Is cognitive–behavioral therapy more effective than other therapies?
A meta-analytic review

David F. Tolin *

The Institute of Living, USA
Yale University School of Medicine, USA

ABSTRACT

Cognitive–behavioral therapy (CBT) is effective for a range of psychiatric disorders. However, it remains unclear whether CBT is superior to other forms of psychotherapy, and previous quantitative reviews on this topic are difficult to interpret. The aim of the present quantitative review was to determine whether CBT yields superior outcomes to alternative forms of psychotherapy, and to examine the relationship between differential outcome and study-specific variables. From a computerized literature search through September 2007 and references from previous reviews, English-language articles were selected that described randomized controlled trials of CBT vs. another form of psychotherapy. Of these, only those in which the CBT and alternative therapy condition were judged to be *bona fide* treatments, rather than "intent-to-fail" conditions, were retained for analysis (28 articles representing 26 studies, $N = 1981$). Four raters identified post-treatment and follow-up effect size estimates, as well as study-specific variables including (but not limited to) type of CBT and other psychotherapy, sample diagnosis, type of outcome measure used, and age group. Studies were rated for methodological adequacy including (but not limited to) the use of reliable and valid measures and independent evaluators. Researcher allegiance was determined by contacting the principal investigators of the source articles. CBT was superior to psychodynamic therapy, although not interpersonal or supportive therapies, at post-treatment and at follow-up. Methodological strength of studies was not associated with larger or smaller differences between CBT and other therapies. Researchers' self-reported allegiance was positively correlated with the strength of CBT's superiority; however, when controlling for allegiance ratings, CBT was still associated with a significant advantage. The superiority of CBT over alternative therapies was evident only among patients with anxiety or depressive disorders. These results argue against previous claims of treatment equivalence and suggest that CBT should be considered a first-line psychosocial treatment of choice, at least for patients with anxiety and depressive disorders.

© 2010 Elsevier Ltd. All rights reserved.

Contents

* The Institute of Living, 200 Retreat Avenue, Hartford, CT 06106, USA. Tel.: +1 860 545 7685; fax: +1 860 545 7156.
E-mail address: dtolin@harthosp.org.

1. Introduction

The diversity of psychotherapeutic approaches in clinical practice (Garfield & Bergin, 1986; Goisman, Warshaw, & Keller, 1999), as well as outcome data that would appear to support many of these approaches (Butler, Chapman, Forman, & Beck, 2006; de Mello, de Jesus Mari, Bacaltchuk, Verdeli, & Neugebauer, 2005; Dobson, 1989; Driessen et al., 2010; Hofmann & Smits, 2008; Leichsenring & Rabung, 2008), unavoidably leads to controversy regarding which form of psychotherapy is most effective. Some authors, for example, have argued that cognitive–behavioral therapy (CBT) is superior to alternative psychotherapies (Eysenck, 1994; Hunsley & Di Giulio, 2002). This argument is bolstered by the results of two limited-scope meta-analyses: Shapiro and Shapiro (1982) examined 143 comparative outcome studies published over a 5-year period and found an effect size (Cohen's *d*) of 1.06 for behavioral treatments vs. 0.40 for dynamic/humanistic treatments. Shadish, Matt, Navarro and Phillips (2000) sampled 90 studies conducted under "clinically representative" conditions; regression analyses indicated that behavioral orientation was a significant predictor of effect size.

On the other hand, some meta-analysts have failed to find evidence for the superiority of one form of psychotherapy over another. The collective argument for psychotherapy equivalence has been commonly termed the "Dodo Bird Verdict" (Luborsky, Singer, & Luborsky, 1975), after the character in *Alice in Wonderland* who proclaimed after a chaotic race that "Everyone has won, and all must have prizes." The seminal meta-analysis by Smith and Glass (1977) revealed few differences between behavior therapy and other forms of psychotherapy: effect size estimates (Cohen's *d*) were 0.8 for behavior therapy and 0.6 for all other therapies combined. This meta-analysis has been criticized in several respects, including the categorization of cognitive interventions as "non-behavioral" therapies, despite their common co-utilization in practice (Hunsley & Di Giulio, 2002; Wilson & Rachman, 1983); failure to include many well-controlled studies of behavior therapy (Rachman & Wilson, 1980); and absence of correction for methodologically weak studies (Wilson & Rachman, 1983). Subsequent re-analyses of Smith and Glass's data, correcting for these concerns, have suggested superior outcomes for behavioral therapies (Andrews & Harvey, 1981; Hunsley & Di Giulio, 2002).

In what has perhaps become the *de facto* last word on the matter, Wampold et al. (1997) used a novel approach to meta-analysis in which the primary outcome measure was heterogeneity of effect sizes across all forms of psychotherapy. Effect sizes were homogeneous around 0, leading to a conclusion of general equivalency across forms of psychotherapy. Criticisms of the Wampold et al. meta-analysis include the use of homogeneity testing, rather than tests of mean effect size differences (Howard, Krause, Saunders & Kopta, 1997); absence of attempt to discriminate among different kinds of outcome measures, different forms of psychotherapy, or different patient populations (Crits-Christoph, 1997; Howard et al., 1997); and the fact that the majority [69% (Crits-Christoph, 1997) to 80% (Hunsley & Di Giulio, 2002)] of selected studies were pre-clinical studies, usually of college students, in which one variant of CBT was compared to another variant of CBT.

