Psychological Bulletin
2005, Vol. 131, No. 5, 785–795

0033-2909/05/$12.00 DOI: 10.1037/0033-2909.131.5.785

Meta-Analysis of Cognitive–Behavioral Treatments for Generalized Anxiety Disorder: A Comparison With Pharmacotherapy

Kristin Mitte
University of Jena

The efficacy of (cognitive) behavioral therapy ([C]BT) for generalized anxiety disorder was investigated and compared with the efficacy of pharmacological therapy using meta-analytic techniques. A total of 65 (C)BT studies and pharmacological studies were included. (C)BT was more effective than control conditions. The results of the comparison between (C)BT and pharmacotherapy varied according to the meta-analytic methods used. Conclusions about differences in efficacy between therapy approaches are limited when all available studies are included owing to a number of factors that influence effect sizes. When only those studies that directly compared both therapies were included in the analysis, there were no significant differences in efficacy. Attrition rates were lower for (C)BT, indicating that it is better tolerated by patients.

Keywords: generalized anxiety disorder, cognitive–behavior therapy, drug therapy, meta-analysis

Generalized anxiety disorder (GAD) is a common mental disorder. Data collected in the United States indicate that approximately 5% of the general population suffers from the disorder at least once in their lifetime (Kessler et al., 1994). Women are more often affected than men, with prevalence rates approximately twice as high (Wittchen, Zhao, Kessler, & Eaton, 1994). Furthermore, Kessler, Mickelson, Barber, and Wang (2001) found that in comparison with those experiencing 25 other common physical conditions and mental disorders, people with GAD reported the highest number of days off work, with an average of 6 days per month prior to taking part in the study.

In the third edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM–III*; American Psychiatric Association, 1980), GAD was defined as uncontrollable and diffuse anxiety or worry with several related psychophysiological symptoms that persists for 1 month or longer. The third edition, revised (*DSM–III–R*; American Psychiatric Association, 1987), and the fourth edition (*DSM–IV*; American Psychiatric Association, 1994) included some changes in the diagnostic criteria. First, the minimum duration of the symptoms was increased from 1 month to 6 months. Second, the cognitive symptoms of chronic worrying became the defining features of GAD. In contrast, in *DSM–III* the most important criteria related to symptoms of arousal, for example, somatic symptoms and muscle tension. However, even now, some authors continue to emphasize the associated symptoms of arousal and question the utility of focusing on worry (see also Rickels & Rynn, 2001).

(Cognitive) behavior therapy ([C]BT) is commonly used to treat patients with GAD. Such treatments include, for example, applied relaxation, cognitive restructuring of dysfunctional beliefs, and cognitive exposure to worry. However, compared with psychopharmacological approaches, research has paid little attention to psychological treatments. Although benzodiazepines have long been used, drugs of the class of azapirones (e.g., buspirone) and antidepressants (e.g., selective serotonin reuptake inhibitors; SSRIs) are now used. Thus, there is a need to address the question of whether (C)BT is effective in the treatment of GAD and also to compare its efficacy with that of pharmacotherapy.

Although a meta-analysis including the results of 35 studies has already been carried out on this issue (Gould, Otto, Pollack, & Yap, 1997), a new one is now required, as a number of additional relevant efficacy studies have since been conducted and new meta-analytic techniques have been developed. The present meta-analysis includes the following improvements. First, it provides a comprehensive quantitative summary of 65 controlled studies. Second, the random-effects model (REM) was used to compute average effect sizes and regression analyses. Two statistical models are used in meta-analyses, which differ in their statistical and sampling assumptions and in the conclusions drawn: the fixed-effects model (FEM) and the REM. In the FEM, the results of a meta-analysis are restricted to the studies included. Results "apply to *this* collection of studies and say nothing about other studies that may be done later, could have been done earlier, or may have already been done but are not included among the observed studies" (Hedges & Vevea, 1998, p. 487). In contrast, when the REM is used, results can be generalized beyond the studies selected and inferences apply to treatment efficacy in general, which is more appropriate in current research syntheses. Third, the present meta-analysis included several sensitivity analyses. In research syntheses, many meta-analytic decisions are made that can affect the results, for example, which studies are included or how effect sizes

The research reported in this article was supported by a personal Grant of the State of Thuringia (Germany). Results were partly presented at the 33rd Congress of the European Association of Behavioural and Cognitive Therapy, September 10–13, 2003, Prague, Czech Republic. I thank Kersten Schäfer for developing the statistical program, Thomas Bär for providing data, Rainer Riemann for providing feedback, and Deidre Winter and Matthew White for proofreading.

Correspondence concerning this article should be addressed to Kristin Mitte, Department of Psychology, University of Jena, Humboldtstr. 11, 07743, Jena, Germany. E-mail: mail@kristin-mitte.de

are computed. Analyses are therefore conducted to evaluate the effects of these decisions on mean effect sizes by changing the meta-analytic methods. That is, sensitivity analyses are performed to examine the robustness of results and conclusions. In the present meta-analysis, sensitivity analyses were conducted on, for example, modifications in methods of calculating effect sizes (including only those effect sizes calculated by means and standard deviations), modifications in the effect size distribution (including outliers), or modifications in the modes of comparing different treatment approaches (comparing [C]BT with all available drug studies vs. comparing [C]BT only with currently used drug classes, i.e., excluding benzodiazepines and any drug not marketed after completion of clinical trials). Fourth, publication bias, that is, a bias against studies with nonsignificant findings, was assessed and taken into account by using trim-and-fill analysis to compare (C)BT and pharmacotherapy. Trim-and-fill analysis (Duval & Tweedie, 2000) is used in meta-analysis to correct the funnel plot by estimating the number of missing studies and the effect sizes of these studies. Trim-and-fill analysis thus allows the calculation of mean effects corrected for publication bias. Fifth, several methodological differences between studies, such as the dropout rate or sample size, were controlled for when different therapy approaches were compared. And finally, a method was used that allows the comparison of studies with different control groups.

Method

An extensive literature search was conducted in various databases (MEDLINE and PsycINFO from the 1st available year to May 2002) using the search terms *general* anxiety*, *treatment*, and **therapy*. Additional titles were identified by a manual search in important journals and lists of references given in previous meta-analyses or primary studies. With a view to reducing file-drawer effects, I attempted to locate unpublished work using the Internet and by contacting researchers and pharmaceutical companies. However, I found only one unpublished study investigating the efficacy of pharmacotherapy (lesopitron). Suitable studies were selected for inclusion according to the following criteria.

1. Studies were published in either English or German.

2. All adult participants had been diagnosed with GAD according to a standardized diagnostic classification system (e.g., *DSM*), or an exact description of the disorder including the duration of symptoms was presented. It was considered insufficient if the patients were described as "anxious" or "neurotic." Studies with children and adolescents or older persons were not included.

3. Participants had received some form of (C)BT or pharmacotherapy. Behavioral therapy was defined as "direct attempts to reduce dysfunctional emotions and behavior by altering behavior" (Brewin, 1996, p. 34) and cognitive therapy as "attempts to reduce dysfunctional emotions and behavior by altering individual appraisals and thinking patterns" (Brewin, 1996, p. 34). Cognitive–behavioral therapy used both methods. Pharmacotherapy was conducted with a minimum duration of 14 days and with a designation of the used substance with the nonproprietary name.

4. No case studies were included.

5. Studies used an adequate control group including waiting lists, pill placebo, or therapy placebo. Therapy placebo was defined as the realization of factors common to all psychological therapies; studies using methods of behavioral or cognitive therapy, including relaxation, as the control condition were excluded. Studies investigating both cognitive–behavior therapy and pharmacological therapy were included.

