Cognitive-Behavioral Therapy for Adult Anxiety Disorders in Clinical Practice: A Meta-Analysis of Effectiveness Studies

Article *in* Journal of Consulting and Clinical Psychology · September 2009

DOI: 10.1037/a0016032 · Source: PubMed

CITATIONS
200

READS
3,565

2 authors, including:

Rebecca E Stewart
University of Pennsylvania
17 PUBLICATIONS **585** CITATIONS

SEE PROFILE

Journal of Consulting and Clinical Psychology
2009, Vol. 77, No. 4, 595–606

0022-006X/09/$12.00 DOI: 10.1037/a0016032

Cognitive–Behavioral Therapy for Adult Anxiety Disorders in Clinical Practice: A Meta-Analysis of Effectiveness Studies

Rebecca E. Stewart and Dianne L. Chambless
University of Pennsylvania

The efficacy of cognitive–behavioral therapy (CBT) for anxiety in adults is well established. In the present study, the authors examined whether CBT tested under well-controlled conditions generalizes to less-controlled, real-world circumstances. Fifty-six effectiveness studies of CBT for adult anxiety disorders were located and synthesized. Meta-analytic effect sizes are presented for disorder-specific symptom measures as well as symptoms of generalized anxiety and depression for each disorder, and benchmarked to results from randomized controlled trials. All pretest–posttest effect sizes for disorder-specific symptom measures were large, suggesting that CBT for adult anxiety disorders is effective in clinically representative conditions. Six studies included a control group, and between-groups comparisons yielded large effect sizes for disorder-specific symptoms in favor of CBT. Benchmarking indicated that results from effectiveness studies were in the range of those obtained in selected efficacy trials. To test whether studies that are more representative of clinical settings have smaller effect sizes, the authors coded studies for 9 criteria for clinical representativeness. Results indicate an inverse relationship between clinical representativeness and outcome, but the magnitude of the relationship is quite small.

Keywords: effectiveness, dissemination, clinical practice, cognitive–behavioral therapy, anxiety disorders

Supplemental materials: http://dx.doi.org/10.1037/a0016032.supp

Cognitive–behavioral therapy (CBT) appears prominently among the empirically supported treatments (ESTs) for adult anxiety disorders (Chambless & Ollendick, 2001). Several meta-analyses of well-controlled clinical trials provide support for the efficacy of CBT for panic disorder, social anxiety disorder, obsessive–compulsive disorder (OCD), generalized anxiety disorder (GAD), and posttraumatic stress disorder (PTSD; see Deacon & Abramowitz, 2004, for a review). In the most recent meta-analysis, Norton and Price (2007) examined the efficacy of CBT across the anxiety spectrum. Results indicated that treatments that used CBT techniques showed significantly larger treatment outcome effect sizes than no treatment or placebo across all of the anxiety disorders. Taken together, these multiple meta-analyses indicate that CBT is an efficacious treatment for adult anxiety disorders.

Editor's Note. Rick E. Ingram served as the action editor for this article.—AML

Rebecca E. Stewart and Dianne L. Chambless, Department of Psychology, University of Pennsylvania.

We thank David B. Wilson and Betsy Becker for assistance with statistical analyses. We also thank John Paul Jameson for coding the studies and Amber Calloway for literature searches, as well as the many respondents to our listserv queries and those authors who provided additional data.

Correspondence concerning this article should be addressed to Rebecca E. Stewart, Department of Psychology, University of Pennsylvania, Solomon Laboratories, 3720 Walnut Street, Philadelphia, PA 19104. E-mail: restewar@psych.upenn.edu

How well does CBT for anxiety disorders hold up in actual clinical practice? This general question of the transportability of efficacious interventions into naturalistic settings has been one of the most contentious issues in the ongoing debate of evidence-based practice in clinical psychology (e.g., Jacobson & Christensen, 1996). Skeptics question whether the procedures used to maximize experimental control in randomized controlled efficacy trials seriously compromise the external validity of the results. Specifically, there are questions about patients, clinicians, and treatments used in research settings, and whether these characteristics of clinical trials are representative, appropriate, or relevant to routine clinical practice.

The claim is often made that research treatments will not work in clinical practice settings because the clients in practice settings are purported to be more severe or to have more comorbid conditions than clients treated in research settings. According to this argument, in research settings patients are recruited specifically for research, and patients with comorbid disorders are often excluded to achieve homogeneous diagnostic samples (Westen & Morrison, 2001). It has been suggested that these highly selected groups are not representative of patients who present in outpatient practice (Silberschatz in Persons & Silberschatz, 1998). It has also been suggested that patient assent to randomization further limits the generalizability of the sample (Seligman, 1995). Moreover, patient expectations of specialist treatment in a research trial might be higher than patient expectations in nonresearch settings (Sanderson, Raue, & Wetzler, 1998), which may further enhance motivation and outcome in research settings. There is disagreement in the literature as to whether populations used in randomized controlled efficacy trials are in fact as selected as assumed (see Stirman, DeRubeis, Crits-Christoph, & Brody, 2003; Stirman, DeRubeis, Crits-Christoph, & Rothman, 2005). Nonetheless,

this suggestion that patients in research settings are less severe, more motivated, and somehow easier may limit the degree to which it is believed that results from research settings can generalize to actual clinical practice.

Another concern about the transportability of ESTs to real world settings is the treatments themselves and the clinicians who provide them. Treatment protocols in randomized controlled trials are manualized and strictly monitored with an emphasis on treatment integrity. Therapy manuals are less likely to be used in clinical practice, and their relevance to practice has been questioned (Seligman, 1995). Furthermore, front-line practitioners typically do not have access to the level of intensive training, monitoring, and supervision available to therapists in research settings (Chambless & Hollon, 1998). Clinicians in research settings are more likely to be expert in the administration of particular treatments and are motivated through adherence measures to stay consistent with the protocol. Moreover, research therapists often have the luxury of focusing exclusively on one type of problem or disorder, whereas average practitioners carry large caseloads covering a wide range of focal problems. In summary, treatments delivered in naturalistic settings may not be as rigorous in terms of content or quality, and this may limit how well results of controlled research trials can generalize to actual clinical practice.

Whether one believes these criticisms are valid or significant (for responses, see Chambless & Ollendick, 2001; Persons in Persons & Silberschatz, 1998), one major implication is that the opponents of controlled psychotherapy research may be unconvinced that efficacy findings are applicable in actual clinical practice. One solution to the perceived shortcomings of the traditional controlled conditions of efficacy research is the treatment effectiveness study (Hoagwood, Hibbs, Brent, & Jensen, 1995). Effectiveness research explores the transportability of efficacious interventions (such as ESTs) to real-world service settings, to examine whether these treatments result in similar, beneficial effects when used in more naturalistic settings. Whereas efficacy studies focus on minimizing threats to a study's internal validity and determining the causal factors of therapeutic change, the emphasis in effectiveness studies is placed on maximizing external validity. Effectiveness studies focus on the effects of psychotherapy conducted in the field, and can include pretest–posttest, quasi-experimental, or experimental designs. External validity is achieved by utilizing one or more of the following clinically representative qualities: clinically representative settings (e.g., private practice or mental health centers), clinically representative therapists (e.g., practicing clinicians for whom provision of services is a substantial part of the job), or clinically representative patients (e.g., few exclusion criteria or patients who refuse randomization).

It is often thought that efficacy and effectiveness studies are mutually exclusive. However, it is more productive and accurate to consider them as studies with different foci on internal and external validity operating on a continuum (Hunsley & Lee, 2007). For example, there is no reason why controlled efficacy trials cannot take place in applied clinical settings with minimal exclusion criteria and clinically representative therapists as hybrid effectiveness–efficacy studies maximizing both internal and external validity (Chambless & Hollon, 1998; for an example, see Blomhoff et al., 2001). Efficacy trials offer a particularly compelling means of testing for causal agency, allowing for confident conclusions regarding the efficacy of active treatments in comparison with control treatments. However, demonstration of efficacy is considered only a first step in treatment research (Argras & Berkowitz, 1980). Effectiveness studies are required to demonstrate the transportability and generalization of efficacious interventions into actual clinical practice.