The question of differential therapeutic efficacy is, therefore, far from settled. As evidence for the efficacy of CBT continues to accumulate and as cognitive–behavioral interventions become more widespread, there is a need to understand more clearly how CBT compares to other forms of psychotherapy. Ideally, such research would follow the prescription of Gordon Paul (1967), who recommended over 40 years ago that "…the question towards which all outcome research should ultimately be directed is the following: *What* treatment, by *whom*, is most effective for *this* individual with *that* specific problem, and under *which* set of circumstances?" (p. 111). Although no single study could ever hope to answer so complicated a question [Beutler (1991) estimated that there would be nearly 1.5 million potential combinations of therapy, therapist, and patient types], the aim of the present study is to capture the spirit of Paul's question using meta-analytic strategies. First, the treatments being used need to be described clearly. The present study examines direct comparisons of CBT vs. other forms of psychotherapy, with specific attention to what kind of CBT is being compared to what kind of alternative psychotherapy. Second, the present study focuses on who is being treated: each study is coded according to the target diagnosis or presenting problem, how the sample was selected (e.g., from a clinic vs. from a pool of student volunteers), and the age range of the patients. Third, the present analyses will attend to how clinical outcome is defined: for example, is outcome defined as a reduction in primary symptoms (such as reduced depression severity in a study of depressed patients), global severity (such as clinician's ratings of overall improvement), or improvement in functional measures (such as quality of life, work productivity, or social adjustment)? Fourth, the present study examines the impact of study quality and methodological adequacy according to criteria proposed by Jadad et al. (1996) and Foa and Meadows (1997). Fifth, the present study includes only comparisons of *bona fide* psychotherapies, which are distinguished from "intent-to-fail" treatments commonly used to control for nonspecific treatment effects; *bona fide* treatments were identified using criteria proposed by Wampold et al. (1997) and Westen, Novotny and Thompson-Brenner (2004). Finally, the present analysis accounts for who conducted the original source studies. In particular, the impact of researcher allegiance to one school of psychotherapy or another is examined. Researcher allegiance is attracting increased attention as a possible complicating factor in comparative outcome studies; previous meta-analyses have suggested that measures of researcher allegiance can account for more than half of the variance in observed differences in outcomes (Gaffan, Tsaousis & Kemp-Wheeler, 1995; Luborsky et al., 1999). The presence of researcher allegiance does not necessarily imply bias, as others have noted (Hollon, 1999; Leykin & DeRubeis, 2009); nevertheless, it is included here as a possible moderator of study outcome. Previous researchers have attempted to quantify researcher allegiance indirectly by examining articles for references to previous published research showing superiority of one treatment, a specific hypothesis or rationale as to why one treatment should be superior, a detailed description of a treatment's procedure and aims, treatments devised or first introduced by one of the authors, or only one treatment being included in the study (Gaffan et al., 1995). These criteria seem inadequate to assess a researcher's personal allegiances, as many of them may simply reflect good science. Instead, the present study includes allegiance ratings collected directly from the principal investigators of each study.
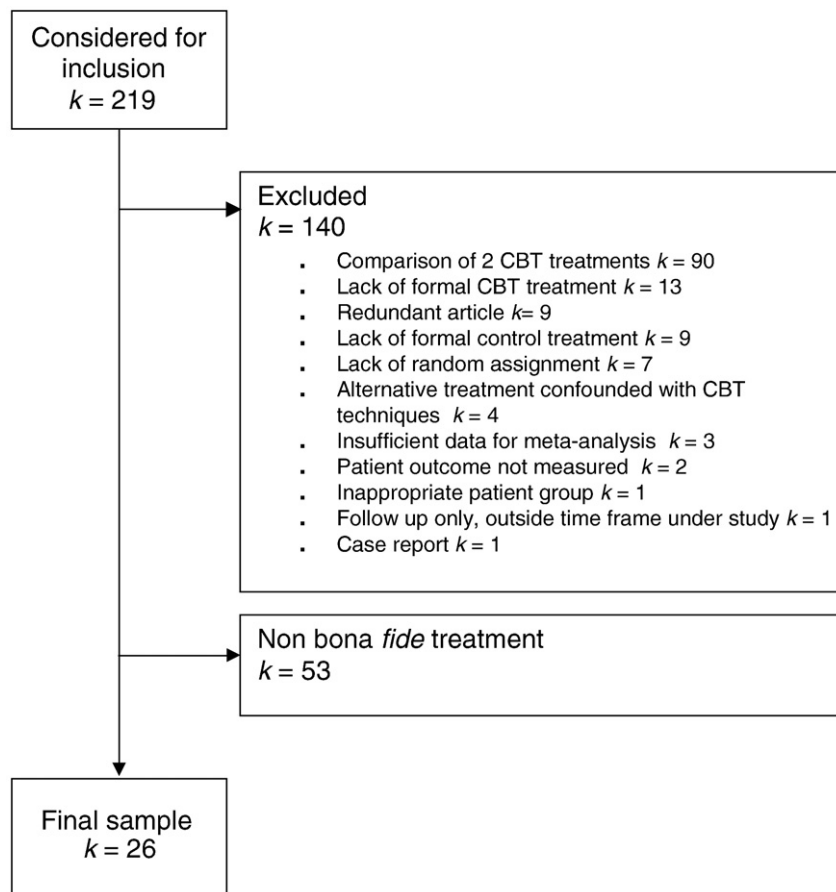
Fig. 1. Selection of articles for meta-analysis. Note: k = number of studies.

2. Method

2.1. Data sources

Journal articles were identified through searches of the Medline and PsycINFO electronic databases through September 2007 using the search terms [(Behavior Therapy or Cognitive Therapy or Cognitive Behavior Therapy or Rational Emotive Behavior Therapy or CBT or Exposure Therapy or Behavior Modification or Skill Learning or Dialectical or Cognitive Restructuring or Cognitive Behavioral Therapy or Schema Therapy) and (Dynamic or Psychodynamic or Psychoanalysis or Interpersonal Psychotherapy or Expressive Psychotherapy or Supportive Psychotherapy or Insight Therapy or Insight-Oriented or Psychoanalytic or Experiential Psychotherapy or Existential Therapy or Client Centered Therapy or Humanistic Psychotherapy or Transactional Analysis or Gestalt Therapy)]. The search was limited to English-language publications, coded as treatment outcome studies or randomized clinical trials. Relevant studies also were identified through references of originally identified articles and published reviews (including previous meta-analyses). This literature search identified 219 articles, which then were examined for inclusion.

2.2. Study selection and data extraction

Randomized trials were considered for inclusion if (1) sufficient information was provided to compute effect sizes (or necessary additional information was supplied by the authors; three authors were contacted for this information and 2 replied); and if they (2) compared some variant of CBT with a non-CBT psychotherapy. The most

common reason for study exclusion (see Fig. 1) was a comparison between two variants of CBT. Some articles were excluded because they represented a second analysis of a data set already included in the meta-analysis (although if new outcome measures were reported in a second paper, these were included with the original study).

Study data (Table 2) were extracted by the author. Four Ph.D.-level clinician–scientists,[1] including the author, co-rated 28 (35%) of the 79 articles meeting basic entry criteria. As seen in the Table, agreement was generally high; disagreements were resolved by mutual discussion until 100% agreement was reached. In addition to basic study information, studies were also coded for methodological strength according to the Jadad et al. (1996) and Foa and Meadows (1997) criteria.

Researcher allegiance of the 79 articles under consideration at this stage was examined by contacting all of the PIs who were still living and could be located. Of 65 PIs who were sent a letter or email, 41 (63% response rate, representing 47 studies dating from 1966–2007) responded and provided allegiance ratings.[2] Allegiance was defined as "the extent to which you believe in a treatment, expect it to succeed beyond the nonspecific effects of treatment, or identify yourself as a

proponent of that treatment." For each CBT vs. alternative treatment comparison (at the time the study was conducted), PIs rated, on a 7-point scale from $+3$ ("strongly favored [name of CBT therapy]") to -3 ("strongly favored [name of alternative therapy]"), with 0 reflecting "no preference at all or equal distribution of allegiance," the allegiance of the PI, the study therapists, and the research team (all key personnel).

From the 79 articles meeting the basic inclusion criteria described above, a second pass was made, using the co-raters' responses, to retain only those studies comparing two or more *bona fide* treatments. In order to distinguish *bona fide* treatments from "intent-to-fail" conditions, we relied on Wampold et al. (1997) operational definition of *bona fide* psychotherapies as those that are (a) delivered by trained therapists and (b) based on psychological principles, (c) offered to the psychotherapy community as books or manuals, or (d) contained specified components. Therefore, a treatment was considered *bona fide* if criterion (a) and at least one of the criteria (b), (c), or (d) was met. Furthermore, in keeping with comments by Westen et al. (2004), any artificial restrictions on therapist activity that would not be consistent with that model of therapy (e.g., restricting discussions of traumatic events in supportive therapy conditions) rendered the treatment non-*bona fide*. These criteria were applied to the CBT and non-CBT conditions. Fifty three (67%) articles were excluded based on the absence of two *bona fide* therapy conditions. The 28 resultant articles (see Table 1) represented 26 distinct studies of CBT vs. other forms of psychotherapy using 1981 patients.