6. No reports on the results of a subsample or double publications were included if they could be clearly assigned to a larger study.

7. Treatment outcome was established by self-report; observer-rated measures; or behavioral tests of anxiety, depression, quality of life, or clinical significance. The latter applied to anxiety only and included responder status, defined as a meaningful improvement (e.g., a 50% change on an assessment scale); end-state status, defined as a comparison with the normative population (e.g., no diagnosis); or both. Data were excluded if they could not be assigned to one of these categories.

8. Studies were required to give sufficient information to permit calculation of effect sizes, including means and standard deviations, t or F values, change scores, frequencies, and probability levels. If these indices were not reported and the study had been published within the past few years, I attempted to contact the authors. Some meta-analysts have included only studies that give means and standard deviations. However, it should be noted that excluding a study because no means and standard deviations are reported may result in a greater bias than if the study is included and the best available estimator is calculated. The reasons why these values have been excluded in the study may plausibly be directly linked to the result of the study (e.g., studies having failed to find a significant effect may tend to exclude means and standard deviations). Excluding such studies would therefore reduce the representativeness of the sample of studies selected for the meta-analysis and thus also the generalizability of its results.

No additional exclusion criteria assessing the quality of the studies were applied. However, most studies had used a randomized design (for two studies this variable was not codable, and one was nonrandomized). Nonrandomized studies were not excluded because the relationship between study quality and effect size is still a point at issue. In the pioneering meta-analysis conducted by Smith, Glass, and Miller (1980), there were no significant differences in effect sizes between randomized studies and nonequivalent control group designs. The same results were found by Lipsey and Wilson (1993) across 74 meta-analyses and Heinsman and Shadish (1996) across 98 studies after controlling for other methodological variables. In contrast, in a review of 100 studies of marital or family psychotherapy, Shadish and Ragsdale (1996) found a significantly greater effect size for randomized studies, even after controlling for covariates. Several authors therefore have recommended that the difference between randomized and nonrandomized studies be investigated in meta-analyses and taken into account when interpreting the findings (e.g., Hunter & Schmidt, 1990). However, the number of randomized studies included in the present meta-analysis was too small to evaluate this difference.

In the present research, a coding form consisting of items assessing methodological and clinical aspects of the studies was used; I completed it. To assess interrater reliability, I and three trained psychology students who were working on their diploma theses on therapy research independently coded 10 randomly selected studies. The coding form had an overall reliability of greater than .75 across all variables and individual item reliabilities of not less than .50 (according to Fleiss, 1981, this is excellent reliability). A total of 19 publications—13 studies comparing (C)BT with a control group and 6 studies comparing (C)BT with pharmacotherapy—met selection criteria for (C)BT trials, the data of 869 patients being included in the meta-analysis. The average age was 37.36 years, the average duration of GAD was 6.31 years, and 61.7% of the patients were women. Table 1 shows the effect sizes of the studies selected. It should be noted that some of the studies included in the meta-analysis by Gould et al. (1997) were not included in the current meta-analysis because of our strict exclusion criteria (e.g., Borkovec et al., 1987, in which the treatment of the control group consisted of relaxation, or Rice, Blanchard, & Purcell, 1993, in which some participants were diagnosed with subclinical GAD). The effect sizes of the (C)BT studies were compared with those of clinical drug trials of GAD (Mitte, Noack, Steil, & Hautzinger, 2005), so that a total of 65 studies with 7,739 patients were included in the analysis.

With regard to statistical analyses, computation of effect sizes was carried out using the standardized mean difference statistic Hedges's g between groups (Hedges's g is the difference between the means divided by the pooled standard deviations; Hedges & Olkin, 1985) or the algebraically equivalent effect sizes computed from the t and F values (Ray &

Table 1
Effect Sizes of Included Studies

Study	Groups	Anxiety effect size	Depression effect size
Barlow et al. (1984)	CBT, waiting list	2.24	—
Barlow et al. (1992)	BT, waiting list	0.93	0.54
	CT, waiting list	1.02	0.85
	CBT, waiting list	0.65	0.41
Biswas et al. (1995)	BT, benzodiazepines	−0.52	—
	CBT, benzodiazepines	0.04	—
Blowers et al. (1987)	Anxiety management training, waiting list	0.65	—
	Anxiety management training, therapy placebo	0.09	—
Borkovec & Costello (1993)	BT, therapy placebo	0.60	0.45
	CBT, therapy placebo	0.68	0.81
Bowman et al. (1997)[a]	Self-examination therapy, waiting list	0.95	—
Butler et al. (1987)	Anxiety management training, waiting list	1.00	1.00
Butler et al. (1991)	BT, waiting list	0.44	0.23
	CBT, waiting list	1.00	0.66
Cragan & Deffenbacher (1984)	Anxiety management training, waiting list	1.14	0.90
	BT, waiting list	1.24	0.73
Jannoun et al. (1982)	Anxiety management training, waiting list	1.07	0.87
Kohli et al. (2000)	Relaxation, pharmacotherapy	0.64	—
Ladouceur et al. (2000)	CBT, waiting list	1.39	1.48
Linden et al. (2002)	CBT, therapy placebo	0.48	—
Lindsay et al. (1987)	Anxiety management therapy, waiting list	0.73	0.42
	Anxiety management therapy, benzodiazepines	0.17	−0.09
	CBT, waiting list	1.27	1.65
	CBT, benzodiazepines	0.77	0.86
Power et al. (1989)	CBT, pill placebo	1.12	—
	CBT, benzodiazepines	0.85	—
Power et al. (1990)	CBT, pill placebo	1.22	0.51
	CBT, benzodiazepines	0.53	—
Sarkar et al. (1999)	BT, pharmacotherapy	−0.06	—
White et al. (1992)	BT, waiting list	0.50	0.83
	BT, therapy placebo	0.38	0.56
	CT, waiting list	0.55	1.07
	CT, therapy placebo	0.30	0.42
	CBT, waiting list	0.44	0.64
	CBT, therapy placebo	0.37	0.33
Woodward & Jones (1980)	BT, waiting list	0.46	—
	CT, waiting list	0.29	—
	CBT, waiting list	0.77	—

Note. CBT = cognitive–behavioral therapy; BT = behavioral therapy; CT = cognitive therapy without behavioral techniques.
[a] Study was excluded in the subsequent analyses because therapy was restricted to use of a booklet; in contrast, other studies in which a self-help treatment was investigated included longer lasting contact with the therapist.

Shadish, 1996). When the relevant means and standard deviations or *t* and *F* values were not reported, effect sizes were calculated by various other methods. Odds ratios were computed and transformed into Hedges's *g* on the basis of dichotomous data (e.g., responder status; Haddock, Rindskopf, & Shadish, 1998; Rosenthal, 1994). In cases in which change scores alone are reported, the correlation between pre- and posttest scores is required for computation of Hedges's *g*. However, most of the studies failed to report these correlations, and it was thus necessary to use an estimator of $r = .81$ (this is the mean retest reliability of 23 frequently used instruments for anxiety disorders). This estimator is higher than that used by Smith et al. (1980), that is, $r = .50$ for a period of 2 to 6 months. However, Ray and Shadish showed that if an estimator of $r = .50$ is used, a higher result is obtained for effect sizes based on change scores than for effect sizes based on means and standard deviations. Because the effect size is higher when computed from a lower pre–post correlation, Ray and Shadish postulated that Smith et al.'s (1980) estimator may well be too small.

