, 2003; Foa et al., 2005; Foa & Rauch, 2004; Marks, Lovell, Noshirvani, Livanou, & Thrasher, 1998). Adding exposure to stress inoculation does not improve outcomes of PTSD patients receiving stress inoculation therapy alone (Foa et al., 1999) or family therapy alone (Glynn et al., 1999). EMDR outcomes are not enhanced by eye movements (Cahill, Carrigan, & Frueh, 1999; Devilly & Spence, 1999; Pitman Altman, Greenwald, Longpre, Macklin, & Poire, 1991; Renfrey & Spates, 1994) or its cognitive components (Devilly & Spence, 1999). The results of the present study and these component studies then suggest the possibility that factors common to all treatments may be responsible for the benefits of these treatments, supporting a common factor model (Wampold, 2001), although the current design does not provide evidence relative to any particular common factor as curative.

4.1. Limitations

Despite inclusion of every direct comparison psychotherapy study of PTSD with two bona fide psychotherapies, the primary limitation of this meta-analysis is the relatively low number of studies included. The number of studies and effects available for the analysis was not large (15 studies, 17 effects), and provides a less than ideal test of heterogeneity. Type II errors are threats to meta-analytic tests of heterogeneity. Specifically, Higgins and Thompson (2002) note that approximately 20 effects are considered appropriate for an adequate test of heterogeneity. However, it should be noted that the estimates of heterogeneity did not approach significance and I^2 values were quite low, suggesting that the failure to find treatment difference was not the result of an underpowered analysis of heterogeneity. In comparison to the present study's inclusion of 17 direct comparisons with 958 participants, previous PTSD psychotherapy meta-analyses included the following in their largest direct comparison within the meta-analysis: Bisson and Andrew (6 direct comparisons, 239 participants), Bisson et al. (7 comparisons, 267

participants), Bradley et al. (13 comparisons, 454 participants), and Sielder and Wagner (7 comparisons, 209 participants). To put this in perspective, the present study included more direct comparisons than any other previous meta-analysis of PTSD, and two-to-four times as many participants. Moreover, it should be recognized that the null hypothesis in the present case is that there are no differences among treatments intended to be therapeutic; if the corpus of studies is insufficient to test this hypothesis, then the null hypothesis should be retained until such time that more studies accumulate.

An assumption of any random effects multilevel model is that the units are drawn from a population of units, as is the case, say, when we assume participants in any study are drawn form a population of participants. Rarely is such a random sampling from a population actually conducted. In the present case, it was assumed that the studies in this meta-analysis were drawn from a population of studies and this extends logically to treatments. Hierarchical linear modeling conclusions are made about treatments in general and not the particular treatments tested in these studies in a random effects model as recommended (see Duncan, Duncan, & Strycker, 2006; Hox, 2002; Kreft & DeLeeuw, 1998; Snijders & Bosker, 1999). However, care must be taken to generalize in such models only to units similar to the units in the sample (see Serlin, Wampold, & Levin, 2003 for a discussion of this in another context). Clearly, the treatments compared in these studies were not randomly selected from all possible treatments but rather were constrained by researchers' allegiance, funding opportunities, feasibility, and so forth. Indeed, it is reasonable to believe that more recent studies tend to compare treatments that might be characterized as having prior evidence bases. Thus, it is inappropriate to conclude that *all* psychotherapy treatments of PTSD intended to be therapeutic are equally efficacious; more appropriately, the conclusion about uniform efficacy applies to treatments similar to the ones used in the primary studies. Nevertheless, the conclusion can be rather general because the corpus of treatments in this meta-analysis were remarkably diverse theoretically and included stress management, psychodynamic treatments, EMDR, hypnotherapy, cognitive behavioral treatment, exposure-based treatment, and treatments designed explicitly to exclude exposure (e.g., present centered therapy; McDonagh et al., 2005).

Only published studies were included in the present analysis. The 'file drawer' issue can be a problem in meta-analysis (Lipsey & Wilson, 1993), yet is likely to be far more problematic in narrative literature reviews (Quintana & Minami, 2006). Additionally, it is likely that any publication bias would bias the analysis towards finding differences between treatments. Specifically, unpublished studies seem more likely to report no differences between treatments as compared to published studies (Rotton, Foos, Van Meek, & Levitt, 1995). Inclusion of unpublished studies, therefore, would likely not lead to a different conclusion.