2.3. Data synthesis

Most studies used multiple dependent measures. When possible, these measures were divided according to the categories listed in Table 2 (primary symptom severity, self-concept, comorbid symp-

Table 1
Articles used in the meta-analysis.

Study	N	Diagnosis	Alternative therapy
Alden (1988)	105	Substance	Other
Areán et al. (1993)	47	Depressive	Other
Barkham et al. (1996)	36	Depressive	Psychodynamic
Barkham, Shapiro, Hardy and Rees (1999)	116	Depressive	Psychodynamic
Beutler et al. (1991)	43	Depressive	Psychodynamic
Carroll et al. (2004)	121	Substance	Interpersonal
Durham et al. (1994, 1999)	80	Anxiety	Psychodynamic
Durham et al. (2003)	41	Psychotic	Supportive
Elkin et al. (1989)	120	Depressive	Interpersonal
Emmelkamp et al. (2006)	40	Personality	Psychodynamic
Fairburn, Kirk, O'Connor and Cooper (1986)	22	Eating	Psychodynamic
Gallagher-Thompson and Steffen (1994)	66	Depressive	Psychodynamic
Gelder, Marks and Wolff (1967)	26	Anxiety	Psychodynamic
Giesen-Bloo et al. (2006)	86	Personality	Psychodynamic
Levy et al. (2006)	59	Personality	Psychodynamic
McIntosh et al. (2005)	56	Eating	Interpersonal, supportive
McLean and Hakistan (1979)	129	Depressive	Psychodynamic
Miller, Barrett, Hampe and Noble (1972)	44	Anxiety	Psychodynamic
Shapiro et al. (1994, 1995)	116	Depressive	Psychodynamic
Shear, Houck, Greeno and Masters (2001)	66	Anxiety	Psychodynamic
Sloane, Staples, Cristol and Yorkston (1975)	60	Unspecified	Psychodynamic
Svartberg, Stiles and Seltzer (2004)	50	Personality	Psychodynamic
Thompson, Gallagher and Breckenridge (1987)	91	Depressive	Psychodynamic
Ward et al. (2000)	260	Depressive	Supportive
Wilfley et al. (1993)	36	Eating	Interpersonal
Woody et al. (1983)	65	Substance	Psychodynamic

toms, global symptom severity, general functioning, mechanism, social adjustment, quality of life). However, many studies used multiple dependent measures within the same category (e.g., a study of PTSD which used the Impact of Events Scale and PTSD Checklist, both categorized as self-report measures of primary symptom severity). In these cases, an aggregate effect size (Rosenthal & Rubin, 1986) was calculated from the mean of the effect size estimates (all expressed as d) and the pooled variance, using the most conservative estimate of correlation among the outcome measures ($r = 1.0$). Larger estimated correlations among outcome measures produce smaller outcome effect size estimates than do smaller estimated correlations; thus, this conservative estimate risks underestimating true differences across treatments. The aggregate effect size was used in overall analyses (e.g., each patient sample contributed only 1 overall effect size for a given outcome type), although the individual measure effect sizes were used in the examination of measure-specific moderator variables (i.e., reliable and valid measures, use of blind independent evaluators, assessor training, source of data).

Data were analyzed using Comprehensive Meta-Analysis v.2.2 software, with supplemental analyses conducted using SPSS v.15. For each comparison of continuous data (e.g., mean scores on a measure of psychopathology), Cohen's d was calculated, weighted by sample size. A d value of 0.0 indicates no difference between CBT and alternative treatment participants; conventionally, 0.2, 0.5, and 0.8 are taken to represent small, medium, and large effects, respectively (Cohen, 1988). For studies reporting dichotomous outcomes (i.e., treatment responder vs. nonresponder), odds ratios (OR) were calculated and converted to d. For other studies, d was obtained from published t or p values, using the most conservative estimate possible when p values were used. The 95% confidence interval (CI), statistical significance (p), and within-group heterogeneity (Q_{within}) was calculated for each effect size estimate. To examine differences between effect size estimates, the mixed-effects between-group heterogeneity ($Q_{between}$) was calculated. Random-effects models were used. To test the file drawer effect (the probability that unpublished null results would eliminate the obtained results), for each significant result the fail-safe N (FSN), or the number of null results that would be needed to overturn a significant result (i.e., $p > 0.05$), was computed. Generally, if the FSN is greater than or equal to 5 times the number of studies in the analysis plus 10 ($FSN > 5k + 10$), the obtained results are considered robust against the file drawer effect (Rosenthal, 1991a). Meta-regression was used to test the impact of moderator variables, using the Z-test of the significance of the slope of the regression line (Borenstein, Hedges, Higgins, & Rothstein, 2009).

3. Results

3.1. Comparisons between CBT and other therapies

Overall, CBT was associated with lower scores on measures of primary symptoms at post-treatment than were alternative treatments (see Table 3). Comparison across different types of alternative treatment did not indicate significant heterogeneity ($Q_{between} = 3.05$, $p = 0.38$). Examination of Table 3 indicates that CBT proved significantly more effective than psychodynamic therapy, but not interpersonal, supportive, or "other" therapy. None of the comparisons were robust against the file drawer effect. Categories containing only 1 study are not reported here due to insufficiency of data for meta-analysis.

CBT was also associated with lower scores than were alternative therapies at 6-month and 1-year follow-up; significant between-group heterogeneity was not detected at 6 months ($Q_{between} = 0.74$, $p = 0.69$) but was detected at 1 year ($Q_{between} = 12.30$, $p = 0.002$). Follow-up results generally paralleled those at post-treatment, and

Table 2
Variables extracted from the studies.