In addition, if the studies included in the present meta-analysis did not publish the exact statistical results of analyses, the effect sizes were inferred from the *t* values of significance levels used to describe the results ("highly significant": $p = .01$; "significant": $p = .05$; "marginally significant": $p = .10$; "nonsignificant": $g = 0$). It should, however, be noted that effect sizes computed by this method may possibly underestimate the real effect (Ray & Shadish, 1996). An additional sensitivity analysis was therefore carried out to examine the impact of these effect sizes on the average effect size across the studies. Only effect sizes based on means and standard deviations were included in this sensitivity analysis.

In the case of the same results being presented with various statistical values (e.g., both means with standard deviations and *t* values), Hedge's *g* was computed using means and standard deviations. The correction for small sample bias was applied (Hedges & Olkin, 1985).

Where data were reported for both completer and intent-to-treat analyses, only the latter was included. Intent-to-treat analysis includes patients having dropped out during the course of the study, for example, because of lack of compliance, lack of treatment success, or the occurrence of serious side effects. Thus, results of an intent-to-treat analysis allow conclusions on the general applicability of a treatment and are regarded as more

meaningful than results of an analysis based on the data of treatment completers only. With a few exceptions, most studies investigating the efficacy of a psychopharmacological therapy used intent-to-treat analysis (of the studies included in this meta-analysis, nearly all [C]BT studies used completer analysis only, as compared with only one fourth of clinical drug trials). However, an intent-to-treat analysis could show lower effect sizes. This seems plausible, given that patients with severe symptomatologies were included—leading to increases in the means and variances of the symptoms. However, Mitte et al. (2005) found no significant differences between the results of completer and intent-to-treat analyses of trials of pharmacotherapy for GAD.

The direction of the effect sizes was standardized so that positive effect sizes always represented a better result for the treatment group (comparison of control and treatment group), for psychotherapy (comparison of pharmaco- and psychotherapy), and for a combination of therapies (comparison of psychotherapy and a combination of psycho- and pharmacotherapy).

Hedges's *g* was calculated separately for each assessment scale used in a study. Then, a mean *g* for the clinical variables was calculated by averaging across all of the dependent measures. Thus, each instrument was equally weighted. There are other approaches to the combination of stochastically dependent effect sizes (Gleser & Olkin, 1994); however, additional information (intercorrelations between the assessment scales) is required. Because these were not available, the more conservative method for computing the mean effect was used.

Some of the studies had used more than one treatment condition to investigate efficacy. In such cases, separate effect sizes were calculated wherever the various treatments represented different techniques. Otherwise, effect sizes were averaged across the treatment groups.

A random-effects analysis was then carried out to compute the mean effect sizes across all studies and the subsequent regression analyses (see, e.g., Erez, Bloom, & Wells, 1995; Hedges & Vevea, 1998; Overton, 1998). FEMs and REMs can be distinguished as follows. In the FEM, the variation between studies results only from the subjects included in the studies (within-study variance). All effect sizes are assumed to be estimates of a common population effect size; results cannot be generalized beyond the included studies. In contrast, the REM includes a variance component (between-studies or random-effects variance) that results from drawing studies from a universe of possible studies in addition to the variation due to the sampling of subjects in the original studies. The results of REM can therefore be generalized to treatment conditions not exactly resembling the conditions in the studies used for the data analyses. The REM takes into account several uncontrollable variables that could influence study effect sizes, such as therapist variables or setting. The REM was adopted both for the computation of the average effect sizes and for the regression analyses. Simulations show that interpretation of the results of the REM is only to be recommended when it is based on the data of at least 5 studies (Hedges & Vevea, 1998). When more than 20 studies are included, the performance of analyses (power of significance test, confidence intervals; CIs) is close to nominal (Hedges & Vevea, 1998). An FEM was therefore carried out in an additional sensitivity analysis to compute average effect sizes (Shadish & Haddock, 1994).

In order to test for variation in study effect sizes, I applied the *Q* test for homogeneity of effect sizes (Hedges, 1994). This test is used to determine whether study effect sizes are all equal or whether at least one effect size differs from the remainder. The *Q* statistic is also used to examine whether the average effect size computed with the FEM is representative of the population effect, which is a main assumption of the FEM. With the REM, the *Q* statistic tests whether the random-effects variance is significantly different from 0. Because the *Q* test has a low statistical power (Harwell, 1997), I tested for significance at a probability level of .10. When the *Q* test is nonsignificant, FEM and REM yield similar results for mean effect size.

In the computation of both the mean effect sizes and the regression analyses, the individual effect sizes of each study were weighted with the reciprocal of the variance components (i.e., with the within-study variance in the FEM and with the sum of the within-study and between-studies variance in the REM; Raudenbush, 1994; Shadish & Haddock, 1994). In view of the association between the sample size and the variance of an effect size, this method was used to ensure that the inaccuracy due to small sample size was taken into consideration. Studies with large sample sizes and thus more precise estimates of effect sizes have greater weights than studies with small sample sizes.

Results

Preliminary Analyses

In the first step of the analysis, a correction of the effect size distribution was performed. This was based on the work of Hunter and Schmidt (1990), who described several artifacts in meta-analyses, including the influence of bad data resulting from a variety of possible data handling errors (e.g., publishing erroneous data). Hunter and Schmidt drew attention to the difficulty of dealing with these artifacts when performing meta-analyses. They proposed excluding outlying values to reduce the impact of the faulty data on the meta-analysis. For the purposes of the present study, outliers were defined as those values that deviate more than two standard deviations from the unweighted mean, and appropriate correction was carried out across all studies included. Although it is not known whether outliers do in fact result from false data, this is the only way to control for this problem. For this analysis of the outliers, the effect sizes were calculated for each individual assessment scale and not averaged across groups. A total of 3.57% of the individual effect sizes for posttest measures were excluded from the outlier analysis comparing (C)BT and a no-treatment control, 1.47% from that comparing (C)BT and a common-factors control, and 3.45% from that comparing psychotherapy with pharmacotherapy. The impact of excluding outliers on the average effect size was investigated in a subsequent sensitivity analysis.

How Effective Is (C)BT?

The majority of the studies comparing (C)BT with a control group investigated a cognitive–behavioral approach, usually in conjunction with some type of exposure. The effect sizes were positive in all studies, that is, the control groups never achieved a better result than the treatment groups. Table 2 shows the average weighted effect sizes, 95% CIs, random-effects variances, and homogeneity statistics for each symptom category. The results of both the REM and the FEM are presented. The results of the REM can be generalized beyond the studies included in the review; however, when only a small number of studies are included, the results of the REM may be biased. As shown in Table 2, (C)BT yielded significant medium-to-large effect sizes in comparison with both a waiting list control and a common-factors control (both psychological and pill placebo). The efficacy of (C)BT therefore exceeds the realization of common factors. This result holds true not only for the main symptom of anxiety but also for associated depressive symptoms and quality of life. The zero random-effects variance indicates that it was not necessary to conduct a regression analysis to investigate the impact of moderator variables. For the same reason, the two statistical models (FEM and REM) yielded

Table 2
Average Effect Sizes for Posttest of the Symptom Categories Comparing Therapy Approaches

Comparison	Symptom category	Random effects		Fixed effects		τ^2	Q	n
		Average effect size	95% CI	Average effect size	95% CI			
(C)BT with no-treatment control	Anxiety	0.82	0.62, 1.01	0.82	0.63, 1.00	0	10.91	19
	Depression	0.76	0.55, 0.98	0.76	0.57, 0.96	0	8.69	15
	Quality of life	0.89	0.57, 1.21	0.89	0.47, 1.31	0	4.04	6
(C)BT with placebo control	Anxiety	0.57	0.30, 0.85	0.57	0.34, 0.80	0	7.28	9
	Depression	0.52	0.15, 0.89	0.52	0.24, 0.80	0	1.06	6
	Clinical significance	0.98	0.38, 1.57	0.98	0.61, 1.35	0	2.80	4
(C)BT with pharmacotherapy	Anxiety	0.33	−0.02, 0.67	0.33	0.04, 0.61	0	6.80	8

Note. CI = confidence interval; τ^2 = random-effects variance; Q = result of the Q test for homogeneity of effect sizes; n = number of effect sizes; (C)BT = (cognitive) behavioral therapy.

equal effect sizes. The CIs for the REM are wider, despite a zero random-effects variance, because the number of studies is used to compute the CIs, and not the number of subjects, as in the FEM.