Since the 1995 special section in the *Journal of Consulting and Clinical Psychology* highlighted the importance of effectiveness research, the literature on outcome research in clinical settings has burgeoned. In an early review, Weisz, Donenberg, Han, and Weiss (1995) proposed a collection of variables that pertain to clinical relevance and utilized these criteria to locate nine clinically representative child and adolescent therapy studies (see also Weisz, Weiss, & Donenberg, 1992). Weisz et al. concluded the effectiveness of clinic therapy was modest or nonsignificant when compared with research therapy. Weisz et al. intended their conclusion to hold only for child and adolescent therapy studies, and they also noted that their results should be interpreted with caution given the small number of studies they located. Moreover, it is critical to note that Weisz et al. combined treatments of all types and disorders of all types, and this finding may not hold when clinic and research therapy are compared for a specific treatment (i.e., CBT) for specific disorders (i.e., anxiety disorders).

Shadish, Matt, Navarro, and Phillips (2000) conducted a meta-analysis of 90 therapy outcome studies drawn from published meta-analyses located in a literature and including Weisz et al.'s (1995) original nine studies. Building upon and expanding Weisz et al.'s criteria for clinically representative studies, the authors utilized a graduated scale of clinical representativeness. Their 10 criteria were based on use in past research (e.g., Weisz et al., 1995), consistency with empirical literature on clinical practice, and face validity. The criteria were as follows:

> (a) clinically representative problems, (b) clinically representative setting, (c) clinically representative referrals, (d) clinically representative therapists, (e) clinically representative structure, (f) clinically representative monitoring, (g) clinically representative problem heterogeneity, (h) pretherapy training, (i) therapy freedom, and (j) flexible number of sessions. (Shadish et al., 2000, p. 514)

Using multiple regression to predict effect size from the clinical representativeness scale total score, Shadish et al. (2000) found that after controlling for confounds—such as therapy dose and outcome specific measures—clinical representativeness was unrelated to effect size. The authors concluded that this study supports the effectiveness of psychotherapy under clinically representative conditions.

We now return to our original question: How well does CBT for anxiety disorders in adults hold up in actual clinical practice? Although important and informative, the inferences that may be drawn from Shadish et al.'s (2000) study for these questions are limited because effectiveness studies of CBT of adult anxiety disorders were not the primary focus of this work. Similar to Weisz et al. (1995), Shadish et al. included psychotherapy treatments of all types and disorders of all types. Accordingly, the first goal of the present study is to conduct a meta-analysis of effectiveness studies for anxiety disorders to determine whether the benefits of CBT tested under well-controlled circumstances generalize to less-controlled, more real-world circumstances. We report analyses of pretest–posttest effect sizes as well as group contrast effect sizes

for those studies that included a control group. We then use a benchmarking strategy to assess whether the pretest–posttest effect sizes achieved in effectiveness studies are comparable with effect sizes obtained in controlled outcome efficacy trials. The second goal of the present study is to expand on the research of Shadish et al. by testing whether the degree of clinical representativeness is related to the effect size of outcome.

Method

Studies

Studies included in this meta-analysis utilized CBT for any adult anxiety disorder encompassed under the current *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; American Psychiatric Association, 1994) nomenclature. CBT was defined broadly and included any treatment with cognitive, behavioral (e.g., exposure), or a combination of components. Effectiveness studies utilizing brief therapy (fewer than six standard sessions) and transdiagnostic CBT were excluded because the efficacy of these forms of CBT for anxiety disorders has not yet been well established in efficacy studies. Although their effectiveness is an important question in its own right, bibliotherapy and computer-directed therapy were excluded because they were considered to be too different from what goes on in actual clinical practice to be pertinent to the goals of the present article. We also excluded treatments that used psychotropic medication as part of the treatment protocol because our interest was in the effectiveness of psychosocial interventions.

We located studies via a search of abstracts in PsycINFO using the following keywords: effectiveness, generalization, dissemination, naturalistic, transporting, private practice, managed care setting, outpatient clinic, community clinic, community mental health center, cognitive–behavioral therapy, cognitive therapy, and behavior therapy. In addition, the major journals publishing effectiveness studies were checked by hand from 1995 to 2008: *Behavior Therapy*, *Behaviour Research and Therapy*, *Cognitive Research and Therapy*, *Cognitive and Behavioural Practice*, *Journal of Anxiety Disorders*, and the *Journal of Consulting and Clinical Psychology*. The year 1995 was selected as the lower limit for journal hand searching because the influential special section on effectiveness research in the *Journal of Consulting and Clinical Psychology* was published in that year. This special section defined effectiveness studies, highlighted its importance for psychotherapy research, and initiated a new direction in the field.

To avoid any studies being missed because of the heterogeneity of descriptor and keyword items, as well as to locate any unpublished work, conference presentations, or works in progress or press, we sent networking e-mails to the electronic mailing lists of the following societies: Association of Behavioral and Cognitive Therapies, Society for the Science and Practice of Clinical Psychology, Society for Psychotherapy Research, and the Academy of Cognitive Therapy. Lastly, reference sections of located articles and other relevant chapters and papers were reviewed for potentially eligible studies. Fifty-four potential studies were located from the PsycINFO, journal, and reference searches. Fifteen potential studies were located through networking. Thirteen studies were excluded on the basis of these criteria: 6 studies did not meet our minimum clinically representative cutoff of three (see below), and 7 studies did not provide sufficient data to include in the meta-analysis. Efforts were made to contact the authors of these studies, but (a) data were not available from the authors, or (b) we were unable to contact or received no response from the authors. In sum, a total of 56 studies were included in these analyses: 17 for panic disorder; 11 each for social anxiety disorder, OCD, and GAD; and 6 for PTSD. One study (Westbrook & Kirk, 2005) included data on CBT outcomes for all disorders with the exception of PTSD, and therefore it is counted more than once. No effectiveness studies were located for specific phobias.

Participants. Available reported patient characteristics of each study are compiled and presented in the supplemental materials on the journal's website. In brief, the majority of patients were female (range = 37%–100%, unweighted *Mdn* = 68.3%) and in their mid-30s (range = 31–71 years, unweighted *Mdn* = 35 years). On average, half or less of the sample had a college education (range = 0%–71%, unweighted *Mdn* = 33%), and the majority of the patients were employed full time (range = 8%–88%, unweighted *Mdn* = 59%). When reported, Axis I comorbidity was common (range = 32%–85%, unweighted *Mdn* = 55.4%). Although the average patient was Caucasian, African Americans or Caribbean Americans of African descent made up at least 20% of the sample in six studies. Latinos were represented at this level in only two studies.

Coding clinical representativeness. Codes were based on Shadish et al. (2000) and modified for the present study. Four of Shadish et al.'s original codes were excluded in the present study on the basis of practicality and theory. By nature of this study's focus on effectiveness studies with anxiety disorders, all studies would meet criteria for clinically representative problems (a). Alternatively, no studies would meet criteria for clinically representative heterogeneity (g) or therapy freedom (i) because we only included data from anxiety disorder patients treated with CBT. We also excluded Shadish et al.'s criteria of flexible number of sessions (j) because we do not agree that a flexible number of sessions is necessarily clinically representative. Managed care often poses strict limitations on the number of sessions a patient may receive. Moreover, given the great percentage of uninsured Americans, many patients pay out of pocket, which may also restrict the number of sessions they can afford and receive. We modified clinically representative structure to include whether strict, flexible, or no manualization was utilized. We added three criteria to the remaining six, on the basis of their usage in effectiveness studies and partly for their face validity: no randomization, clinically representative patients (i.e., no exclusion criteria aside from psychosis, suicidality, organic brain disease, or substance abuse), and allowance of medication.[1] The resultant nine criteria are as follows: clinically representative settings, clinically representative referrals, clinically representative therapists, clinically representative structure, clinically representative monitoring, no pretherapy training of therapists, no randomization, clin-

[1] Although we excluded studies that used medication as part of the treatment protocol, it is common for anxiety patients who present in clinical practice to be on varying levels of psychotropic medications, as prescribed by psychiatrists or general medical practitioners. As a result, we included the allowance of medication as a criterion for clinical representativeness to express this feature of clinical practice. In contrast, in most randomized controlled trials of psychological treatments, patients are withdrawn from medication before initiation of the study protocol or required to maintain a stable dosage throughout treatment.

ically representative patients, and allowance of medication. Scores on these criteria were summed to yield a total clinical representativeness score for each study. A coding manual was developed and is presented in the Appendix.