Variable	Comment	Match Rate
Study author, title, and year		–
Sample size, group means and standard deviations (for continuous data) or frequency counts (for categorical data)	In cases where these data were not provided, effect size estimates were derived from statistical comparisons (t or p values). When p ranges, rather than values, were reported (e.g., $p<0.05$), the most conservative estimate was used (in this case, $p = 0.049$). When authors reported only that a comparison was not statistically significant, p was conservatively assumed to be 1.0. When more than 2 treatments were compared within a single study, each CBT vs. alternative therapy comparison was treated as independent, a procedure which can lead to errors of significance testing but is generally considered acceptable for purposes of effect size estimation (Glass, McGaw, & Smith, 1981; Rosenthal, 1991b; Smith & Glass, 1977).	–
Time of measurement	Follow-up duration was recoded using assessments closest to 6 months and 1 year after treatment discontinuation (in cases when two assessments were equidistant from one of these time points, the later assessment was used). Post-maintenance assessments, in which minimal additional treatment was provided over a maintenance period, were considered follow-up assessments.	–
Type of CBT	Coded for the presence or absence of well established components (Goldfried & Davison, 1994) including *relaxation training* (including progressive muscle relaxation, meditation, or breathing retraining), *exposure therapy* (imaginal or *in vivo* exposure, including flooding and implosive therapy), *behavior rehearsal* (behavioral training in social skills, habit reversal, or problem solving), *cognitive restructuring* (including direct strategies to identify and alter maladaptive thought processes), or *operant procedures* (systematic manipulation of reinforcers or punishers for behavior, including behavioral activation). When unclear from the published manuscript, the cited treatment manual (if applicable) was consulted.	86%
Type of other treatment	Classified as *supportive* (any nondirective therapy, typically based on principles such as expressing empathy, establishing therapeutic relationship), *psychodynamic* (treatment that emphasizes uncovering unconscious conflicts, personality dynamics, or working through transference reactions), *interpersonal* (emphasizing understanding interpersonal conflicts and developing more adaptive ways to relate to others), or *other* (any treatments not elsewhere classified, including hypnosis and spiritual counseling).	97%
Concurrent treatment (e.g., medications)		88%
Diagnosis or diagnoses of the sample	Condensed into superordinate diagnoses of *anxiety disorders*, *depressive disorders* (including complicated grief), *eating disorders*, *substance use disorders*, *psychotic disorders*, *developmental disorders*, *habit disorders*, *personality disorders*, *martial distress*, *academic problems*, and *other or unspecified*. Another group, called *subclinical*, was reserved for samples that were not described as having any psychiatric disturbance (e.g., at-risk populations).	100%
Population	Coded as *students* (any sample of students drawn from a school population, regardless of symptom severity), *recruited patients* (patients who were recruited via advertisement, letter, or other means who would likely not have sought treatment from the investigative team had there not been some kind of marketing), or *clinic patients* (patients who sought treatment and were then recruited into the study, would have been in treatment with the investigative team even if there was no study or marketing).	87%
Treatment duration (weeks) and number of sessions		91%
Completer vs. intent-to-treat (ITT) analyses	When both were provided, ITT analyses were selected for aggregate analyses.	94%
Type of measure	Coded as *primary symptom severity* (measures one or more symptoms of the superordinate diagnosis), *self-concept* (any self-referential beliefs, e.g., self esteem, self efficacy), *comorbid symptoms* (psychiatric symptoms pertaining to a specific disorder that are not part of the superordinate diagnosis), *global symptom severity* (an overall measure of psychiatric symptoms that is not limited to a specific diagnosis or cluster of diagnoses), *general functioning* (an overall measure of functional impairment that is not limited to a specific domain of functioning), *mechanism* (a measure that is not intended to be an outcome measure but rather relates to a specific model of treatment and is expected to show different mechanisms across treatments), *social adjustment* (a measure of social functioning or social skill), or *quality of life* (an overall measure of life satisfaction that is not limited to a specific domain of life satisfaction).	96%
Source of data	Coded as *self-report*, *rater* (any trained rater who is part of the research or clinical team), *parent*, or *corroborative* (objective data, e.g., weight).	100%
Age group	Coded as *child*, *adolescent*, or *adult*.	100%
Individual vs. group therapy		94%
Manualized and replicable CBT	Present if the CBT description included reference to a treatment manual; reference to a clinical trial that did not provide sufficient detail to replicate the basic treatment did not satisfy this criterion.	97%
Manualized and replicable alternative therapy	As above.	79%
Diagnosis required for inclusion	Present if participants were required to meet criteria for a psychiatric disorder according to DSM, ICD, or other recognized scheme.	94%
Minimum severity criterion for inclusion	Present if participants had to have some minimal level of symptom severity to be included.	97%
Clearly defined inclusion/exclusion criteria for study entry		91%
Reliable and valid measure	Present if there was any reference to acceptable psychometric properties such as adequate inter-rater reliability or correlation with related measures.	94%
Use of independent evaluators blind to treatment condition		91%
Assessor training	Present if there was a specific reference to the fact that assessors were trained to use the study measures.	91%
Acceptable CBT fidelity	Present if adherence to the CBT protocol was assessed by having a rater review tapes of the therapy sessions. Ratings must show at least adequate fidelity. Supervision without such ratings or previous training in the treatment did not satisfy this criterion.	97%
Acceptable alternative therapy fidelity	As above.	94%
Trained therapists	Present if, in both treatment conditions, at least some therapists must hold a doctoral degree, at least most must not be trainees, and at least most must have some mention of prior experience in the treatment provided.	100%
CBT therapist allegiance	Present if therapists reported a preference in, positive expectations about, or specialization in the CBT treatment.	88%

Table 2 (continued)

Variable	Comment	Match Rate
Alternative therapy therapist allegiance	As above.	88%
Adequate randomization	Present if specific randomization procedures are mentioned and adequate.	88%
Participants blind to condition		100%
Description of dropouts	Present if the authors described the number of patients who dropped out, as well as the reasons for dropouts.	63%
CBT *bona fide*	Present if: treatment delivered by trained therapists, no artificial restrictions on therapist activity that would not be consistent with that model of therapy, and at least one of: (1) based on psychological principles, (2) offered to the psychotherapy community as books or manuals, or (3) contained specified components.	100%
Alternative therapy *bona fide*	As above.	88%

the 6-month follow-up effects for psychodynamic and overall therapy were robust against the file drawer effect.

3.2. Comparisons between CBT and other therapies for specific conditions

Significant heterogeneity was not detected across clinical conditions for primary symptom measures at post-treatment ($Q_{between} = 3.28$, $p = 0.77$). As shown in Table 4, significant effects were found for anxiety and depressive disorders, although neither was robust against the file drawer effect. Examination of the 21 comparisons for anxiety or depressive disorders (the two most frequently-studied clinical conditions) revealed the following post-treatment d and CI values for CBT vs. specific alternative therapies: interpersonal ($k = 1$, $d = -0.14$, CI $= -0.52$–0.24), other ($k = 1$, $d = 0.53$, CI $= -0.07$–1.13), psychodynamic ($k = 18$, $d = 0.31$, CI $= 0.14$–0.48), supportive ($k = 1$, $d = 0.04$, CI $= -0.20$–0.29), total ($k = 21$, $d = 0.26$, CI $= 0.10$–0.41). CBT was associated with significantly ($p < 0.05$) lower post-treatment scores than were psychodynamic and "other" therapies, but not interpersonal or supportive therapies.

3.3. Examining type of outcome measure

When post-treatment effect sizes were compared across different types of outcome measure, no significant heterogeneity was found ($Q_{between} = 3.124$, $p = 0.873$). Table 5 shows that CBT was associated with lower post-treatment scores than were alternative therapies for symptoms of the primary illness as well as general functioning and global symptom severity. Small but nonsignificant effects for comorbid symptoms and quality of life, favoring CBT, were also noted. CBT was not associated with lower post-treatment scores on measures of self-concept or social adjustment.

3.4. Type of cognitive–behavioral therapy

Studies were identified as including one or more commonly-accepted CBT variants (Goldfried & Davison, 1994); most contained more than one. For each CBT variant, studies were divided into two groups: those that included the variant, and those that did not. These two groups were then examined for effect size relative to alternative psychotherapy conditions. Effect sizes (vs. alternative psychotherapies) were significant, moderate, and comparable to one another for treatments that included relaxation ($k = 6$, $d = 0.32$, CI $= 0.01$–0.64, $p = 0.046$) vs. no relaxation ($k = 26$, $d = 0.20$, CI $= 0.06$–0.34, $p = 0.005$) ($Q_{between} = 0.506$, $p = 0.477$), exposure ($k = 5$, $d = 0.15$, CI $= -0.12$–0.43, $p < 0.270$) vs. no exposure ($k = 27$, $d = 0.24$, CI $= 0.09$–0.38, $p = 0.001$) ($Q_{between} = 0.279$, $p = 0.597$), behavioral rehearsal ($k = 19$, $d = 0.28$, CI $= 0.11$–0.45, $p = 0.001$) vs. no behavioral rehearsal ($k = 13$, $d = 0.15$, CI $= -0.04$–0.34, $p = 0.123$) ($Q_{between} = 1.051$, $p = 0.305$), cognitive restructuring ($k = 23$, $d = 0.27$, CI $= 0.09$–0.45, $p = 0.002$) vs. no cognitive restructuring ($k = 9$, $d = 0.13$, CI $=$

Table 3
Effect size estimates for cognitive–behavioral therapy compared to control therapies using post-treatment measures of primary symptoms.