4.2. Clinical implications

The present findings have important clinical implications. Because the number of PTSD cases is rising, access to effective treatments is increasingly critical. Availability of treatment seems especially important given the debilitating effects of PTSD and stigma attached to seeking mental health services, especially for those serving in the military (see Hoge et al., 2004). Although prolonged exposure has been selected by the Substance Abuse and Mental Health Services Administration (SAMHSA) as a model treatment for dissemination to clinics nationwide (Nemeroff et al., 2006), the meta-analytic findings do not suggest that any particular therapy is superior to another.

Given the lack of differential efficacy between treatments, it seems scientifically questionable to recommend one particular treatment over others that appear to be of comparable effectiveness. Alternatively, patients may be better served by increasing access to as many bona fide psychotherapy treatments as possible. Although research on patient by treatment interactions is lacking, patients do present to therapy with varying motives, symptom clusters, abilities, worldviews, cultural backgrounds, and life experiences. It is plausible that these factors have an influence on patient preference and tolerability of treatment and may be helpful in guiding treatment selection. Specifically, there is some evidence to suggest that patients whose explanations for their distress are congruent with the treatment rationale may benefit from treatment more than patients whose explanations do not (Wampold, 2001).

Consider dropout rates and our NNT analysis. Although the upper bound found in this study for treatment differences was approximately 0.16, let us imagine the effect if the true difference between treatments was 0.16. In this scenario, approximately eight out of 100 patients might do better in a superior treatment versus another. Due to the fact that about one-quarter of patients drop out of psychotherapy treatment for PTSD (Hembree et al., 2003) and dropout appears to be related to symptom change (Bradley et al., 2005), keeping patients in treatment would appear to be more important in achieving desired outcomes than would prescribing a particular type of psychotherapy. Having several psychotherapies to

choose from may enable a better match of patients to type of psychotherapy that fits the patient's worldview and is more tolerable to that particular patient.

References[1]

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders*, 4th edition Washington, DC: American Psychiatric Association.

Baskin, T. W., Tierney, S. C., Minami, T., & Wampold, B. E. (2003). Establishing specificity in psychotherapy: A meta-analysis of structural equivalence of placebo controls. *Journal of Consulting and Clinical Psychology*, *71*, 973−979.

Bisson, J., & Andrew, M. (2005). Psychological treatment of post-traumatic stress disorder (PTSD). *Cochrane Database of Systematic Reviews*, *18*.

Bisson, J. I., Ehlers, A., Matthews, R., Pilling, S., Richards, D., & Turner, S. (2007). Psychological treatments for chronic post-traumatic stress disorder. Systematic review and meta-analysis. *British Journal of Psychiatry*, *190*, 97−104.

Bradley, R., Greene, J., Russ, E., Dutra, L., & Westen, D. (2005). A multidimensional meta-analysis of psychotherapy for PTSD. *American Journal of Psychiatry*, *162*, 214−227.

*Brom, D., Kleber, R. J., & Defares, P. B. (1989). Brief psychotherapy for posttraumatic stress disorders. *Journal of Consulting and Clinical Psychology*, *57*, 607−612.

Brown, T. A., Campbell, L. A., Lehman, C. L., Grisham, J. R., & Mancil, R. B. (2001). Current and lifetime comorbidity of the DSM-IV anxiety and mood disorders in a large clinical sample. *Journal of Abnormal Psychology*, *110*, 585−599.

Bryant, R. A., Moulds, M. L., Guthrie, R. M., Dang, S. T., & Nixon, R. D. V. (2003). Imaginal exposure alone and imaginal exposure with cognitive restructuring in treatment of posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology*, *71*, 706−712.

Bryk, A. S., Raudenbush, S. W., & Congdon, R. T. (1996). HLM: Hierarchical linear and nonlinear modeling with the HLM/2L and HLM/3L programs Chicago: Scientific Software International.

Cahill, S. P., Carrigan, M. H., & Frueh, B. C. (1999). Does EMDR work? And if so, why? A critical review of controlled outcome and dismantling research. *Journal of Anxiety Disorders*, *13*, 5−33.

Cramer, H. (1946). Mathematical methods of statistics Princeton, NJ: Princeton University.

Crits-Christoph, P. (1997). Limitations of the dodo bird verdict and the role of clinical trials in psychotherapy research: Comment on Wampold et al. (1997). *Psychological Bulletin*, *122*, 216−220.