In a first sensitivity analysis, all effect sizes from studies that did not provide means and standard deviations were excluded. This step was based on findings described by Ray and Shadish (1996), who reported notable differences between various computational variants of Hedges's g. They found, for example, that mean effect sizes and variances based on means and standard deviations tend to be larger than those based on probability levels and concluded that researchers should pay more attention to this in their analyses. In the sensitivity analysis conducted in the present meta-analysis, in which only effect sizes computed with means and standard deviations were included, results for the main category anxiety were robust for the comparison of (C)BT with a waiting list ($g = 0.81$, 95% CI = 0.59, 1.03, $\tau^2 = 0$; FEM: $g = 0.81$, 95% CI = 0.61, 1.01), $\chi^2(16, N = 17) = 10.38$, and that of (C)BT with a common-factors control ($g = 0.61$, 95% CI = 0.32, 0.90, $\tau^2 = 0$; FEM: $g = 0.61$, 95% CI = 0.37, 0.85), $\chi^2(7, N = 8) = 5.94$.

In a second sensitivity analysis, effect sizes defined as outliers were also included, that is, no effect size was excluded. As expected, for the comparison with waiting list controls, the mean effect size for the main category of anxiety was slightly higher than in the analysis excluding outliers ($g = 0.87$, 95% CI = 0.67, 1.07, $\tau^2 = 0$; FEM: $g = 0.87$, 95% CI = 0.68, 1.05), $\chi^2(19, N = 20) = 12.39$. However, the differences between the mean effect sizes found with these two methods were only minor. There were no outliers for the main category of anxiety in the comparison with a common-factors control.

It should be noted that the common-factors control in the primary analysis also included control groups using pill placebo. There has been some discussion as to whether pill placebo and psychological placebo are comparable controls (e.g., Crits-Christoph, 1997; Heimberg, 1997; Klein, 1997). To address this issue, I conducted a third sensitivity analysis, which excluded all effect sizes from studies in which (C)BT was compared with a pill placebo. The resulting average effect size was lower than in the primary analysis ($g = 0.44$, 95% CI = 0.13, 0.76, $\tau^2 = 0$; FEM: $g = 0.44$, 95% CI = 0.19, 0.70), $\chi^2(6, N = 7) = 1.60$, but the CIs overlap, indicating that the difference was not significant.

Comparison of (C)BT and Pharmacotherapy I

Effect sizes were computed from studies directly comparing both conditions. Six studies compared (C)BT and pharmacotherapy, which in most studies consisted of the administration of benzodiazepines. The average effect sizes are presented in Table 2. The REM showed no significant difference between these therapy approaches. On the basis solely of the studies included in the analysis, (C)BT proved to be more effective than pharmacotherapy in treating GAD; however, this result cannot be generalized beyond these studies, because it was based on a fixed-effects analysis. Again, the CIs obtained with the REM were wider. Although some trials showed greater efficacy for pharmacotherapy (smallest $g = -0.52$), others showed (C)BT to be more effective (highest $g = 0.85$), but the random-effects variance between effect sizes was 0 and thus not important.

However, the results of the first sensitivity analysis, from which all effect sizes not computed by means and standard deviations had been excluded, did not yield the same results as the primary analysis ($g = 0.28$, 95% CI = −0.10, 0.66, $\tau^2 = 0$; FEM: $g = 0.28$, 95% CI = −0.02, 0.59), $\chi^2(6, N = 7) = 6.15$. That is, in this first sensitivity analysis neither REM nor FEM revealed significant differences between the approaches.

The second sensitivity analysis also included effect sizes that were defined as outliers. No difference was found between the symptom category of anxiety in this sensitivity analysis and that in the primary analysis ($g = 0.31$, 95% CI = −0.05, 0.68, $\tau^2 = 0.01$; FEM: $g = 0.31$, 95% CI = 0.02, 0.60), $\chi^2(7, N = 8) = 7.49$.

Only one study provided data on the efficacy of psychotherapy compared with a combination of psychosocial therapy and pharmacotherapy. An effect size of 0.46 (for the combination with diazepam) was found for the anxiety measures.

Comparison of (C)BT and Pharmacotherapy II

A second strategy for comparing (C)BT and pharmacotherapy is to compare the results of the studies in which the two conditions were tested against a control group (i.e., to compare studies investigating (C)BT and waiting list or placebo control with studies investigating pharmacotherapy and pill placebo). However, the problem with this approach is that the control groups differ in

efficacy. Because, for instance, common factors (e.g., expectations and hope) are realized in pill placebo and not in the waiting list condition, pill placebo proves more effective than a waiting list. As a consequence of this difference between control groups, given, for example, the hypothetical case that both (C)BT and pharmacotherapy were equally effective, the effect sizes for pharmacotherapy compared with pill placebo would be lower than those for (C)BT when compared with the waiting list.

Direct comparison of the effect sizes from studies with different types of control groups is therefore not appropriate, and a method for adjusting the effect sizes to render comparison possible is required. I have adopted and modified a method based on procedures suggested by Becker (1988) and Kirsch and Sapirstein (1998). First, the within-group, pretest-to-posttest effect sizes of the control groups were computed. Then, using weighted regression analysis, the difference between these within-group effect sizes of the different control groups was computed as a measure of the differing "efficacies" of the control groups. In a third step, the effect sizes obtained from the comparison of (C)BT with the waiting list controls were adjusted by subtracting the differences between the pretest–posttest effect sizes of the waiting list and pill placebo groups from the between-groups effect sizes. The appropriateness of this method has been verified.[1] An additional advantage of this method is that treatment effects can be differentiated into common factors and specific treatment effects, even for pharmacotherapy. It should be noted that the results of these analyses are only interpretable if the same retest effect is assumed both for studies investigating (C)BT and for those investigating pharmacotherapy. This is a basic assumption underlying all pre–post comparisons.

Effect sizes obtained for comparisons between (C)BT and a control group in the current meta-analysis were then compared with the effect sizes from a meta-analysis on the efficacy of pharmacological therapy of GAD (Mitte et al., 2005). This meta-analysis integrated effect sizes from 48 studies comparing pharmacotherapy with pill placebo (including 2 studies from the aforementioned analyses in which [C]BT, pharmacotherapy, and pill placebo were given). The most frequently used drug classes were benzodiazepines and azapirones. Mitte et al. (2005) found a mean effect size across all drugs of 0.31.