Comparison	# Studies	d	95% CI	FSN	Q_{within}
Post-treatment					
Interpersonal	4	0.07	−0.31–0.45	–	7.40
Other	2	0.24	−0.24–0.73	–	1.66
Psychodynamic	24	0.28*	0.12–0.44	99	37.43
Supportive	2	0.05	−0.18–0.28	–	0.00
Total	32	0.22**	0.09–0.35	136	49.37*
Approximate 6 month follow-up					
Interpersonal	0	–	–	–	–
Other	1	–[††]	–	–	–
Psychodynamic	18	0.50**	0.29–0.71	111[†]	21.24
Supportive	1	–[††]	–	–	–
Total	20	0.47**	0.29–0.66	132[†]	21.98
Approximate 1 year follow-up					
Interpersonal	–	–	–	–	–
Other	1	–[††]	–	–	–
Psychodynamic	7	0.55**	0.30–0.81	25	4.17
Supportive	1	–[††]	–	–	–
Total	9	0.34*	0.06–0.62	73	16.46*

Note: d = Cohen's d (effect size). CI = Confidence Interval. FSN = Fail-Safe *N*. Q_{within} = within-group heterogeneity.
* $p < 0.05$.
** $p < 0.001$.
[†] Robust against file drawer effect.
[††] Categories with $k = 1$ are not reported.

Table 4
Effect size estimates for cognitive–behavioral therapy compared to control therapies for specific conditions using post-treatment measures of primary symptoms.

Condition	# Studies	d	95% CI	FSN	Q_{within}
Academic	0	–	–	–	–
Anxiety	5	0.43*	0.14–0.72	7	2.96
Depressive	16	0.21*	0.04–0.39	29	23.54
Developmental	0	–	–	–	–
Eating	3	0.14	−0.44–0.72	–	4.07
Habit	0	–	–	–	–
Marital	0	–	–	–	–
Personality	3	0.34	−0.08–0.77	–	3.35
Psychotic	1	–[††]	–	–	–
Subclinical	0	–	–	–	–
Substance	3	0.01	−0.57–0.59	–	10.47*
Unspecified	1	–[††]	–	–	–
Total	32	0.22**	0.09–0.35	136	49.37*

Note: d = Cohen's d (effect size). CI = Confidence Interval. FSN = Fail-Safe *N*. Q_{within} = within-group heterogeneity. [†]Robust against file drawer effect.
* $p < 0.05$.
** $p < 0.001$.
[††] Categories with $k = 1$ are not reported.

Table 5
Effect size estimates for cognitive–behavioral therapy compared to control therapies using different categories of post-treatment measures.

Measure type	# Studies	d	95% CI	FSN	Q_{within}
Comorbid symptoms	11	0.24	−0.02–0.50	–	19.41
General functioning	6	0.27*	0.02–0.51	5	5.56
Global symptoms	14	0.21*	0.01–0.42	13	19.57
Primary symptoms	32	0.22**	0.09–0.35	136	49.37*
Quality of life	3	0.21	−0.03–0.46	–	1.98
Self-concept	10	0.11	−0.14–0.36	–	6.55
Social adjustment	19	0.13	0.00–0.26	–	14.70

Note: d = Cohen's d (effect size). CI = Confidence Interval. FSN = Fail-Safe *N*. Q_{within} = within-group heterogeneity. †Robust against file drawer effect.
 * $p<0.05$.
 ** $p<0.001$.

−0.02–0.28, $p=0.099$) ($Q_{between}=1.430$, $p=0.232$), and operant procedures ($k=16, d=0.22$, CI=0.09–0.35, $p=0.001$) vs. no operant procedures ($k=16, d=0.18$, CI=0.01–0.34, $p=0.042$) ($Q_{between}=0.580$, $p=0.446$).

3.5. Moderator effects

Moderator analyses were conducted on measures of primary symptom severity at post-treatment. Indices of methodological adequacy were obtained by assigning each study 1 point for each of the criteria listed by Jadad et al. (1996) and Foa and Meadows (1997). Scores on the Jadad scale ($\alpha=0.31$) ranged from 1–4; scores on the Foa and Meadows scale ($\alpha=0.68$) ranged from 1–13. As shown in Fig. 2, neither index was a significant predictor of between-group effect size.

Effect size for CBT vs. alternative therapy was not significantly associated with study year ($Z=1.70$, $p=0.09$), treatment duration ($Z=1.46$, $p=0.14$), or number of sessions ($Z=1.51$, $p=0.13$). Twenty-nine comparisons of individual therapy yielded an average effect size of 0.23 (CI=0.10–0.36, $p=0.001$); these did not differ significantly ($Q_{between}=0.18$, $p=0.67$) from the 3 comparisons of group therapy ($d=0.12$, CI=−0.36–0.60, $p=0.62$). Effects did not differ significantly ($Q_{between}=0.44$, $p=0.80$) for studies that used clinic patients ($k=6$, $d=0.16$, CI=−0.11–0.42, $p=0.109$) vs. recruited patients ($k=22, d=0.26$, CI=0.08–0.44, $p=0.005$); none of the retained studies used students. Corroborative sources ($k=6$, $d=0.31$, CI=−0.10–0.51, $p=0.003$) produced the greatest effects for CBT vs. alternative therapies, followed by patient self-report ($k=85$, $d=0.21$, CI=0.12–0.29, $p<0.001$) and professional raters ($k=38$, $d=0.17$, CI=0.05–0.29, $p=0.004$). No significant between-group homogeneity was found ($Q_{between}=1.34$, $p=0.85$). Due to small number of studies, moderator analyses could not be conducted based on age of the sample or using parent report of child's treatment response.