As noted earlier, studies fall on a continuum from efficacy to effectiveness in nature. Any determination of where on that continuum a study must fall to be classified as an effectiveness study is necessarily arbitrary. For the purposes of this study, we selected an a priori cutoff score of three on the clinical representativeness scale constructed. Such a score would be achieved, for example, if a study was conducted in a clinically representative setting, with clinically representative patients treated by clinically representative therapists.

Coding of the clinically representativeness criteria followed a rigorous examination of the methods in each study. Many of the studies reviewed were clear in their explanations of these characteristics. In the few cases in which the information was not reported, we contacted the study authors. Rebecca E. Stewart coded all studies, and a second coder independently coded 24 out of the 56 studies. Reliability for the total score was excellent, $\rho_I(3,1) = .83$ (Shrout & Fleiss, 1979). The codes for each study are available in the supplemental materials on the journal's website.

Effect Size Calculation and Statistical Procedures

Standardized mean gain. Standardized mean gain (pretest–posttest) effect sizes were computed for diagnosis-specific outcome measures. In addition, because generalized anxiety and depression symptoms are common complaints of patients with anxiety disorders and were often assessed, we computed effect sizes for these measures as well. For panic disorder, there were three disorder-specific symptom constructs: frequency of attacks, fear of fear, and avoidance measures. For OCD, social anxiety disorder, and PTSD, there was one disorder-specific construct. In the special case of GAD, generalized anxiety measures were used as diagnosis-specific outcome measures. Because of the paucity of intent-to-treat data (4 studies out of 56), completer data were used in this study. We calculated Cohen's *d* for the pretest–posttest effect sizes using the pooled standard deviation (see Dunlap, Cortina, Vaslow, & Burke, 1996).

The effect size was adjusted to yield Hedges's *g* (Hedges, 1981) and weighted to account for sample size. The weights were based on the standard errors of effect size (Lipsey & Wilson, 2001). The standard error formula for repeated measures requires the use of the correlation *r* between pretest and posttest measures. This value was never reported in the studies, although it is possible to derive a value of *r* from the means, standard deviations, and the paired *t*-test value with the following formula (D. Wilson, personal communication, February 27, 2008):

$$r = \frac{(s_1^2 t^2 + s_2^2 t^2) - (\overline{X_1} - \overline{X_2})^2}{2 s_1 s_2 t^2}.$$

However, only one fourth of the studies reported a paired *t*-test between pretest and posttest conditions. These *r*s were calculated, converted to *z*, averaged, and converted back to *r* to deduce an overall correlation *r* to be used in following calculations ($r = .41$). We also did a sensitivity analysis by repeating the analyses with correlation *r*s of .2 and .6, and we found that the results did not differ substantially. The results of the sensitivity analysis are presented in the supplemental materials on the journal's website. On the basis of the averaged *r*, the standard error for each effect size for each study was calculated as follows (Lipsey & Wilson, 2001):

$$SE = \sqrt{\frac{2(1 - r)}{n} + \frac{ES^2}{2n}}.$$

Lastly, we calculated the weights of each effect size using the inverse variant weight, which is the reciprocal of the squared standard error:

$$w = \frac{1}{SE^2}.$$

Given the heterogeneity of the sample (see below), a priori random effects meta-analyses proceeded as follows. The effect size, standard error, and inverse variance weights were calculated for each construct measured in the study: the disorder-specific constructs, as well as generalized anxiety and depression. The weighted mean effect size for each construct was computed for each disorder, according to the formula:

$$\overline{ES} = \frac{\Sigma(w_i ES_i)}{\Sigma w_i}.$$

The standard error of each weighted mean effect size was also calculated:

$$SE = \sqrt{\frac{1}{\Sigma w_i}}.$$

Individual studies often reported multiple measures on a given construct. Multiple measures on one construct would violate assumptions of independence, inflate the sample size, and distort standard error estimates. Therefore, a single effect size was calculated for each construct for each study by averaging the multiple measures to result in a single effect size for each construct for each study.

Homogeneity analysis. Statistical tests based on the *Q* statistic (Hedges & Olkin, 1985) and I^2 (Higgins & Thompson, 2002) indicated significant heterogeneity among panic disorder, social anxiety disorder, and PTSD effect sizes. This was not surprising given differences in methods across these studies, such as the lack of common measures. Accordingly, we adopted a random effects model for the analyses. Random effects analyses have the advantage of allowing generalization to the potential population of studies. Heterogeneity analyses for OCD and GAD indicated that the distributions of observed effect sizes were homogenous. However, to gain the greater generalizability to a potential population of studies permitted by the random effects approach, random effects models were adopted for all analyses.

Standardized mean difference. The standardized mean difference effect size was calculated from posttreatment data to evaluate between-groups differences for those studies that included a control group. This analysis was completed to examine whether CBT treatment groups in clinically representative studies yield significantly improved outcomes when compared with control groups. These included waiting list ($n = 3$), treatment as usual ($n = 2$), and contact control ($n = 1$) groups. Hedges's *g* was calculated as

before. A positive sign indicates the effect size favors the CBT treatment group over the control group, whereas a negative sign denotes that the control group has an advantage over the treatment group. The standard error for these analyses was computed as follows (Lipsey & Wilson, 2001):

$$SE = \sqrt{\frac{n_{g1} + n_{g2}}{n_{g1}n_{g2}} + \frac{ES^2}{2(n_{g1} + n_{g2})}}.$$

The inverse variance weights were calculated as above. Heterogeneity analyses based on the Q statistic and I^2 indicated significant heterogeneity, and random effects meta-analyses proceeded as described above. Because there were only six studies that included a control group, we collapsed across disorder and used the disorder-specific symptom measures for each construct for each disorder. In the case of panic disorder (for which there were three constructs), these constructs were averaged to yield one disorder-specific construct.

Selection of benchmarking studies. We adopted a benchmarking strategy to assess the transportability of CBT in tightly controlled experimental studies to clinically representative studies. The strategy allows us to determine whether the magnitude of improvement in clinically representative studies is comparable with that obtained in research settings. We selected three efficacy studies per disorder to use as benchmarks against which to compare our standardized mean gain (pretest–posttest) effect sizes. We culled all appropriate efficacy studies from the most recent meta-analysis of CBT for adult anxiety disorders (Norton & Price, 2007) and determined the sample size for each study. Many studies included multiple treatment conditions that meet our definition of CBT treatment (e.g., applied relaxation, stress inoculation training). In these cases, we included all CBT-like treatment conditions, and we utilized a weighted average (dependent on sample size) across treatments to yield one effect size for each disorder. Because larger samples provide the most stable estimates of effect size, we selected for each disorder the three studies with the largest samples—provided that the study yielded the appropriate data—and calculated the pretest–posttest effect sizes for completer analyses in the manner reported above. For each disorder, this yielded a range of effect sizes from randomized controlled studies against which to benchmark our results.

Clinical representativeness. The second goal of the present study was to examine the relationship of clinical representativeness to outcome. One disorder-specific pretest–posttest effect size was calculated for each study as described above. In the case of panic disorder (for which there were three constructs), these constructs were averaged to yield one disorder-specific construct. We used an a priori linear contrast to examine the relationship between clinical representativeness and effect size (Rosenthal & Rosnow, 1991).

File Drawer Problem

One concern for meta-analyses is the so-called file drawer problem, which describes systematic upward bias due to the omission of studies that have been conducted but never reported because of trivial or null results (Rosenthal, 1979). Our networking strategy to locate unpublished work is one potential resolution to the file drawer problem. Another solution is to examine the scatterplot of effect size by sample size. Examination of the funnel plot revealed a classic funnel pattern, indicating greater variability in effect sizes from studies with smaller samples. This suggests that the effect sizes come from an unbiased distribution and that small sample data with negative results have not been suppressed (Light, Singer, & Willett, 1994).