The different efficacies of the control groups were evaluated by calculating the pre–post effect sizes. An average difference in pre–post effect sizes of 0.94 was found for measures of anxiety, and an effect size of 0.85 was found for measures of depression, with larger effect sizes for pill placebo compared with no-treatment control. However, it should be noted that only a small number of drug trials reported sufficient data for computation (ns = 20 and 7, respectively). No comparison was carried out between pill placebo and psychological placebo used in studies using (C)BT, owing to insufficient data. After adjustment of the individual effect sizes, a single weighted regression analysis showed the effect sizes for pharmacotherapy for both anxiety and depression to be higher than for (C)BT ($B = -0.31$, $p < .01$, $n = 92$; $B = -0.34$, $p < .01$, $n = 42$). Regression coefficients are unstandardized and are the amount of change in effect size associated with 1 unit of change in the predictor.

To assess whether the results of these analyses were influenced by different methodological variables associated with different therapy approaches, a multiple weighted regression analysis was computed (using the method of moments estimators; Raudenbush, 1994) controlling for dropout rate, sample size, and diagnostic system. Studies whose dropout rates were not reported and could not be estimated were excluded from the regression analysis. It was assumed that two reasons related to the dropout rate could be responsible for missing information: (a) No patient dropped out, or (b) numerous patients dropped out. An association between the dropout rate and the probability of missing information was therefore assumed, and thus no estimator was used in place of missing information.

If the exact diagnostic system used in a study was not stated, it was assumed that the version of the *DSM* in use 2 years before publication (taking into account the time between performance and publication of a study) had been used.

Again, drug therapy was more effective than (C)BT for anxiety and depression ($B = -0.43$, $p < .01$, $n = 82$; $B = -0.54$, $p < .01$, $n = 39$), whereas the other three variables had no significant impact on effect size. To explore these results further, less common treatments were excluded. This applied to any drug that was not marketed after completion of the trials and all benzodiazepines, given their reduced use in recent years. Thus the azapirones, for example, buspirone, were the most investigated drug class in this analysis. After these exclusions, the differences between therapeutic approaches remained constant for symptoms of anxiety and increased for depression ($B = -0.43$, $p < .01$, $n = 56$; $B = -0.75$, $p < .05$, $n = 27$).

Attrition

Acceptance of the respective types of therapy was investigated by analyzing the dropout rates. It is, of course, possible that the reasons why some participants dropped out were completely independent of factors pertaining to the therapy, for example, a patient moved to another area. However, because such individual dropout reasons were not reported in all studies, the total dropout rate was explored.

The average dropout rate for (C)BT was 9.0% ($SD = 10.43$, $n = 25$). The treatment is thus well tolerated. Across all drugs, Mitte et al. (2005) found that approximately one fourth of participants dropped out. For computation of the average dropout rate for pharmacotherapy, both kinds of studies were included, that is, studies in which pharmacotherapy was compared with (C)BT and studies in which pharmacotherapy was compared with pill placebo. Comparison of psychotherapy and pharmacotherapy revealed a significantly lower dropout rate for (C)BT, $t(64.8) = 5.57$, $p < .01$. Because differences in dropout rate could result in differences in treatment efficacy, the reader is reminded that in earlier com-

[1] Across 100 randomly selected cases with both pre- and posttest means and standard deviations available for all anxiety disorders, no difference was found between (a) effect sizes calculated between groups at posttest and (b) effect sizes calculated using the differences between the pre–post effect sizes of the two groups. The mean effect size for (a) was 0.28 ($\sigma^2 = 0.007$, 95% CI = 0.11, 0.43), and for (b) it was 0.31 ($\sigma^2 = 0.011$, 95% CI = 0.10, 0.52); the difference was not significant ($B = 0.03$, $R^2 = .00$). The correlation between the two methods was .79, and the average square error was 0.37. The method performed well. The mean and variance estimates of the two methods were quite similar.

parisons of (C)BT and pharmacotherapy across separate studies I controlled for the dropout rate.

Publication Bias

Results of meta-analyses may be subject to publication bias. Publication bias means that the publication of a study depends on its results, that is, studies with nonsignificant results are less likely be published than those leading to significant results (the file drawer problem). Results of a meta-analysis are then biased in a positive direction. Because power and sample size are associated and the variance of an effect size is inversely related to sample size, publication bias is investigated by estimating the associations between effect sizes and their variances. I performed the rank correlation test as a formal test of funnel plot asymmetry to examine the hypothesis that no publication bias exists[2] (Begg, 1994), selecting a probability level of 10%. No significant associations between effect sizes and their variances were found for the comparison of (C)BT with a waiting list ($Z = 0.31$), the comparison of (C)BT and a placebo control ($Z = 0.74$), or for the comparison of (C)BT with pharmacotherapy from studies in which both conditions were investigated ($Z = 0.25$). There was therefore no evidence of a publication bias. However, it should be noted that the power of the rank correlation test is low when only a small number of studies are included, a problem inherent in all methods for the detection of publication bias.

In contrast, Mitte et al. (2005) found a publication bias for the studies on pharmacotherapy included in this review, which suggests that there are a meaningful number of unpublished studies on pharmacotherapy with nonsignificant results. To investigate the influence of this bias on the comparison of (C)BT and pharmacotherapy, I conducted a trim-and-fill analysis (Duval & Tweedie, 2000). In trim-and-fill analysis, the number of missing (unpublished) studies is estimated and the effect size distribution is adjusted in such a way that the funnel plot is rendered symmetrical. The "true" mean effect size is then estimated with these new effect sizes. When these adjusted effect sizes were used for pharmacotherapy, the difference between (C)BT and pharmacotherapy decreased but was still significant in a single regression analysis of studies on anxiety and depression ($B = -0.27$, $p < .01$, $n = 92$; $B = -0.30$, $p < .01$, $n = 42$) as well as in a multiple regression analysis of studies on anxiety that controlled for sample size, diagnostic system, and dropout rate ($B = -0.33$, $p < .01$, $n = 82$). In contrast, in a multiple regression analysis of effect sizes for depression, no significant difference between (C)BT and pharmacotherapy was found after adjusting for publication bias ($B = -0.18$, $n = 39$). In both analyses four effect sizes were recomputed. It should be noted that the aim of trim-and-fill analysis is to obtain a symmetrical funnel plot. In the current analysis of findings for anxiety, the new effect sizes were thus actually negative rather than nonsignificant (down to $g = -0.94$), with the result that a total of 40% of the drug trials had negative or small effect sizes (up to $g = 0.20$).

Common Factors and Specific Treatment Effects

If one assumes an additive model of retest effects, effects of common factors, and specific treatment effects, changes in symptoms are caused by a retest effect in the case of the waiting list, by a retest effect plus common factors for therapy placebo, and by all three components for the treatment group. The effect sizes found for the comparisons of (C)BT with a no-treatment control ($g = 0.82$; cf. Table 2) and a placebo group ($g = 0.57$; cf. Table 2) suggest a relatively large specific treatment effect, explaining 70% of the symptomatic change of patients with GAD. For pharmacotherapy, there was no study comparing the effects of the drug under investigation with those of a waiting list. To estimate the retest effect, I therefore used the aforementioned regression analysis to compare the pre–post effect sizes for pill placebo (from drug trials) with pre–post effect sizes for waiting lists (from [C]BT studies). The effect sizes for drug versus pill placebo ($g = 0.31$; Mitte et al., 2005) and pill placebo versus waiting list conditions ($g = 0.94$) indicate that the factors responsible for the efficacy of pharmacotherapy can be broken down into 75% common factors and 25% specific factors.