3.6. Allegiance effects

Allegiance effects were examined using linear regression analyses predicting effect size (d) of measures of primary symptom severity at post-treatment from the three allegiance ratings provided by the PIs. PI allegiance ratings, but not therapist or research team allegiance ratings, were mildly in favor of CBT (PI mean = 0.75, range = −2 to 2; therapist mean = 0.00, range −1 to 1; research team mean = −0.02, range −1 to 1). PI allegiance ratings, but not therapist or research team allegiance ratings, correlated significantly with study year (PI $r=0.30$, $p=0.045$; therapist $r=0.05$, $p=0.752$; research team $r=-0.05$, $p=0.743$). Fig. 3 depicts the results of the regression analyses. PI allegiance, but not therapist or research team allegiance, was a significant and positive predictor of effect size at post-treatment. The impact of allegiance was examined further using a univariate GLM with and without the allegiance ratings as covariates. When the studies for which allegiance effects were available were examined without covariates, the mean unweighted effect size ($d=0.29$) was significantly different from 0 ($F_{1,43}=17.93$, $p<0.001$). When PI allegiance, therapist allegiance, and research team allegiance were included as covariates, the mean unweighted effect size was unchanged ($d=0.29$, $F_{3,40}=4.88$, $p=0.006$).
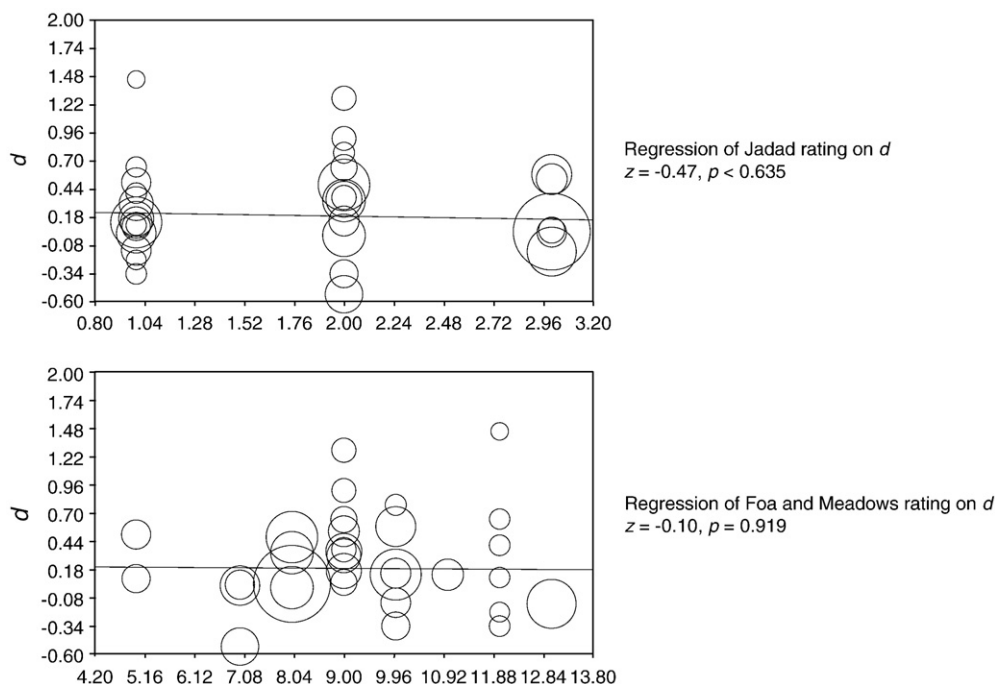
Fig. 2. Fixed-effect regressions of ratings of study methodological strength on effect size (d). Note: *N* for each study is depicted by the size of the circle.
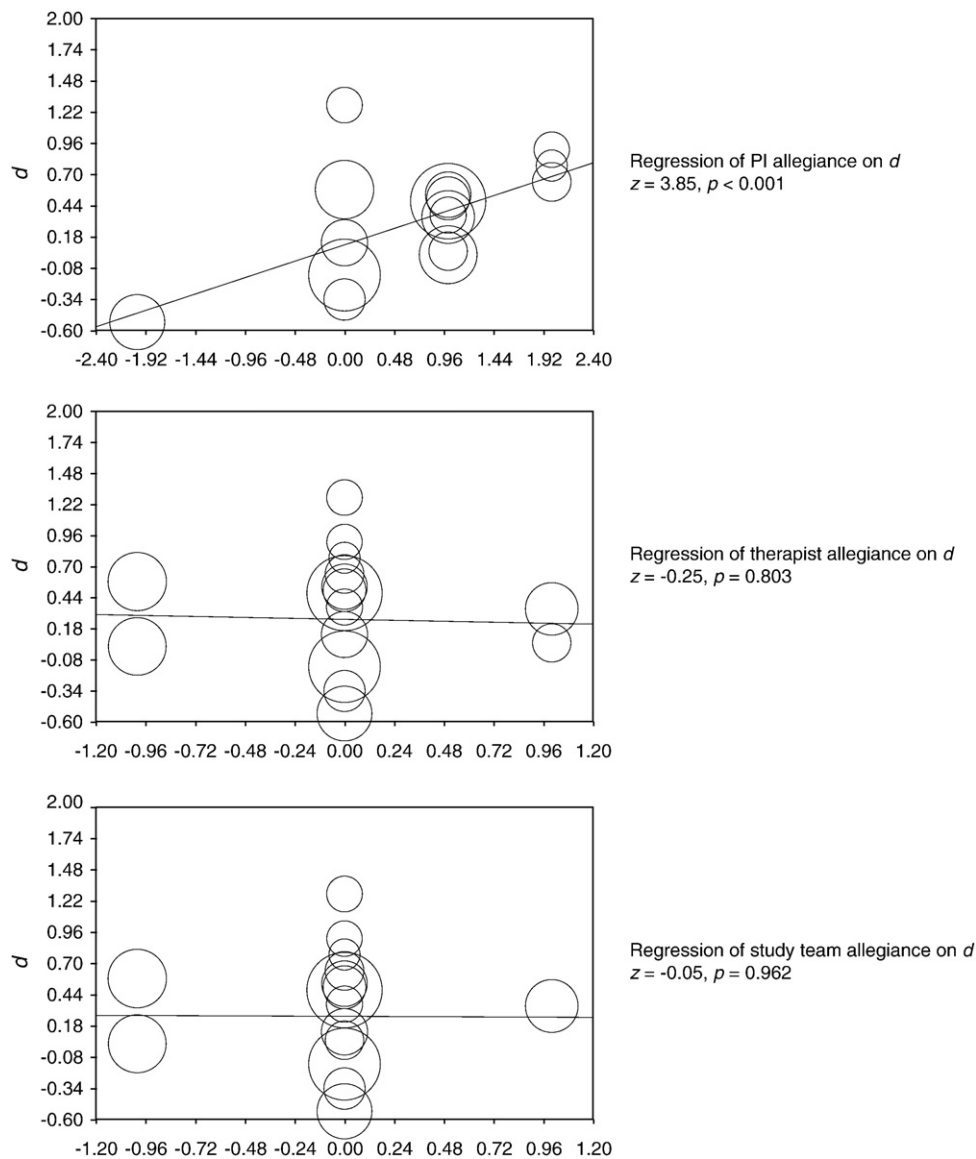
Fig. 3. Fixed-effect regressions of principal investigator (PI), therapist, and study team allegiance ratings on effect size (d). Note: N for each study is depicted by the size of the circle.