Rosenthal (1979) responded to the file-drawer problem by developing a statistic to estimate the threat posed by potential unpublished studies. The fail-safe N estimates the number of unpublished studies reporting null results that would be necessary to reduce the cumulative effect across the meta-analytic studies to the point of nonsignificance. Orwin (1983) calculated the formula for application to standardized mean difference effect sizes. This formula can be applied to our pretest–posttest meta-analysis (Lipsey & Wilson, 2001), where k_0 is the number of effect sizes with a value of 0 needed to reduce the mean effect size to ES_c, the criterion effect size level. ES_k is the weighted mean effect size, and k is the number of studies utilized to calculate the mean effect size.

$$k_0 = k\left[\frac{\overline{ES_k}}{\overline{ES_c}} - 1\right].$$

The criterion effect size level is the effect size at which we would no longer consider our results significant. We selected a value of 0.40, which resulted from a between-groups meta-analysis (fixed effects) of the four studies that included a no treatment (waiting list) control. Results indicate that 122 studies with null effects would be necessary to offset our overall pretest–posttest effect size. It is important to note that our criterion effect size level of 0.40, although substantially smaller than our weighted mean effect size (1.29), is still of a moderate size and statistically significant. Thus, this is a conservative estimate of the fail-safe N. Hence, it is likely that our conclusions are not in error because of publication bias.

Results

How Effective Is CBT for Anxiety Disorders in Clinically Representative Conditions?

Table 1 summarizes the uncontrolled pretest–posttest effect size estimates for each disorder. Positive effect sizes indicate improvement from pretreatment to posttreatment. All disorder-specific effect sizes (range = 0.83–2.59) are large (Cohen, 1998). Effect sizes for reductions in depression and generalized anxiety symptoms were only calculated if there was a sufficient number of studies (five or more) reporting these data. Aside from GAD (in which generalized anxiety is considered disorder-specific), only the panic disorder studies reported enough data from generalized anxiety measures to calculate an effect size. The effect size for reduction in generalized anxiety symptoms for panic disorder is large. All depression symptom effect sizes (range = 0.73–1.60) are large, with the exception of social anxiety disorder, for which the effect size for depression symptoms is medium. All effect sizes are significant at $p < .0001$.

Is CBT More Effective Than Control Groups?

Six studies included a control group against which to compare treatment outcomes: two panic studies, one OCD study, and three

Table 1
Effectiveness Pretest–Posttest Effect Sizes by Disorder

Disorder/symptoms	*n*	Effect size	*SE*	95% confidence interval	*z*	*p*
Panic disorder						
Attacks	9	1.01	0.12	0.77–1.25	8.32	<.0001
Avoidance	14	0.83	0.12	0.60–1.06	7.09	<.0001
Fear of fear	11	1.23	0.16	0.92–1.54	7.84	<.0001
Depression	17	1.01	0.08	0.86–1.17	13.01	<.0001
Generalized anxiety	14	1.02	0.13	0.77–1.26	8.04	<.0001
Social anxiety disorder						
Social anxiety symptoms	11	1.04	0.13	0.79–1.29	8.14	<.0001
Depression	8	0.73	0.09	0.55–0.91	7.90	<.0001
Posttraumatic stress disorder (PTSD)						
PTSD symptoms	6	2.59	0.27	2.06–3.13	9.50	<.0001
Depression	5	1.62	0.32	0.99–2.25	5.06	<.0001
Generalized anxiety disorder						
Generalized anxiety	11	0.92	0.08	0.77–1.07	11.94	<.0001
Depression	5	0.89	0.09	0.70–1.07	9.39	<.0001
Obsessive–compulsive disorder (OCD)						
OCD symptoms	11	1.32	0.07	1.19–1.45	20.15	<.0001
Depression	9	0.89	0.06	0.77–1.01	14.65	<.0001

GAD studies. For the disorder-specific measures, random effects analyses produced an effect size of $g = 1.29$, $p < .001$, confidence interval = 0.76–1.83. The positive sign indicates an advantage of CBT treatments over control groups, with a large effect size. Following Rosenthal and Rosnow (1991), we converted this effect size into a binomial effect size display (BESD) to yield a more clinically meaningful metric. The BESD translates a continuous effect size into its dichotomous equivalent, in this case improved versus not improved. The degree of superiority of CBT over control conditions indicated by a *g* of 1.29 is comparable with a 78% improvement rate for CBT patients versus a 22% improvement rate for patients in the control conditions.

Benchmarking Effectiveness Studies Against Efficacy Studies

Three benchmarking studies for each disorder were identified, and pretest–posttest effect sizes were calculated as described above (see Table 2). Foa et al.'s (2005) study is noteworthy because it included samples from both an academic and community clinic for the treatment of PTSD. The community clinic data were included in the meta-analysis, and the sample size at the academic clinic designated this study as one of our benchmarks against which to compare our meta-analytic results for PTSD. Overall, the effectiveness effect sizes from this study were in the range of the effect sizes obtained in our selected efficacy trials. Specifically, the present study's effect sizes for social anxiety disorder, OCD, and GAD are contained within the range of the efficacy benchmarking studies, whereas the effectiveness effect size for PTSD exceeds the range of the controlled studies. However, the effect size for panic disorder falls somewhat below the range of the benchmarking studies.

Does Clinical Representativeness Predict Effect Size?

Scores on the clinical representativeness scale ranged from 3 to 9 ($M = 6.62$, $SD = 1.76$). A linear contrast was significant with a small effect size, $t(2526) = -2.03$, $p < .05$, $d = -0.08$. As degree of clinical representativeness gets larger, the outcome improvement effect size becomes smaller, but the relationship is slight. Given the overall significant relationship between clinical representativeness and outcome, exploratory post hoc contrast analyses were conducted to gain insight into which individual components in the overall score might lead to this finding. Reliability was examined for the individual ratings and deemed acceptable if the intraclass correlation coefficient exceeded .65 (see the Method section for coding specifics). Three of the nine variables (clinically representative referrals, therapists, and patients) were not used for further analyses because of unacceptable reliability. Reliabilities for the remaining six variables was adequate to excellent [$\rho_I(3,1) = .66–.88$, $Mdn = .77$]. A Bonferroni correction was applied to protect against Type I error, yielding an effective *p* value of .008 for significance (.05/6).

Clinically representative setting was not significantly related to outcome, $t(2525) = 2.29$, $p = .03$, $d = 0.06$. Outcome effect sizes increased significantly when medication was allowed, $t(2525) = 5.87$, $p \leq .0001$, $d = 0.17$, and when patients were not randomized to treatments, $t(2392) = 3.69$, $p < .0002$, $d = 0.11$. In contrast, outcome effect sizes decreased significantly when there was no training for therapists, $t(2392) = -10.83$, $p < .0001$, $d = -0.31$; when therapists were not asked to follow a manual (clinically representative structure), $t(2392) = -5.49$, $p < .0001$, $d = -0.16$; and when there was little or no monitoring to make sure the treatment was followed (clinically representative monitoring), $t(2392) = -6.24$, $p < .0001$, $d = -0.18$. All effects are small.

Discussion

The primary objective of this study was to determine whether CBT for anxiety disorders works in actual clinical practice. A secondary goal was to test whether clinical representativeness in these studies was related to the effect size in a partial replication of Shadish et al. (2000).

Table 2
Benchmark Efficacy Studies by Disorder Versus Current Effectiveness Results: Pretest–Posttest Effect Sizes

Disorder/study	CBT treatments	*n*	Effect size
Panic disorder			
Kenardy et al. (2003)	CBT	42	1.53
Öst and Westling (1995)	AR, CT	36	1.23
Barlow et al. (2000)	CBT	56	1.43
Effectiveness			1.02
Social anxiety disorder			
Stangier et al. (2003)	CBT, CBGT	40	0.89
Clark et al. (2006)	CT, ERP	42	1.75
Davidson et al. (2004)	CBT	48	1.47
Effectiveness			1.04
Posttraumatic stress disorder			
Foa et al. (2005)	PE, PE/CR	56	2.50
Resick et al. (2002)	CPT, PE	81	2.49
Foa et al. (1999)	PE, SIT, PE/SIT	64	1.90
Effectiveness			2.59
Generalized anxiety disorder			
Borkovec and Costello (1993)	AR, CBT	36	2.26
Borkovec et al. (2002)	CT, SCD, CBT	58	1.96
Barlow et al. (1992)	AR, CT, AR/CT	34	0.84
Effectiveness			0.92
Obsessive–compulsive disorder			
McLean et al. (2001)	CBT, ERP	63	1.15
Whittal et al. (2005)	CBT, ERP	59	1.88
Van Oppen et al. (1995)	CT, ERP	57	1.24
Effectiveness			1.45

Note. CBT = cognitive–behavioral therapy; AR = applied relaxation; CT = cognitive therapy; CBGT = cognitive–behavioral group therapy; ERP = exposure and response prevention; PE = prolonged exposure; CR = cognitive restructuring; CPT = cognitive processing therapy; SIT = stress inoculation training; SCD = self-control desensitization.