Discussion

The results of the present meta-analysis indicate that (C)BT is a highly effective treatment of GAD, reducing not only the main symptoms of anxiety but also the associated depressive symptoms and subsequently improving quality of life. (C)BT was not only more effective than a waiting list but also a common-factors control. The results suggest that the characteristics of an individual study—for example, patient variables, method, or therapy techniques—have no important impact on the efficacy of the treatment. However, the stringent exclusion criteria used to select the studies for the meta-analysis could have resulted in a decrease in variance and thus have led to the lack of differences detected. In addition, it should be noted that recent developments in the treatment of GAD were not included in the meta-analysis because no study investigating them was found that met the inclusion criteria. Examples of such developments could be the inclusion of a metacognitive perspective (e.g., Wells, 2000), mindfulness approaches (e.g., Roemer & Orsillo, 2002), or interpersonal therapy (see also Borkovec, Newman, Pincus, & Lytle, 2002). Furthermore, only a few studies used appropriate instruments to assess quality of life. Future studies should address this aspect.

Although the current results provide support for the efficacy of (C)BT in the treatment of GAD, the effect sizes found may overestimate the real effect. The main problem seems to be the substantial use of completer analysis, that is, the data of patients who dropped out were excluded. In contrast, an intent-to-treat analysis could result in lower effect sizes because patients with more severe symptomatologies would be included. The results of a regression analysis of pharmacotherapy studies (Mitte et al., 2005), which failed to indicate any differences between the analyses, are difficult to generalize to (C)BT. One possible explanation of the pharmacological findings is that differences between the dropout rates for drug and placebo groups were not taken into consideration. However, if more patients withdraw from the placebo group, a higher effect size can be expected for the intent-to-treat analysis. This assumption is supported by findings reported by Gould, Otto, and Pollack (1995), who derived their observa-

[2] A simple funnel plot analysis was not used because (a) the number of studies was low and (b) this method relies on subjective interpretation.

tions from four studies. Differences between intent-to-treat analysis and completer analysis may therefore be diminished because there are drug trials with higher dropout rates for pill placebo as well as studies with higher dropout rates for the active substance class. In contrast, in studies on (C)BT the dropout rate for the treatment group is usually higher than that for the no-treatment control. Differences between intent-to-treat and completer analyses may therefore be expected for (C)BT.

The results of the comparison between (C)BT and pharmacotherapy varied according to the computational methods used in the analysis. When only studies that had investigated both treatment approaches were included, there was a significantly greater effect for (C)BT. However, this finding cannot be generalized beyond the studies included in the analysis and in fact disappeared when the study population was modified in the sensitivity analyses. It would therefore seem that (C)BT is at least as effective as pharmacotherapy. The advantages of this method, in which therapy approaches are compared directly with each other, are that methodological variables are controlled for and no additional assumptions are necessary. However, it should be noted that in most cases (C)BT was compared with benzodiazepines. It is well-known that there is an increasing risk of physical dependence and withdrawal after long-term treatment with benzodiazepines, as several authors have already noted in their reviews (e.g., Hollister, Müller-Oerlinghausen, Rickels, & Shader, 1993; Woods, Katz, & Winger, 1992). In addition, other severe adverse effects such as memory impairment can result from long-term use of benzodiazepines (e.g., Woods et al., 1992). Use of this class of drugs is therefore limited. Future research is needed to evaluate the efficacy of (C)BT and currently used drug classes in studies that compare the two approaches directly.

In contrast, an indirect method of comparing the two therapy approaches, that is, looking at effect sizes found by studies comparing (C)BT with a waiting list control and pharmacotherapy with a pill placebo control, showed pharmacotherapy to be superior to (C)BT. The primary analysis included all available studies and therefore also classes of drugs that are not used in practice. Hence, in a second comparison, I contrasted (C)BT only with drugs currently used, such as azapirones and SSRIs. The result again shows a difference between (C)BT and pharmacotherapy. In interpreting this result, it is important to emphasize that although it is less likely to be influenced by potentially unpublished studies on (C)BT, the possibility that unpublished drug trials have an impact cannot be ruled out. To overcome this problem, I used trim-and-fill analysis and found that the differences between (C)BT and pharmacotherapy decreased for depressive symptoms, but not for anxiety. Thus, the limited availability of (nonsignificant) studies of pharmacotherapy is indeed an important factor to consider. Nevertheless, none of the statistical methods used in attempts to correct publication bias can solve the substantial problem that published studies, and particularly those on pharmacotherapy, are not representative of evidence from research as a whole. It should be emphasized that the estimators calculated by these methods used to deal with publication bias are not the true effect sizes of a population. Thus, in the face of (systematic) suppression of findings, the inherent limitation of meta-analyses is revealed. Here, the clinical trial registry, supported, for example, by the American Medical Association,[3] will increase transparency of research and therefore validity of research synthesis.

What are the strengths and limitations of the method of indirect comparison? An advantage of this method is that it increases the number of studies included in the meta-analysis. Only 6 studies on GAD were found in which (C)BT was compared with pharmacotherapy (mainly benzodiazepines). Clearly, randomized studies comparing the two therapy approaches directly are to be preferred. However, currently too few head-to-head studies have been carried out. Only when therapy approaches are investigated in various studies with different samples, methods, and researchers can the results be generalizable. Indirect comparison of different types of therapy may thus be an alternative. Furthermore, the impact of the allegiance effect on results is reduced. Results of studies in which (C)BT and pharmacotherapy are compared directly may be biased by researchers who are convinced of the efficacy of one of the forms of therapy.[4]

On the other hand, several limitations were found in the studies included in the meta-analysis that influence the results. A disadvantage of this method compared with the method of direct comparisons is that not all methodological variables are controlled across therapy conditions. Differences in effect sizes between forms of therapy could be due to confounding variables. In addition, there is no means of checking whether the underlying assumption of equal retest effects (defined here as all changes between two occasions without effects of treatment or common factors) is indeed correct.

In reviewing the studies selected, it is clear that there are many differences in methodological variables between the psychological and pharmacological studies that could have an influence on the effect size. One factor associated with effect sizes is the measures used. It is well-known that different measures assess change with differing levels of sensitivity. For example, in studies on the treatment of depression (see Lambert & Hill, 1994), it was found that the Hamilton Depression Scale (HAMD; Hamilton, 1960) tends to yield larger effect sizes than the Beck Depression Inventory (BDI; Beck, Epstein, Brown, & Steer, 1988). Approximately half of the (C)BT studies included in the current review used the BDI, whereas none of the pharmacotherapy trials used this scale, the majority having used the HAMD. Hence, even if pharmacotherapy and (C)BT are equally effective in the treatment of GAD, studies using the BDI (i.e., studies in which [C]BT was used) will show a lower effect size. The possibility cannot be ruled out that assessment scales used to measure anxiety symptoms also have different sensitivity levels.

Furthermore, various aspects of anxiety were assessed with different weights in studies on the different types of treatment. The majority of the drug trials used the Hamilton Anxiety Scale (HAMA; Hamilton, 1959). Although the HAMA assesses both

[3] I thank an anonymous reviewer for this helpful comment.