Davidson, P. R., & Parker, K. C. (2001). Eye movement desensitization and reprocessing (EMDR): A meta-analysis. *Journal of consulting and clinical psychology*, *69*, 305−316.

DeRubeis, R. J., Brotman, M. A., & Gibbons, C. J. (2005). A conceptual and methodological analysis of the nonspecifics argument. *Clinical Psychology: Science and Practice*, *12*, 174−183.

*Devilly, G. J., & Spence, S. H. (1999). The relative efficacy and treatment distress of EMDR and a cognitive–behavior trauma treatment protocol in the amelioration of posttraumatic stress disorder. *Journal of Anxiety Disorders*, *13*, 131−157.

Duncan, T. E., Duncan, S. C., & Strycker, L. A. (2006). An introduction to latent variable growth curve modeling Mahwah, NJ: Lawrence Erlbaum.

Engdahl, B., Dikel, T. N., Eberly, R., & Blank, A. (1998). Comorbidity and course of psychiatric disorders in a community sample of former prisoners of war. *American Journal of Psychiatry*, *155*, 1740−1745.

*Foa, E. B., Dancu, C. V., Hembree, E. A., Jaycox, L. H., Meadows, E. A., & Street, G. P. (1999). A comparison of exposure therapy, stress inoculation training, and their combination for reducing posttraumatic stress disorder in female assault victims. *Journal of Consulting and Clinical Psychology*, *67*, 194−200.

Foa, E. B., Hembree, E. A., Cahill, S. P., Rauch, S. A. M., Riggs, D. S., Feeny, N. C., et al. (2005). Randomized trial of prolonged exposure for posttraumatic stress disorder with and without cognitive restructuring: Outcome at academic and community clinics. *Journal of Consulting and Clinical Psychology*, *73*, 953−964.

Foa, E. B., Keane, T. M., & Friedman, M. J. (Eds.). (2004). *Effective treatments for PTSD: Practice guidelines from the International Society for Traumatic Stress Studies*. New York: Guilford Press.

Foa, E. B., & Rauch, S. A. M. (2004). Cognitive changes during prolonged exposure alone versus prolonged exposure and cognitive restructuring in female assault survivors with PTSD. *Journal of Consulting and Clinical Psychology*, *70*, 879−884.

*Foa , E. B., Rothbaum, B. O., Riggs, D. S., & Murdock, T. B. (1991). Treatment of posttraumatic stress disorder in rape victims: A comparison between cognitive–behavioral procedures and counseling. *Journal of Consulting and Clinical Psychology*, *59*, 715−723.

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*, 181−191.

Glynn, S. M., Randolph, E. T., Foy, D. W., Urbaitis, M., Boxer, L., Paz, G. G., et al. (1999). A test of behavioral family therapy to augment exposure for combat-related posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology*, *67*, 243−251.

Green, B. L., & Kaltman, S. I. (2002). Recent research findings on the diagnosis of posttraumatic stress disorder: Prevalence, course, comorbidity, and risk. In R. I. Simon (Ed.), *Posttraumatic stress disorder in litigation: Guidelines for forensic assessment, 2nd edition* (pp. 19−39). Orlando, FL: American Psychiatric Press, Inc.Hedges, L. V., & Olkin, I. (2002). *Statistical methods for meta-analysis* (pp. 19−39). Washington, DC: Academic Press.

Hedges, L. V., & Olkin, I. (1985). Statistical methods for meta-analysis therapy outcome studies: A replication and refinement. *Psychological Bulletin*, *92*, 581−604.

[1] References marked with an asterisk indicate studies included in the meta-analysis.

Hembree, E. A., Foa, E. B., Dorfan, N. M., Street, G. P., Kowalski, J., & Tu, X. (2003). Do patients drop out prematurely from exposure therapy for PTSD? *Journal of Traumatic Stress*, *16*, 555−562.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*, 1539−1558.

Hoge, I. W., Castro, C. A., Messer, S. C., McGurk, D., Cotting, D. I., & Koffman, R. L. (2004). Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. *The New England Journal of Medicine*, *351*, 13−22.

Howard, K. I., Krause, M. S., Saunders, S. M., & Kopta, S. M. (1997). Trials and tribulations in the meta-analysis of treatment differences: Comment on Wampold et al. (1997). *Psychological Bulletin*, *122*, 221−225.