Does CBT for adult anxiety disorders work in actual clinical practice? The results from this meta-analysis provide evidence supporting the effectiveness of CBT techniques with the anxiety disorders. All pretest–posttest effect sizes for disorder-specific symptom measures were large, indicating that patients treated with CBT in clinically representative studies improved significantly and substantially from pretest when they completed treatment. Moreover, CBT for anxiety disorders produced significant pretest–posttest reductions in depression symptoms with large effects across panic disorder, PTSD, GAD, and OCD, and a medium effect for social anxiety disorder. Six of our studies included a control group against which to compare posttreatment outcomes, and the effect size was large for disorder-specific symptom measures across the anxiety disorders. Although the latter findings are limited (on the basis of six studies), the results suggest that treatments using CBT techniques in clinically representative conditions lead to significantly larger treatment effect sizes than waiting list or treatment as usual control groups. The effect size comparing CBT with control conditions, translated into a BESD, was comparable with a substantial difference in improvement rates—78% for CBT versus 22% for control groups. Overall, these results indicate that CBT is effective in clinical settings.

We utilized a benchmarking strategy to evaluate whether the magnitude of improvement in clinically representative studies is in the range of selected efficacy studies. Overall, this was the case. The clinically representative studies of social anxiety disorder, OCD, and GAD generated effect size estimates that were within the range of effect sizes from our selected efficacy trials. In the case of PTSD, the effectiveness effect sizes exceeded the range of the benchmarking studies. Only for panic disorder was the effect size smaller than the smallest of the benchmarking effect sizes, although it is important to note that this effect size is still large. The discrepant finding for panic disorder may have been due to the noise introduced by the great variety of symptom measures used for this disorder. In general, these results provide initial evidence for the generalizability of CBT for adult anxiety disorders from highly controlled research settings to typical clinical settings.

Is clinical representativeness related to treatment outcome? At face value, our results do not replicate Shadish et al. (2000), in that we found a slight inverse relationship between clinical representativeness and effect size. These results suggest that effect sizes are lower in the more clinically representative studies, which is consistent with the results of Weisz et al. (1995). However, it is critical to note that although the test was statistically significant, the effect size in the present study was extremely small ($d = -0.08$), suggesting that the impact of clinical representativeness on effect size is minor. In contrast, Weisz et al. found an effect size of $d = 0.62$ in favor of research settings over clinic settings.

Post hoc exploratory analyses of the individual variables contributing to the overall clinical representativeness score indicated that outcome effect sizes increase when patients are not randomized to treatments and when medication is permitted. The latter is not surprising given that the potential additive effects of anxiolytic and antidepressant medication with psychotherapy. On the other hand, our results indicate that when therapists are not trained, do not use manuals, and are not monitored to ensure they are carrying out the intended treatment, outcome effect sizes decrease. This may be a result of patients improving less, greater error in these studies because of the uncontrolled variance in therapists' behavior (Crits-Christoph et al., 1991), or some combination of both. These results are consistent with prior controlled research indicating that patients with anxiety disorders may improve more with standardized treatments than with therapist-generated treatment plans (Schulte, Kunzel, Pepping, & Schulte-Bahrenberg, 1992) and that therapists become more skilled at delivering an EST when they receive supervised training (Sholomskas et al., 2005). Nonetheless, these post hoc analyses must be interpreted with caution because of their exploratory nature and the correlational nature of these results.

Limitations in this study suggest caution in interpretation. The first limitation is a construct validity issue: Do our codes assess legitimate and central measures of clinical representativeness? Although our codes were built on past research (e.g., Shadish et al., 2000; Weisz et al., 1995), it is likely that there will be disagreements about key elements we omitted or added to build the current coding schema. It is also important to note that the codes were developed (in current and past research) by academics who value empirical testing of psychological treatments. The real-world mental health practitioner may not agree that studies included in this meta-analysis are clinically representative. More research and input is needed from front-line practitioners so that we can better approximate and codify clinically representative conditions. In addition, although reliability for the total clinically representative score was excellent, three of the nine clinically representative characteristics (clinically representative referrals, therapists, and

patients) were not reliably coded, limiting further analyses. It is not clear why this was the case. However, studies differed on the quality of reporting this information, most notably for these three criteria, and coders likely made judgment calls on the basis of the limited information provided. Improved reporting and greater detailed methodology in effectiveness studies should be encouraged so that the reader may understand how closely these studies estimate clinically representative conditions.

Another concern in meta-analysis is the quality of included research studies and the measures utilized to represent constructs. Although we selected studies that met our inclusion criteria, there was still variability in the overall quality of studies, the quality of measures, and reporting from which we had to infer codes. In their meta-analysis of efficacy studies, Norton and Price (2007) used only established measures as selected by Antony, Orsillo, and Roemer (2001) in their compendium of assessment instruments for anxiety disorders. We found this inclusion criterion unnecessarily prohibitive. Owing to the small number of clinically representative studies in the literature, we prioritized including studies over eliminating studies that may not have used measures included by Antony et al. In the current study, we determined on a case-by-case basis whether to include a measure if we found it to adequately assess the construct of interest, and not all measures had undergone extensive psychometric evaluation.

Another limitation of this sample of studies is the lack of measures on clinical or applied significance of the effects of CBT (e.g., reliable change statistics, quality of life, disability). Although our BESD demonstration was one effort to display the practical importance of our results in a manner that is meaningful to researchers and clinicians, there remains little information in these papers about what these changes mean for the average patient's functioning. Such measures are helpful in determining whether CBT is making a genuine difference in the lives of recipients. More research addressing the applied importance of CBT in clinically representative conditions is sorely needed.

A fourth issue that complicates interpretation is the use of pretest–posttest analyses. As previously mentioned, because of limitations in cost, organizational structure, ethical concerns, feasibility, and the focus on external validity, many effectiveness studies do not include a control group. Hence, our findings must be interpreted with some caution because we cannot rule out alternative explanations for patients' improvement—such as external events, the passage of time, regression to the mean, and the effects of the assessment itself. One solution to this problem is our analysis of the six studies that included a control group, thus permitting the conclusion that the intervention was responsible for the change. There is also a body of efficacy studies of anxiety disorders indicating that patients receiving no treatment respond very minimally. Norton and Price (2007) found an average effect size of $d = 0.25$ in the placebo conditions, which is substantially less than any CBT effect size produced by our meta-analysis. Taken together, these two pieces of evidence suggest that it was likely the CBT, rather than alternative factors, that produced the effects reported in this article.

Nonetheless, additional controlled research is needed to permit a more definitive statement about the causal role of CBT in patients' change. Although such research may be difficult to conduct (Weisz & Addis, 2006), the importance of such studies in evaluating arguments for or against dissemination of ESTs is clear. At the same time, it would be a mistake to overlook the importance of uncontrolled effectiveness research. One of the most contentious issues in evidence-based practice is the extent to which results from randomized controlled trials can be generalized to routine clinical practice. Uncontrolled effectiveness research permits the researcher to maximize external validity by testing treatments (with prior supporting efficacy research) in all types of naturalistic circumstances to evaluate whether these treatments translate well to the clinical setting. For example, such studies permit us to determine whether patients who have not and, in some cases, would not agree to randomization in a research trial nonetheless improve when they receive an EST—a point that has been in some dispute in the literature (e.g., Seligman, 1995). In our view, an argument is most powerful when results from research emphasizing external validity and research emphasizing internal validity converge as they do here.