[4] An anonymous reviewer noted that there should be no systematic effect of allegiance across studies because it could be exerted in either direction (pharmacotherapy or [C]BT). I acknowledge that this is an important point. However, it depends on the number of available studies. When only a small number of studies are included, a systematic error may occur. In a meta-analysis on depression, Gaffan, Tsaousis, and Kemp-Wheeler (1995) found a substantial allegiance effect for cognitive therapy compared with other treatment approaches (other psychosocial treatments and pharmacotherapy) and an important impact of this effect on effect sizes.

somatic and cognitive features of GAD, the latter aspect is incorporated in only a few items, fewer than would be required to establish fulfillment of *DSM–III–R* or *DSM–IV* criteria. On the other hand, the (C)BT studies frequently used measures of cognition and worry such as the Penn State Worry Questionnaire (Meyer, Miller, Metzger, & Borkovec, 1990). Thus, results could plausibly have been influenced by the "apples and oranges" problem: Effect sizes may have been compared that were based on different concepts.

The characteristics of the patients seem to differ less between the (C)BT and drug studies included in the current meta-analysis, which is partly due to the inclusion criteria used. However, it should be noted that approximately two thirds of the drug studies included did not state whether the diagnosis was made by structured interview.

Although it remains an open question which treatment is most effective for GAD, it does seem clear that (C)BT is better tolerated than pharmacotherapy. Although psychotherapeutic approaches to treating GAD have attracted relatively little research compared with other anxiety disorders, (C)BT seems a valuable alternative to psychotropic drugs. The current results highlight the need for more studies using a direct comparison of different forms of therapy.

In addition, the impact of the common and specific factors of the different types of treatment proved to be inconsistent across (C)BT and pharmacotherapy trials. For (C)BT there was a high impact of specific treatment factors. Thus, patients with GAD appear to profit less from the common factors of a therapy. In a meta-analysis in which the magnitude of common and specific factors across all disorders was computed directly (i.e., only studies with a treatment group, a common-factors control, and a no-treatment control were included), Stevens, Hynan, and Allen (2000) found that the impact of common factors is lower for less circumscribed disorders. The current results support their findings. In contrast, I found a strong impact of common factors on the efficacy of pharmacotherapy.

A potential explanation for this finding deserves consideration. Waiting list, pill placebo, and therapy conditions are needed to compute common and specific factors; however, no studies were included in which pill placebo and waiting list conditions were directly compared, and too few studies allowed a comparison of pre–post effect sizes. Because none of the drug trials included a waiting list, the pre–post effect sizes for no-treatment control came from (C)BT studies. As noted earlier, is difficult to compare effect sizes from different types of treatment ([C]BT and pharmacotherapy). It may therefore be that the results are confounded with methodological variables such as different assessment instruments and that the additional assumption of equal retest effects cannot be evaluated. Thus, the estimation of common factors and specific treatment effects could be biased for pharmacotherapy.

The results of meta-analyses are only as reliable and valid as the studies on which they depend. Because only a few studies performed direct comparisons of the two forms of therapy directly, it is not possible to give generalizable results. I have also found some evidence that there is a bias for studies on generalized anxiety disorder that may, in addition to the publication bias for pharmacological studies, have contributed to the differences found between (C)BT and pharmacotherapy in the indirect comparison. It is therefore possible that the superiority of pharmacotherapy found in this analysis may result from factors other than therapy itself, such as methodological variables of the studies analyzed. Future research would benefit from application of the same methodological standards in psychosocial and pharmacological studies. Such standards should, for example, include using structured interviews to diagnose patients and using the same measures to assess GAD and associated symptoms, as well as using measures of quality of life.

References

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

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.

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

*Barlow, D. H., Cohen, A. S., Waddell, M. T., Vermilyea, B. B., Klosko, J. S., Blanchard, E. B., & DiNardo, P. A. (1984). Panic and generalized anxiety disorder: Nature and treatment. *Behavior Therapy, 15,* 431–449.

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

Beck, A. T., Epstein, N., Brown, G., & Steer, R. A. (1988). An inventory for measuring clinical anxiety. *Journal of Consulting and Clinical Psychology, 56,* 893–897.

Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology, 41,* 257–278.

Begg, C. B. (1994). Publication bias. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 399–410). New York: Russell Sage Foundation.

*Biswas, A., Biswas, D., & Chattopadhyay, P. K. (1995). Cognitive behaviour therapy in generalised anxiety disorder. *Indian Journal of Clinical Psychology, 22,* 1–10.

*Blowers, C., Cobb, J., & Mathews, A. (1987). Generalised anxiety: A controlled treatment study. *Behaviour Research and Therapy, 25,* 493–502.

*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., Mathews, A. M., Chambers, A., Ebrahimi, S., Lytle, R., & Nelson, R. (1987). The effects of relaxation training with cognitive or nondirective therapy and the role of relaxation-induced anxiety in the treatment of generalized anxiety. *Journal of Consulting and Clinical Psychology, 55,* 883–888.

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.

*Bowman, D., Scogin, F., Floyd, M., Patton, E., & Gist, L. (1997). Efficacy of self-examination therapy in the treatment of generalized anxiety disorder. *Journal of Counseling Psychology, 44,* 267–273.

Brewin, C. R. (1996). Theoretical foundations of cognitive–behavior therapy for anxiety and depression. *Annual Review of Psychology, 47,* 33–57.

*Butler, G., Cullington, A., Hibbert, G., Klimes, I., & Gelder, M. (1987). Anxiety management for persistent generalised anxiety. *British Journal of Psychiatry, 151,* 535–542.

*Butler, G., Fennell, M., Robson, P., & Gelder, M. (1991). Comparison of behavior therapy and cognitive behavior therapy in the treatment of generalized anxiety disorder. *Journal of Consulting and Clinical Psychology, 59,* 167–175.

*Cragan, M. K., & Deffenbacher, J. L. (1984). Anxiety management training and relaxation as self-control in the treatment of generalized anxiety in medical outpatients. *Journal of Counseling Psychology, 31,* 123–131.

Crits-Christoph, P. (1997, September 22). Control groups in psychotherapy research revisited. *Treatment, 1,* Comment 1. Available from http://journals.apa.org/prevention/volume1/97_c1-97_a1.html

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 50,* 455–463.

Erez, A., Bloom, M. C., & Wells, M. T. (1995). *On a proper meta-analytic model for correlations* (Working Paper No. 95-11). Ithaca, NY: Cornell University.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions* (2nd ed.). New York: Wiley.

Gaffan, E. A., Tsaousis, J., & Kemp-Wheeler, S. M. (1995). Researcher allegiance and meta-analysis: The case of cognitive therapy for depression. *Journal of Consulting and Clinical Psychology, 63,* 966–980.

Gleser, L. J., & Olkin, I. (1994). Stochastically dependent effect sizes. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 340–355). New York: Russell Sage Foundation.

Gould, R. A., Otto, M. W., & Pollack, M. H. (1995). A meta-analysis of treatment outcome for panic disorder. *Clinical Psychology Review, 15,* 819–844.

Gould, R. A., Otto, M. W., Pollack, M. H., & Yap, L. (1997). Cognitive behavioral and pharmacological treatment of generalized anxiety disorder: A preliminary meta-analysis. *Behavior Therapy, 28,* 285–305.

Haddock, C. K., Rindskopf, D., & Shadish, W. R. (1998). Using odds ratio as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods, 3,* 339–353.

Hamilton, M. C. (1959). The assessment of anxiety states by rating. *British Journal of Medical Psychology, 32,* 50–55.

Hamilton, M. C. (1960). A rating scale for depression. *Journal of Neurology, Neurosurgery & Psychiatry, 23,* 56–62.

Harwell, M. (1997). An empirical study of Hedges's homogeneity test. *Psychological Methods, 2,* 219–231.