4. Discussion

Several limitations of the present analyses should be acknowledged. It is important to recognize that only pairwise comparisons between CBT and other therapies were conducted. The present results cannot, therefore, be used to infer similarities or differences among the alternative therapies (e.g., between interpersonal vs. psychodynamic therapy) or among variants of CBT (e.g., between exposure therapy vs. cognitive restructuring). Although the overall number of studies is reasonable (26 studies with 1981 patients), the small number of studies for certain sub-analyses merits caution of interpretation. Many of the sub-analyses lacked statistical power; in particular, only tentative conclusions can be drawn regarding the effect of CBT vs. other therapies for conditions other than anxiety or depressive disorders. The general lack of robust findings (with the exception of CBT vs. psychodynamic therapy at 6-month follow-up) is noteworthy; despite the substantial number of studies comparing CBT to other psychotherapies, the findings could be overturned by unpublished null results. Although 62% of PIs were located and responded to queries about researcher allegiance, it is possible that

the remaining 38% (who are disproportionately associated with older studies) would have responded differently. The choice to use all available outcome measures could be interpreted as a strength or a weakness of the meta-analysis, depending on one's perspective. A reasonable argument could be made to restrict analyses to only those measures deemed reliable and valid. Conversely, it was decided to include all pertinent data and to examine measurement validity as a possible moderator variable, although it is acknowledged that the inclusion of less valid measures could have introduced error variance into the aggregate analyses. Finally, the present analyses inform us of the relative, but not absolute, efficacy of treatments; other meta-analytic reviews suggest that CBT appears effective for some conditions (Butler et al., 2006; Dobson, 1989; Hofmann & Smits, 2008) but not for others (Lynch, Laws, & McKenna, 2009).

4.1. Returning to the critical Paul's question

Returning to the critical question of Paul (1967), "*What* treatment, by *whom*, is most effective for *this* individual with *that* specific problem, and under *which* set of circumstances?" (p. 111), the

number of available studies is not large enough to investigate the vast number of potentially interesting interaction effects (Beutler, 1991). Nevertheless, the obtained pairwise comparisons can provide specific guidance for planning of psychosocial treatment.

4.2. What treatment

Broadly speaking, CBT outperforms other forms of psychotherapy at post-treatment. That is, patients receiving CBT (regardless of specific techniques) tend to exhibit lower primary symptom severity at post-treatment than do patients receiving other psychotherapies, particularly psychodynamic therapy. The apparent superiority of CBT is equally or more evident 6 months and 1 year after treatment discontinuation, suggesting that the advantages of CBT are not time-limited. From a clinical perspective, the present results suggest that CBT should be considered the first-line psychosocial treatment of choice for many, if not most, patients.

It is noted that the advantage for CBT over alternative therapies falls in the small to moderate range of effect size, and one may wonder about the practical significance of the advantage. Using Rosenthal and Rubin (1982) binomial effect size display to provide context to the effect sizes, the advantage of CBT over alternative therapies at post-treatment ($d = 0.22$) is roughly equivalent to that of a treatment with a 55% success rate over one with a 45% success rate. The advantage of CBT over alternative therapies at 6-month follow-up ($d = 0.47$) is roughly equivalent to that of a treatment with a 61% success rate over one with a 39% success rate. Far from trivial, these results argue strongly against the "Dodo Bird Verdict" (Luborsky et al., 1975), and contrast with prior meta-analyses which purportedly supported the conclusion of treatment equivalence (Smith & Glass, 1977; Wampold et al., 1997) [although, as noted previously, re-analyses of those data also suggested an advantage for CBT (Andrews & Harvey, 1981; Hunsley & Di Giulio, 2002)]. The (nonsignificantly) larger effect sizes at 6-month and 1-year follow-up, compared to post-treatment, might suggest a tendency for CBT gains to be maintained (or improved) to a greater extent than are gains from other psychotherapies. Such a conclusion would be consistent with models of CBT based on learning, although clearly more research on long-term maintenance and relapse processes is needed.

The finding of superior results for patients receiving CBT is not without qualifiers, however. Although CBT proved superior to psychodynamic therapy, there is no evidence at this time that CBT is superior to bona fide IPT or supportive therapies. IPT has primarily been investigated as a treatment for depression (de Mello et al., 2005), although it has also been explored as a treatment for substance abuse (Carroll, Rounsaville, & Gawin, 1991) and eating disorders (McIntosh et al., 2005; Wilfley et al., 2002). IPT appears more similar in many respects to CBT than to supportive or psychodynamic therapies, e.g., its time-limited and focused structure, focus on specific symptoms of illness, emphasis on present rather than historical events, and deemphasis of unconscious and transference-based foci (Bleiberg & Markowitz, 2008). It may be, therefore, that the advantages of CBT over supportive and psychodynamic therapies derive in part from factors shared with IPT. The small number (2) of bona fide supportive therapy studies (rather than intent-to-fail conditions labeled "supportive") precludes clear interpretation, although further study is warranted.

4.3. By whom

Although examination of specific qualities of the therapist is beyond the scope of this review, the present data suggest that different research teams, using different strategies, can reach different conclusions. Methodological strength, variously described, was not associated with larger or smaller differences between CBT and other therapies. However, researcher allegiance does predict outcome. For the most part, comparative psychotherapy trials tend to be conducted by PIs, therapists, and research teams with a reportedly mild allegiance to CBT over the alternative therapy. The strength of PI allegiance is positively correlated with the strength of CBT's superiority; however, even when controlling for allegiance ratings CBT is still associated with a significant advantage. Therefore, the observed superiority of CBT cannot be attributed solely to researchers' theoretical allegiance.

4.4. For this individual

The present findings are equivocal regarding differential treatment response according to basic demographics. Due to the small number of studies, it is not clear whether CBT is superior to alternative therapies for child and adolescent patients. Unlike Wampold et al. (1997) previous meta-analysis, in which nonclinical student volunteers were oversampled (Crits-Christoph, 1997), the present results found an advantage for CBT with clinic patients and recruited patients, suggesting that distinct differences among psychotherapies might be most readily detected when presenting problems are in the clinical range.

4.5. With that specific problem

CBT's superiority is most clearly evident for patients with anxiety and depressive disorders. The small number of studies likely precludes clear understanding of how CBT compares to other therapies for personality disorders, although the preliminary results are at least suggestive of treatment specificity. It is noted that CBT has not been demonstrated to be superior to other therapies for patients with psychotic or substance use disorders, although there have been relatively few studies (1 and 3, respectively) of these groups.

Proponents of non-CBT psychotherapies (e.g., psychodynamic therapy) might argue that because CBT is more directly symptom-focused than are other therapies, the body of research using measures of primary symptom severity are inherently biased in favor of CBT. There is merit to this assertion, given the absence of significant findings in quality of life, self-concept, and social adjustment. However, it is important to note that other psychotherapies have not consistently demonstrated an advantage over CBT using any category of outcome measure, and that CBT appears to yield superior outcomes in general functioning measures as well as symptom-specific measures.

4.6. And under which set of circumstances

How outcome is defined also appears to impact findings of treatment specificity. CBT was demonstrably superior to other therapies when outcome was defined as reduction in symptoms of the primary illness or global symptoms, or as improvements in general functioning. When outcome is defined in terms of improvements in self-concept or social adjustment, however, CBT no longer appears superior. The meaning of this pattern is unclear, and further study is needed in order to determine more fully the parameters of CBT's efficacy. Most of the studies that assessed self-concept were comparisons of CBT vs. psychodynamic therapy, and it is possible that the relative emphasis on self-concept in that therapy (compared to the more symptom-focused emphasis in CBT) leads to greater improvements in self-concept. Other circumstances of the research studies were not demonstrated to impact the advantage of CBT over other therapies, although the small number of studies precludes clear interpretation in some cases. Although CBT fared well against alternative therapies in individual settings, it is less clear whether the same pattern holds for group therapy. The advantage of CBT over other therapies appears to have neither increased nor decreased over the years. One noteworthy null result is the absence of a relationship between treatment duration and differential treatment effect. Although the present results do not address the question of

whether longer CBT is more effective than is briefer CBT, they do suggest that the differential efficacy is unrelated to treatment duration: briefer CBT appears to be superior to briefer alternative psychotherapies, and longer CBT appears to be superior to longer alternative psychotherapies.