Another issue complicating interpretation of the results is the use of completer data in this meta-analysis. Few studies (4 out of 56) included reports of intention-to-treat data. Completer analyses may inflate the apparent results of treatment, in that those who dropped out of treatment are not represented in such samples. It is not the case that all of those who drop out of treatment are treatment failures (Bados, Balaguer, & Saldaña, 2007); indeed, some patients leave therapy early because they believe they are doing well. Nonetheless, it is undoubtedly true that patients who leave treatment frequently do so because it is unacceptable to them or because they are not benefiting. Most high-quality journals now require authors of randomized controlled trials to report intention-to-treat analyses. Effectiveness researchers should be encouraged to do so as well.

Many authors of the effectiveness studies included here also failed to report key characteristics of their samples. It is critical that researchers provide tables or summaries of patient characteristics so that readers can determine how generalizable findings might be to patients of interest to them. Such data as were reported attest to the need for more research with African American and especially Latino patients. The dearth of such research is of particular concern given the growing number presence of Latinos and other ethnic minorities in the U.S. population (U.S. Census Bureau, 2001).

A final limitation of this article is that we did not complete a formal meta-analytic comparison of effectiveness effect sizes with efficacy effect sizes. Such an effort was beyond the scope of the present research, but a formal meta-analytic comparison between efficacy and effectiveness data would be a valuable next step.

In summary, this meta-analysis provides initial evidence that CBT in clinically representative conditions is robustly effective across the adult anxiety disorders. Moreover, the effect sizes found in conditions approximating actual clinical practice are, on the whole, within the range of effect sizes produced in our selected efficacy benchmark studies. More research is needed to determine whether this effectiveness generalizes to CBT treatments for children with anxiety disorders, other ESTs for anxiety, or to CBT for other disorders in children and adults.

References

References marked with an asterisk indicate studies included in the meta-analysis. References marked with a plus indicate benchmarking studies.

*Addis, M. E., Hatgis, C., Krasnow, A. D., Jacob, K., Bourne, L., & Mansfield, A. (2004). Effectiveness of cognitive–behavioral treatment for panic disorder versus treatment as usual in a managed care setting. *Journal of Consulting and Clinical Psychology, 72,* 625–635.

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

Antony, M. M., Orsillo, S. M., & Roemer, L. (2001). *Practitioner's guide to empirically-based measures of anxiety.* New York: Kluwer Academic/Plenum.

Argras, W. S., & Berkowitz, R. (1980). Clinical research in behavior therapy: Halfway there? *Behavior Therapy, 11,* 472–488.

*Arntz, A. (2003). Cognitive therapy versus applied relaxation as treatment of generalized anxiety disorder. *Behaviour Research and Therapy, 41,* 633–646.

Bados, A., Balaguer, G., & Saldaña, C. (2007). The efficacy of cognitive–behavioral therapy and the problem of drop-out. *Journal of Clinical Psychology, 63,* 585–592.

*Bakker, A., Spinhoven, P., van Balkom, A. J. L. M., Vleugel, L., & van Dyke, R. (2000). Cognitive therapy by allocation versus cognitive therapy by preference in the treatment of panic disorder. *Psychotherapy and Psychosomatics, 69,* 240–243.

+Barlow, D. H., Gorman, J. M., Shear, M. K., & Woods, S. W. (2000). Cognitive–behavioral therapy, imipramine, or their combination for panic disorder: A randomized controlled trial. *Journal of the American Medical Association, 19,* 2529–2536.

+Barlow, D., Rapee, R. M., & Brown, T. (1992). Behavioral treatment of generalized anxiety disorder. *Behavior Therapy, 23,* 551–570.

Blomhoff, S., Haug, T. T., Hellstrom, K., Holme, I., Humble, M., Madsbu, H. P., et al. (2001). Randomized controlled general practice trial of sertraline, exposure therapy and combined treatment in generalized social phobia. *British Journal of Psychiatry, 179,* 23–30.

+Borkovec, T. D., & Costello, E. (1993). Efficacy of applied relaxation and cognitive–behavioral therapy in the treatment of generalized anxiety disorder. *Journal of Consulting and Clinical Psychology, 61,* 611–619.

+Borkovec, T. D., Newman, M. G., Pincus, A. L., & Lytle, R. (2002). A component analysis of cognitive–behavioral therapy for generalized anxiety disorder and the role of interpersonal problems. *Journal of Consulting and Clinical Psychology, 70,* 288–298.

*Burke, M., Drummond, L. M., & Johnston, D. W. (1997). Treatment choice for agoraphobic women: Exposure or cognitive–behaviour therapy? *British Journal of Clinical Psychology, 36,* 409–420.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66,* 7–18.

Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology, 52,* 685–716.

*Chambless, D. L., Tran, G. Q., & Glass, C. R. (1997). Predictors of response to cognitive–behavioral group therapy for social phobia. *Journal of Anxiety Disorders, 11,* 221–239.

+Clark, D. M., Ehlers, A., Hackmann, A., McManus, F., Fennell, M., Grey, N., et al. (2006). Cognitive therapy versus exposure and applied relaxation in social phobia: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 74,* 569–578.

Cohen, J. (1998). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Crits-Christoph, P., Baranackie, K., Kurcias, J., Beck, A. T., Carroll, K., Perry, K., et al. (1991). Meta-analysis of therapist effects in psychotherapy outcome studies. *Psychotherapy Research, 1,* 81–91.

+Davidson, J. R. T., Foa, E. B., Huppert, J. D., Keefe, F. J., Franklin, M. E., Compton, J. S., et al. (2004). Fluoxetine, comprehensive cognitive behavioral therapy, and placebo in generalized social phobia. *Archives of General Psychiatry, 61,* 1005–1013.

Deacon, B. J., & Abramowitz, J. S. (2004). Cognitive and behavioral treatments for anxiety disorders: A review of meta-analytic findings. *Journal of Clinical Psychology, 60,* 429–441.

*Deacon, B., & Abramowitz, J. (2006). A pilot study of two-day cognitive–behavioral therapy for panic disorder. *Behaviour Research and Therapy, 44,* 807–817.

Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods, 2,* 170–177.

*Durham, R. C., Fisher, P. L., Dow, M. G. T., Sharp, D., Power, K. G., Swan, J. S., & Morton, R. V. (2004). Cognitive behaviour therapy for good and poor prognosis generalized anxiety disorder: A clinical effectiveness study. *Clinical Psychology and Psychotherapy, 11,* 145–157.

*Ellis, T. E., & Lehman, G. L. (1994). *A naturalistic outcome study of group cognitive behavioral treatment for panic disorder.* Paper presented at the 28th Annual Convention of the Association for Advancement of Behavior Therapy, San Diego, CA.

*Feske, U. (2001). Treatment of low-income and African-American women with posttraumatic stress disorder: A case series. *Behavior Therapy, 32,* 585–601.

+Foa, E. B., Dancu, C. V., Hembree, E. A., Jaycox, L. H., Meadows, E. A., & Street, G. P. (1999). 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. A., 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.

*Franklin, M. E., Abramowitz, J. S., Kozak, M. J., Levitt, J. T., & Foa, E. B. (2000). Effectiveness of exposure and ritual prevention for obsessive–compulsive disorder: Randomized compared with nonrandomized samples. *Journal of Consulting and Clinical Psychology, 69,* 594–602.

*Friedman, S., Braunstein, J. W., & Halpern, B. (2006). Cognitive behavioral treatment of panic disorder and agoraphobia in a multiethnic urban outpatient clinic: Initial presentation and treatment outcome. *Cognitive and Behavioral Practice, 13,* 282–292.

*Friedman, S., Smith, L. C., Halpern, B., Levine, C., Paradis, C., Viswanathan, R., et al. (2003). Obsessive–compulsive disorder in a multi-ethnic urban outpatient clinic: Initial presentation and treatment outcome with exposure and ritual prevention. *Behavior Therapy, 34,* 397–410.

*García-Palacios, A., Botella, C., Robert, C., Baños, R., Perpiñá, C., Quero, S., & Ballester, R. (2002). Clinical utility of cognitive–behavioural treatment for panic disorder. Results obtained in different settings: A research center and a public mental health care unit. *Clinical Psychology and Psychotherapy, 9,* 373–383.

*Gaston, J. E., Abbott, M. J., Rapee, R. M., & Neary, S. A. (2006). Do empirically supported treatments generalize to private practice? A benchmark study of a cognitive–behavioural group treatment programme for social phobia. *British Journal of Clinical Psychology, 45,* 33–48.