Hedges, L. V. (1994). Fixed effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 285–299). New York: Russell Sage Foundation.

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

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3,* 486–504.

Heimberg, R. G. (1997, September 22). The use of pill placebo as a control strategy in the evaluation of psychotherapy: Reply to Klein. *Treatment, 1,* Comment 2. Available from http://journals.apa.org/prevention/volume1/97_c2-97_a1.html

Heinsman, D. T., & Shadish, W. R. (1996). Assignment methods in experimentation: When do nonrandomized experiments approximate answers from randomized experiments? *Psychological Methods, 1,* 154–169.

Hollister, L. E., Müller-Oerlinghausen, B., Rickels, K., & Shader, R. I. (1993). Clinical uses of benzodiazepines. *Journal of Clinical Psychopharmacology, 13,* 1S–169S.

Hunter, J. E. & Schmidt, F. L. (1990). *Methods of meta-analysis.* Newbury Park, CA: Sage.

*Jannoun, L., Oppenheimer, C., & Gelder, M. (1982). A self-help treatment program for anxiety state patients. *Behavior Therapy, 13,* 103–111.

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., et al. (1994). Lifetime and 12-month prevalence of *DSM–III–R* psychiatric disorders in the United States: Results from the National Comorbidity Survey. *Archives of General Psychiatry, 51,* 8–19.

Kessler, R. C., Mickelson, K. D., Barber, C., & Wang, P. (2001). The association between chronic medical conditions and work impairment. In A. S. Rossi (Ed.), *Caring and doing for others: Social responsibility in the domain of the family, work, and community* (pp. 403–426). Chicago: University of Chicago Press.

Kirsch, I., & Sapirstein, G. (1998). Listening to Prozac but hearing placebo: A meta-analysis of antidepressant medication. *Prevention & Treatment, 1,* Article 0002a. Retrieved May 18, 2001, from http://journals.apa.org/prevention/volume1/pre0010002a.html

Klein, D. F. (1997, September 22). Control groups in pharmacotherapy and psychotherapy evaluations. *Treatment, 1,* Article 1. Available from http://journals.apa.org/prevention/volume1/97_a1.html

*Kohli, A., Nehra, V., & Nehra, R. (2000). Comparison of efficacy of psychorelaxation and pharmacotherapy in generalized anxiety disorder. *Journal of Personality and Clinical Studies, 16,* 43–48.

*Ladouceur, R., Dugas, M. J., Freeston, M. H., Léger, E., Gagnon, F., & Thibodeau, N. (2000). Efficacy of cognitive–behavioral treatment of generalized anxiety disorder: Evaluation in a controlled clinical trial. *Journal of Consulting and Clinical Psychology, 68,* 957–964.

Lambert, M. J., & Hill, C. (1994). Assessing psychotherapy outcomes and processes. In A. E. Bergin & S. L. Garfield (Eds.), *Handbook of psychotherapy and behaviour change* (pp. 72–113). New York: Wiley.

*Linden, M., Bär, T., Zubrägel, D., Ahrens, B., & Schlattmann, P. (2002). Wirksamkeit der kognitiven Verhaltenstherapie bei generalisierten Angsterkrankungen. Ergebnisse der Berliner KVT-GAD-Studie [Effectiveness of cognitive behavior therapy in the treatment of generalized anxiety disorders—Results of the Berlin GAD-KVT Study]. *Verhaltenstherapie, 12,* 173–181.

*Lindsay, W. R., Gamsu, C. V., McLaughlin, E., Hood, E. M., & Espie, C. A. (1987). A controlled trial of treatments for generalized anxiety. *British Journal of Clinical Psychology, 26,* 3–15.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment. *American Psychologist, 18,* 1181–1209.

Meyer, T. J., Miller, M. L., Metzger, R. L., & Borkovec, T. D. (1990). Development and validation of the Penn State Worry Questionnaire. *Behaviour Research and Therapy, 28,* 487–495.

Mitte, K., Noack, P., Steil, R., & Hautzinger, M. (2005). A meta-analytic review of the efficacy of drug treatment in generalized anxiety disorder. *Journal of Clinical Psychopharmacology, 25,* 141–150.

Overton, R. C. (1998). A comparison of fixed-effects and mixed (random effects) models for meta-analysis tests of moderator variable effects. *Psychological Methods, 3,* 354–379.

*Power, K. G., Jerrom, D. W. A., Simpson, R. J., Mitchell, M. J., & Swanson, V. (1989). A controlled comparison of cognitive–behaviour therapy, diazepam and placebo in the treatment of generalized anxiety. *Behavioural Psychotherapy, 17,* 1–14.

*Power, K. G., Simpson, R. J., Swanson, V., & Wallace, L. A. (1990). A controlled comparison of cognitive–behaviour therapy, diazepam, and placebo, alone and in combination, for the treatment of generalised anxiety disorder. *Journal of Anxiety Disorders, 4,* 267–292.

Raudenbush, S. W. (1994). Random effects models. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 301–321). New York: Russell Sage Foundation.

Ray, J. W., & Shadish, W. R. (1996). How interchangeable are different estimators of effect size? *Journal of Consulting and Clinical Psychology, 64,* 1316–1325.

Rice, K. M., Blanchard, E. B., & Purcell, M. (1993). Biofeedback treatments for generalized anxiety disorder: Preliminary results. *Biofeedback and Self-Regulation, 18,* 93–105.

Rickels, K., & Rynn, M. A. (2001). What is generalized anxiety disorder? *Journal of Clinical Psychiatry, 62*(Suppl. 11), 4–12.

Roemer, L., & Orsillo, S. M. (2002). Expanding our conceptualization of and treatment for generalized anxiety disorder: Integrating mindfulness/acceptance-based approaches with existing cognitive–behavioral models. *Clinical Psychology: Science and Practice, 9,* 54–68.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 232–234). New York: Russell Sage Foundation.

*Sarkar, P., Rathee, S. P., & Neera, N. (1999). Comparative efficacy of pharmacotherapy and biofeedback among cases of generalised anxiety disorder. *Journal of Projective Psychology and Mental Health, 6,* 69–77.

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 261–281). New York: Russell Sage Foundation.

Shadish, W. R., & Ragsdale, K. (1996). Random versus non-random assignment in controlled experiments: Do you get the same answer? *Journal of Consulting and Clinical Psychology, 4,* 1290–1303.

Smith, M. L., Glass, G. V., & Miller, T. I. (1980). *The benefits of psychotherapy.* Baltimore: John Hopkins University Press.

Stevens, S. E., Hynan, M. T., & Allen, M. (2000). A meta-analysis of common factor and specific treatment effects across the outcome domains of the phase model of psychotherapy. *Clinical Psychology: Science and Practice, 7,* 273–290.

Wells, A. (2000). *Emotional disorders and metacognition: Innovative cognitive therapy.* New York: Wiley.

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

Wittchen, H.-U., Zhao, S., Kessler, R. C., & Eaton, W. W. (1994). *DSM–III–R* generalized anxiety disorder in the National Comorbidity Survey. *Archives of General Psychiatry, 51,* 355–364.

Woods, J. H., Katz, J. L., & Winger, G. (1992). Benzodiazepines: Use, abuse, and consequences. *Pharmacology Review, 44,* 151–347.

*Woodward, R., & Jones, R. B. (1980). Cognitive restructuring treatment: A controlled trial with anxious patients. *Behaviour Research and Therapy, 18,* 401–407.

Received August 26, 2004
Revision received January 28, 2005
Accepted April 11, 2005 ■