Hox, J. (2002). Multilevel analysis Mahwah, NJ: Lawrence Erlbaum.

Huedo-Medina, T. B., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I2 index? *Pyschological Methods*, *11*(2), 193−206.

*Ironson, G., Freund, B., Strauss, J. I., & Williams, J. (2002). Comparison of two treatments for traumatic stress: A community-based study of EMDR and prolonged exposure. *Journal of Clinical Psychology*, *58*, 113−128.

Kessler, R. C. (2000). Posttraumatic stress disorder: The burden to the individual and to society. *Journal of Clinical Psychiatry*, *61*, 4−12.

Kessler, R. C., Sonnega, A., Bromet, E., Hughes, M., & Nelson, C. (1995). Posttraumatic stress disorder in the National Comorbidity Survey. *Archives of General Psychiatry*, *52*, 1048−1060.

Kraemer, H. C., & Kupfer, D. J. (2006). Size of treatment effects and their importance to clinical research and practice. *Biological Psychiatry*, *59*, 990−996.

Kreft, I., & DeLeeuw, J. (1998). Introducing multilevel modeling Oaks, CA: Sage.

Lambert, M. J., & Ogles, B. (2004). The efficacy and effectiveness of psychotherapy. In M. J. Lambert (Ed.), *Handbook of psychotherapy and Behavior Change (5th ed.)* (pp. 139−193). New York: John Wiley.

*Lee, C., Gavriel, H., Drummond, P., Richards, J., & Greenwald, R. (2002). Treatment of PTSD: Stress inoculation training with prolonged exposure compared to EMDR. *Journal of Clinical Psychology*, *58*, 1071−1089.

Lee, C. W., Taylor, G., & Drummond, P. D. (2006). The active ingredient in EMDR: Is it traditional exposure or dual focus of attention? *Clinical Psychology and Psychotherapy*, *13*, 97−107.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment. Confirmation from meta-analysis. *Journal of American Psychology*, *48*(12), 1181−1209.

*Marks , I., Lovell, K., Noshirvani, H., Livanou, M., & Thrasher, S. (1998). Treatment of posttraumatic stress disorder by exposure and/or cognitive restructuring: A controlled study. *Archives of General Psychiatry*, *55*, 317−325.

*McDonagh, A., Friedman, M., McHugo, G., Ford, J., Sengupta, A., Mueser, K., et al. (2005). Randomized trial of cognitive-behavioral therapy for chronic posttraumatic stress disorder in adult female survivors of childhood sexual abuse. *Journal of Consulting and Clinical Psychology*, *73*(3), 515−524.

McFarlane, A. C., & Yehuda, R. (2000). Clinical treatment of posttraumatic stress disorder: Conceptual challenges raised by recent research. *Australian and New Zealand Journal of Psychiatry*, *34*, 940−953.

Nemeroff, C. B., Bremner, J. D., Foa, E. B., Mayberg, H. S., North, C. S., & Stein, M. B. (2006). Posttraumatic stress disorder: A state-of-the-science review. *Journal of Psychiatric Research*, *40*(1), 1−21.

*Paunovic, N., & Ost, L. (2001). Cognitive-behavior therapy vs exposure therapy in the treatment of PTSE in refugees. *Behavior Research and Therapy*, *39*, 1183−1197.

Pitman, R. K., Altman, B., Greenwald, E., Longpre, R. E., Macklin, M. L., & Poire, R. E. (1991). Psychiatric complications during flooding therapy for post-traumatic stress disorder. *Journal of Clinical Psychiatry*, *52*, 17−20.

*Power, K., McGoldrick, T., Brown, K., Buchanan, R., Sharp, D., Swanson, V., et al. (2002). A controlled comparison of eye movement desensitization and reprocessing versus exposure plus cognitive restructuring versus waiting list in the treatment of post-traumatic stress disorder. *Clinical Psychology and Psychotherapy*, *9*, 229−318.

Quintana, S. M., & Minami, T. (2006). Guidelines for meta-analyses of counseling psychology research. *The Counseling Psychologist*, *34*(6), 839−877.

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods Newbury Park, CA: Sage.

Renfrey, G., & Spates, C. R. (1994). Eye movement desensitization: A partial dismantling study. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(3), 231−239.