*Gillespie, K., Duffy, M., Hackmann, A., & Clark, D. M. (2002). Community based cognitive therapy in the treatment of post-traumatic disorder following the Omagh bomb. *Behaviour Research and Therapy, 40,* 345–357.

*Hahlweg, K., Fiegenbaum, W., Frank, M., Schroeder, B., & von Witzleben, I. (2001). Short- and long-term effectiveness of an empirically supported treatment for agoraphobia. *Journal of Consulting and Clinical Psychology, 69,* 375–382.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6,* 107–128.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

*Heldt, E., Manfro, G. G., Kipper, L., Blaya, C., Maltz, S., Isolan, L., et al. (2003). Treating medication-resistant panic disorder: Predictors and outcome of cognitive–behavior therapy in a Brazilian public hospital. *Psychotherapy and Psychosomatics, 72,* 43–48.

*Herbert, J. D., Gaudiano, B. A., Rheingold, A. A., Myers, V. H., Dalrymple, K., & Nolan, E. M. (2005). Social skills training augments the effectiveness of cognitive behavioral group therapy for social anxiety disorder. *Behavior Therapy, 36,* 125–138.

*Herbert, J. D., Rheingold, A. A., Gaudiano, B. A., & Myers, V. H. (2004). Standard versus extended cognitive behavior therapy for social anxiety disorder: A randomized-controlled trial. *Behavioural and Cognitive Psychotherapy, 32,* 131–147.

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

Hoagwood, K., Hibbs, E., Brent, D., & Jensen, P. (1995). Introduction to the special section: Efficacy and effectiveness in studies of child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63,* 683–687.

Hunsley, J., & Lee, C. M. (2007). Research-informed benchmarks for psychological treatments, efficacy studies, effectiveness studies, and beyond. *Professional Psychology: Research and Practice, 38,* 21–33.

*Hunt, C., & Andrews, G. (1998). Long-term outcome of panic disorder and social phobia. *Journal of Anxiety Disorders, 12,* 395–406.

Jacobson, N. S., & Christensen, A. (1996). Studying the effectiveness of psychotherapy: How well can clinical trials do the job? *American Psychologist, 51,* 1031–1039.

+Kenardy, J. A., Dow, M. G. T., Johnston, D. W., Newman, M. G., Thomson, A., & Taylor, C. B. (2003). A comparison of delivery methods of cognitive–behavioral therapy for panic disorder: An international multicenter trial. *Journal of Consulting and Clinical Psychology, 71,* 1068–1075.

*Levitt, J. T., Malta, L. S., Martin, A., Davis, L., & Cloitre, M. (2007). The flexible application of a manualized treatment for PTSD symptoms and functional impairment related to the 9/11 World Trade Center attack. *Behaviour Research and Therapy, 45,* 1419–1433.

Light, R. J., Singer, J. D., & Willett, J. B. (1994). The visual presentation and interpretation of meta-analyses. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 439–453). New York: Russell Sage Foundation.

*Lincoln, T. M., Rief, W., Hahlweg, K., Frank, M., von Witzleben, I., Schroeder, B., & Fiegenbaum, W. (2003). Effectiveness of an empirically supported treatment for social phobia in the field. *Behaviour Research and Therapy, 41,* 1251–1269.

*Linden, M., Zubraegel, D., Franke, U., & Schlattmann, P. (2005). Efficacy of cognitive behaviour therapy in generalized anxiety disorders. *Psychotherapy and Psychosomatics, 74,* 36–42.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage.

*Marom, S., Gilboa-Schechtman, E., Aderka, I. M., Weizman, A., & Hermesh, H. (2009). Impact of depression on treatment effectiveness and gains maintenance in social phobia: A naturalistic study of cognitive behavior group therapy. *Depression and Anxiety, 26,* 289–300.

*Martinsen, E. W., Olsen, T., Tønset, E., Nyland, K., & Aarre, T. F. (1998). Cognitive–behavioral group therapy for panic disorder in the general clinical setting: A naturalistic study with 1-year follow-up. *Journal of Clinical Psychiatry, 59,* 437–442.

*McEvoy, P. M. (2007). Effectiveness of cognitive behavioural group therapy for social phobia in a community clinic: A benchmarking study. *Behaviour Research and Therapy, 45,* 3030–3040.

+McLean, P. D., Whittal, M. L., Thordarson, D. S., Taylor, S., Söchting, I., Koch, W. J., et al. (2001). Cognitive versus behavior therapy in the group treatment of obsessive–compulsive disorder. *Journal of Consulting and Clinical Psychology, 69,* 205–214.

*Mortberg, E., Berglund, G., & Sundin, O. (2005). Intensive cognitive behavioural group treatment of social phobia: A pilot study. *Cognitive Behaviour Therapy, 34,* 41–49.

*Mortberg, E., Karlsson, A., Fyring, C., & Sundin, O. (2006). Intensive cognitive–behavioral group treatment (CBGT) of social probia: A randomized controlled study. *Anxiety Disorders, 20,* 646–660.

*Norberg, M. M., Calamari, J. E., Cohen, R. J., & Riemann, B. C. (2008). Quality of life in obsessive–compulsive disorder: An evaluation of impairment and a preliminary analysis of the ameliorating effects of treatment. *Depression and Anxiety, 25,* 248–259.

Norton, P. J., & Price, E. C. (2007). A meta-analytic review of adult cognitive–behavioral treatment outcome across the anxiety disorders. *The Journal of Nervous and Mental Disease, 195,* 521–531.

Orwin, R. G. (1983). A fail-safe *N* for effect size in meta-analysis. *Journal of Educational Statistics, 8,* 157–159.

+Öst, L., & Westling, B. E. (1995). Applied relaxation versus cognitive behavior therapy in the treatment of panic disorder. *Behavior Research and Therapy, 33,* 145–158.

Persons, J. B., & Silberschatz, G. (1998). Are results of randomized controlled trials useful to psychotherapists? *Journal of Counseling and Clinical Psychology, 66,* 126–135.

+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 list condition for the treatment of chronic posttraumatic stress disorder in female rape victims. *Journal of Consulting and Clinical Psychology, 70,* 867–879.

*Rosenberg, N. K., & Hougaard, E. (2005). Cognitive–behavioural group treatment of panic disorder and agoraphobia in a psychiatric setting: A naturalistic study of effectiveness. *Nordic Journal of Psychiatry, 59,* 198–204.

Rosenthal, R. (1979). The "file drawer problem" and tolerance for null results. *Psychological Bulletin, 86,* 638–641.

Rosenthal, R., & Rosnow, R. L. (1991). *Essentials of behavioral research: Methods and data analysis* (2nd ed.). New York: McGraw-Hill.

*Rothbaum, B. O. (2000). Behavioral treatment of obsessive–compulsive disorder in a naturalistic setting. *Cognitive and Behavioral Practice, 7,* 262–270.

*Sanderson, W. C., Raue, P. J., & Wetzler, S. (1998). The generalizability of cognitive behavior therapy for panic disorder. *Journal of Cognitive Psychotherapy: An International Quarterly, 12,* 323–330.

Schulte, D., Kunzel, R., Pepping, G., & Schulte-Bahrenberg, T. (1992). Tailor-made versus standardized therapy of phobic patients. *Advances in Behaviour Research and Therapy, 14,* 67–92.

*Schultz, P. M., Resick, P. A., Huber, L. C., & Griffin, M. G. (2006). The effectiveness of cognitive processing therapy for PTSD with refugees in a community setting. *Cognitive and Behavioral Practice, 13,* 322–331.

Seligman, M. E. P. (1995). The effectiveness of psychotherapy: The Consumer Reports study. *American Psychologist, 50*(12), 965–974.

Shadish, W. R., Matt, G. E., Navarro, A. M., & Phillips, G. (2000). The effects of psychological therapies under clinically representative conditions: A meta-analysis. *Psychological Bulletin, 126,* 512–529.

*Sharp, D. M., Power, K. G., & Swanson, V. (2000). Reducing therapist contact in cognitive behaviour therapy for panic disorder and agoraphobia in primary care: Global measures of outcome in a randomized controlled trial. *British Journal of General Practice, 40,* 963–968.