*Resick, P. A., Nishith, P., Weaver, T. L., Astin, M. C., & Feuer, C. A. (2002). A comparison of cognitive-processing therapy with prolonged exposure and a waiting condition for the treatment of chronic post-traumatic stress disorder in female rape victims. *Bulletin of Menninger Clinic*, *61*, 317−334.

Robertson, M., Humphreys, L., & Ray, R. (2004). Psychological treatments for posttraumatic stress disorder: Recommendations for the clinician based on a review of the literature. *Journal of Psychiatric Practice*, *10*, 106−118.

Rosenthal, D., & Frank, J. D. (1956). Psychotherapy and the placebo effect. *Psychological Bulletin*, *53*, 294−302.

*Rothbaum, B. O., Astin, M. C., & Marsteller, F. (2005). Prolonged exposure versus eye movement desensitization and reprocessing (EMDR) for PTSD rape victims. *Journal of Traumatic Stress*, *18*, 607−616.

Rotton, L., Foos, P. W., Van Meek, L., & Levitt, M. (1995). Publication practices and the file drawer problem: A survey of published authors. *Journal of Social Behavior and Personality*, *10*, 1−13.

Schnurr, P. P., Friedman, M. J., Foy, D. W., Shea, M. T., Heish, F. Y., Lavori, P. W., Glynn, S. M., et al. (2003). Randomized trial of trauma-focused group therapy for post-traumatic stress disorder. *Archives of General Psychiatry*, *60*, 481−489.

Serlin, R. C., Wampold, B. E., & Levin, J. R. (2003). Should providers of treatment be regarded as a random factor? If it ain't broke, don't "fix" it: Comment on Siemer & Joormann (2003). *Psychological Methods*, *8*, 524−534.

Shadish, W. R., Jr., & Sweeney, R. B. (1991). Mediators and moderators in meta-analysis: There's a reason we don't let Dodo birds tell us which psychotherapies should have prizes. *Journal of Consulting and Clinical Psychology*, *59*, 883−893.

Siedler, G. H., & Wagner, F. E. (2006). Comparing the efficacy of EMDR and trauma focused cognitive behavioral therapy in the treatment of PTSD: A meta-analytic study. *Psychological Medicine*, *36*, 1515−1522.

Smith, M. L., & Glass, G. V. (1977). Meta-analysis of psychotherapy outcome studies. *American Psychologist*, *32*, 752−760.

Snijders, T., & Bosker, R. (1999). Multilevel analysis Thousand Oaks, CA: Sage.

Stapleton, J. A., Taylor, S., & Asmundson, G. J. (2006). Effects of three PTSD treatments on anger and guilt: Exposure therapy, eye movement desensitization and reprocessing, and relaxation training. *Journal of Traumatic Stress*, *19*(1), 19−28.

*Tarrier , N., Pilgrim, H., Sommerfield, C., Faragher, B., Reynolds, M., Graham, E., et al. (1999). A randomized trial of cognitive therapy and imaginal exposure in the treatment of chronic posttraumatic stress disorder. *Journal of Consulting and Clinical Psychology*, *67*, 13−18.

*Taylor, S., Thordarson, D. S., Maxfield, L., Fedoroff, I. C., Lovell, K., & Ogrodniczuk, J. (2003). Comparative efficacy, speed, and adverse effects of three PTSD treatments: Exposure therapy, EMDR, and relaxation training. *Journal of Consulting and Clinical Psychology*, *71*(2), 330−338.

Van Etten, M. L., & Taylor, S. (1998). Comparative efficacy of treatments for post-traumatic stress disorder: A meta-analysis. *Clinical Psychology and Psychotherapy*, *5*, 126−144.

Vaughan, K., Armstrong, M. S., Gold, R., O'Connor, N., Jenneke, W., & Tarrier, N. (1994). A trial of eye movement desensitization compared to image habituation training and applied muscle relaxation in post-traumatic stress disorder. *Journal of Behavior Therapy and Experimental Psychiatry*, *24*, 283−291.

Wampold, B. E. (2001). Th\e great psychotherapy debate: Model, methods and findings Mahwah, NJ: Lawrence Erlbaum.

Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes". *Psychological Bulletin*, *122*, 203−215.

Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: Assumptions, findings, and reporting in controlled clinical trials. *Psychological Bulletin*, *130*, 631−663.