*Sharp, D. M., Power, K. G., & Swanson, V. (2004). A comparison of the efficacy and acceptability of group versus individual cognitive behaviour therapy in the treatment of panic disorder and agoraphobia in primary care. *Clinical Psychology and Psychotherapy, 11,* 73–82.

Sholomskas, D. E., Syracuse-Siewert, G., Rounsaville, B. J., Ball, S. A., Nuro, K. F., & Carroll, K. M. (2005). We don't train in vain: A dissemination trial of three strategies of training clinicians in cognitive–

behavioral therapy. *Journal of Consulting and Clinical Psychology, 73,* 106–115.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin, 86,* 420–428.

[+]Stangier, U., Heidenreich, T., Peitz, M., Lauterbach, W., & Clark, D. M. (2003). Cognitive therapy for social phobia: Individual versus group treatment. *Behaviour Research and Therapy, 41,* 991–1007.

*Stanley, M. A., Hopko, D. R., Diefenbach, G. J., Bourland, S. L., Rodriguez, H., & Wagener, P. (2003). Cognitive–behavior therapy for late-life generalized anxiety disorder in primary care. *American Journal of Geriatric Psychiatry, 11,* 92–96.

Stirman, S. W., DeRubeis, R. J., Crits-Christoph, P., & Brody, P. E. (2003). Are samples in randomized controlled trials of psychotherapy representative of community outpatients? A new methodology and initial findings. *Journal of Consulting and Clinical Psychology, 71,* 963–972.

Stirman, S. W., DeRubeis, R. J., Crits-Christoph, P., & Rothman, A. (2005). Can the randomized controlled trial literature generalize to nonrandomized patients? *Journal of Consulting and Clinical Psychology, 73,* 127–135.

*Tolin, D. F., Hannan, S., Maltby, N., Diefenbach, G. J., Worhunsky, P., & Brady, R. E. (2007). A randomized controlled trail of self-directed versus therapist-directed cognitive–behavioral therapy for obsessive–compulsive disorder patients with prior medication trials. *Behavior Therapy, 38,* 179–191.

*Tolin, D. F., Maltby, N., Diefenbach, G. J., Hannan, S. E., & Worhunsky, P. (2004). Cognitive–behavioral therapy for medication nonresponders with obsessive–compulsive disorder: A wait-list-controlled open trial. *Journal of Clinical Psychiatry, 65,* 922–931.

U.S. Census Bureau. (2001). *Profile of general demographic characteristics: 2000 census of population and housing, United States.* Retrieved February 16, 2009, from http://www.census.gov/population/www/socdemo/race/race.html

*Van Noppen, B., Steketee, G., McCorkle, B. H., & Pato, M. (1997). Group and multifamily behavioral treatment for obsessive compulsive disorder: A pilot study. *Journal of Anxiety Disorders, 11,* 431–446.

[+]Van Oppen, P., de Haan, E., van Balkom, A. J. L. M., Spinhoven, P., Hoogduin, K., & van Dyck, R. (1995). Cognitive therapy and exposure in vivo in the treatment of obsessive compulsive disorder. *Behavior Research and Therapy, 33,* 379–390.

*Wade, W. A., Treat, T. A., & Stuart, G. L. (1998). Transporting an empirically supported treatment for panic disorder to a service clinic setting: A benchmarking strategy. *Journal of Consulting and Clinical Psychology, 66,* 231–239.

*Warren, R., Strand, J., & Thomas, J. C. (2008). Panic control treatment in private practice: Immediate and long term results. *Manuscript in preparation.*

*Warren, R., & Thomas, J. C. (2001). Cognitive–behavior therapy of obsessive–compulsive disorder in private practice: An effectiveness study. *Anxiety Disorders, 15,* 277–285.

Weisz, J. R., & Addis, M. E. (2006). The research-practice tango and other choreographic challenges: Using and testing evidence-based psychotherapies in clinical care settings. In C. D. Goodheart, A. E. Kazdin, & R. J. Sternberg (Eds.), *Evidence-based psychotherapy: Where practice and research meet* (pp. 179–206). Washington, DC: American Psychological Association.

Weisz, J. R., Donenberg, G. R., Han, S. S., & Weiss, B. (1995). Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63,* 699–701.

Weisz, J. R., Weiss, B., & Donenberg, G. R. (1992). The lab versus the clinic: Effects of child and adolescent psychotherapy. *American Psychologist, 47,* 1578–1585.

*Westbrook, D., & Kirk, J. (2005). The clinical effectiveness of cognitive behaviour therapy: Outcome for a large sample of adults treated in routine practice. *Behaviour Research and Therapy, 43,* 1243–1261.

Westen, D., & Morrison, K. (2001). A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: An empirical examination of the status of empirically supported therapies. *Journal of Consulting and Clinical Psychology, 69,* 875–899.

Westen, D., Novotny, C. M., & Thompson-Brenner, H. (2004). The empirical status of empirically supported psychotherapies: Assumptions, findings, and reportings in controlled clinical trials. *Psychological Bulletin, 4,* 631–663.

*Wetzel, C., Bents, H., & Florin, I. (1999). High-density exposure therapy for obsessive–compulsive inpatients: A 1-year follow-up. *Psychotherapy and Psychosomatics, 68,* 186–192.

*White, J., & Keenan, M. (1992). Stress control: A controlled comparative investigation of large group therapy for generalized anxiety disorder. *Behavioural Psychotherapy, 20,* 97–114.

[+]Whittal, M. L., Thordarson, D. S., & McLean, P. D. (2005). Treatment of obsessive–compulsive disorder: Cognitive behavior therapy vs. exposure and response prevention. *Behaviour Research and Therapy, 43,* 1559–1576.

(*Appendix follows*)

Appendix

Clinical Representativeness Coding Manual

Criterion for clinical representativeness	Definition and code
Clinically representative setting	1: Setting where clinical services are commonly provided in nonuniversity settings (e.g., outpatient mental health clinics, CMHCs, general hospitals, private practices, prisons, school systems) in which the primary purpose is clinical care. ½: Patient setting affiliated with a university research program but primary purpose of setting is clinical care. 0: University research setting.
Clinically representative referrals	1: Clients initially referred through usual clinical routes (e.g., MDs). ½: Clients referred through usual clinical routes (e.g., MDs) *and* recruitment solicitations/advertisements. 0: Clients referred by recruitment solicitations/advertisements.
Clinically representative therapists	1: Practicing and experienced clinicians with regular caseloads for whom provision of services is a substantial part of job. The primary therapists at the clinical setting—interns are included. 0: Research therapists.
Clinically representative structure	1: Treatment either with a structure used in clinical practice or not structured in a detailed and uniform way according to a manual (e.g., general CBT). ½: Treatment "based on" or "inspired by" a clinical manual, which can be used flexibly (e.g., include treatment procedures not in the manual, flexible number of sessions). 0: Strict manualization, or manual-driven treatment (e.g., set number of sessions).
Clinically representative monitoring	1: Implementation of treatment was not monitored in any way that could influence therapist behavior (*no* formal adherence checks and *no* supervision). ½: Only monitoring was supervision and/or outlines containing information to be covered in session (*no* formal adherence checks). 0: Supervision and adherence checks.
No pretherapy training	1: Therapists did *not* receive special training immediately before study in specific techniques to be used. Of course, therapists may have received training in that treatment at some point in their career but not for purposes of study in which they participated. ½: Only novice therapists are trained, and they represent less than 25% of therapists. 0: Pretherapy training; also give 0 when percentage of novice therapists not specified.
No randomization	1: Patients were not part of a randomized trial, did not meet criteria for a particular trial, or refused randomization. 0: Patients were randomized to treatments.
Clinically representative patients	1: No exclusionary criteria aside from psychosis, suicidality, organic brain disease, or substance dependence if patient meets criteria for disorder under study. 0: Exclusion criteria beyond those above, including comorbidity and medication.
Medications allowed	1: Medication is allowed. If no specific mention in exclusion criteria, assume medication is allowed. 0: Patients on medication are forced to go off medication or are excluded from study.

Note. CMHCs = community mental health centers; MDs = medical doctors; CBT = cognitive–behavioral therapy.

Received June 9, 2008
Revision received February 19, 2009
Accepted April 3, 2009 ■