Acknowledgements

References

Alden, L. E. (1988). Behavioral self-management controlled-drinking strategies in a context of secondary prevention. *Journal of Consulting and Clinical Psychology*, *56*, 280—286.

Andrews, G., & Harvey, R. (1981). Does psychotherapy benefit neurotic patients? *Archives of General Psychiatry*, *38*, 1203—1208.

Areán, P. A., Perri, M. G., Nezu, A. M., Schein, R. L., Christopher, F., & Joseph, T. X. (1993). Comparative effectiveness of social problem-solving therapy and reminiscence therapy as treatments for depression in older adults. *Journal of Consulting and Clinical Psychology*, *61*, 1003—1010.

Barkham, M., Rees, A., Shapiro, D. A., Stiles, W. B., Agnew, R. M., Halstead, J., et al. (1996). Outcomes of time-limited psychotherapy in applied settings: Replicating the Second Sheffield Psychotherapy Project. *Journal of Consulting and Clinical Psychology*, *64*, 1079—1085.

Barkham, M., Shapiro, D. A., Hardy, G. E., & Rees, A. (1999). Psychotherapy in two-plus-one sessions: outcomes of a randomized controlled trial of cognitive–behavioral and psychodynamic–interpersonal therapy for subsyndromal depression. *Journal of Consulting and Clinical Psychology*, *67*, 201—211.

Beutler, L. E. (1991). Have all won and must all have prizes? Revisiting Luborsky et al'.s verdict. *Journal of Consulting and Clinical Psychology*, *59*, 226—232.

Beutler, L. E., Engle, D., Mohr, D., Daldrup, R. J., Bergan, J., Meredith, K., et al. (1991). Predictors of differential response to cognitive, experiential, and self-directed psychotherapeutic procedures. *Journal of Consulting and Clinical Psychology*, *59*, 333—340.

Bleiberg, K. L., & Markowitz, J. C. (2008). Interpersonal psychotherapy for depression. In D. H. Barlow (Ed.), *Clinical handbook of psychological disorders: A step-by-step treatment manual* (pp. 306—327)., 4th ed. New York: Guilford Press.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis.* Chichester, England: Wiley.

Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The empirical status of cognitive–behavioral therapy: A review of meta-analyses. *Clinical Psychology Review*, *26*, 17—31.

Carroll, K. M., Rounsaville, B. J., & Gawin, F. H. (1991). A comparative trial of psychotherapies for ambulatory cocaine abusers: Relapse prevention and inter-personal psychotherapy. *American Journal of Drug and Alcohol Abuse*, *17*, 229—247.

Carroll, K. M., Fenton, L. R., Ball, S. A., Nich, C., Frankforter, T. L., Shi, J., et al. (2004). Efficacy of disulfiram and cognitive behavior therapy in cocaine-dependent outpatients: A randomized placebo-controlled trial. *Archives of General Psychiatry*, *61*, 264—272.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*, 2 ed. Hillsdale, NJ: Lawrence Erlbaum Associates.

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

de Mello, M. F., de Jesus Mari, J., Bacaltchuk, J., Verdeli, H., & Neugebauer, R. (2005). A systematic review of research findings on the efficacy of interpersonal therapy for depressive disorders. *European Archives of Psychiatry and Clinical Neuroscience*, *255*, 75—82.

Dobson, K. S. (1989). A meta-analysis of the efficacy of cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, *57*, 414—419.

Driessen, E., Cuijpers, P., de Maat, S. C., Abbass, A. A., de Jonghe, F., & Dekker, J. J. (2010). The efficacy of short-term psychodynamic psychotherapy for depression: a meta-analysis. *Clinical Psychology Review*, *30*, 25—36.

Durham, R. C., Murphy, T., Allan, T., Richard, K., Treliving, L. R., & Fenton, G. W. (1994). Cognitive therapy, analytic psychotherapy and anxiety management training for generalised anxiety disorder. *British Journal of Psychiatry*, *165*, 315—323.

Durham, R. C., Fisher, P. L., Treliving, L. R., Hau, C. M., Richard, K., & Stewart, J. B. (1999). One year follow-up of cognitive therapy, analytic psychotherapy and anxiety management training for generalized anxiety disorder: Symptom change, medication usage and attitudes to treatment. *Behavioural and Cognitive Psychotherapy*, *27*, 19—35.

Durham, R. C., Guthrie, M., Morton, R. V., Reid, D. A., Treliving, L. R., Fowler, D., et al. (2003). Tayside-Fife clinical trial of cognitive–behavioural therapy for medication-resistant psychotic symptoms. Results to 3-month follow-up. *British Journal of Psychiatry*, *182*, 303—311.

Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., et al. (1989). National Institute of Mental Health Treatment of Depression Collaborative Research Program. General effectiveness of treatments. *Archives of General Psychiatry*, *46*, 971—982 discussion 983.

Emmelkamp, P. M., Benner, A., Kuipers, A., Feiertag, G. A., Koster, H. C., & van Apeldoorn, F. J. (2006). Comparison of brief dynamic and cognitive–behavioural therapies in avoidant personality disorder. *British Journal of Psychiatry*, *189*, 60—64.

Eysenck, H. J. (1994). The outcome problem in psychotherapy: What have we learned? *Behaviour Research and Therapy*, *32*, 477—495.

Fairburn, C. G., Kirk, J., O'Connor, M., & Cooper, P. J. (1986). A comparison of two psychological treatments for bulimia nervosa. *Behaviour Research and Therapy*, *24*, 629—643.

Foa, E. B., & Meadows, E. A. (1997). Psychosocial treatments for posttraumatic stress disorder: A critical review. *Annual Review of Psychology*, *48*, 449—480.

Gaffan, E. A., Tsaousis, I., & 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.

Gallagher-Thompson, D., & Steffen, A. M. (1994). Comparative effects of cognitive–behavioral and brief psychodynamic psychotherapies for depressed family caregivers. *Journal of Consulting and Clinical Psychology*, *62*, 543—549.

Garfield, S. L., & Bergin, A. E. (Eds.). (1986). *Handbook of psychotherapy and behavior change*, 3rd ed. . New York: Wiley.

Gelder, M. G., Marks, I. M., & Wolff, H. H. (1967). Desensitization and psychotherapy in the treatment of phobic states: A controlled inquiry. *British Journal of Psychiatry*, *113*, 53—81.

Giesen-Bloo, J., van Dyck, R., Spinhoven, P., van Tilburg, W., Dirksen, C., van Asselt, T., et al. (2006). Outpatient psychotherapy for borderline personality disorder: Randomized trial of schema-focused therapy vs transference-focused psychotherapy. *Archives of General Psychiatry*, *63*, 649—658.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* London: Sage Publications.

Goisman, R. M., Warshaw, M. G., & Keller, M. B. (1999). Psychosocial treatment prescriptions for generalized anxiety disorder, panic disorder, and social phobia, 1991–1996. *American Journal of Psychiatry*, *156*, 1819—1821.

Goldfried, M. R., & Davison, G. C. (1994). *Clinical behavior therapy.* New York: Wiley.

Hofmann, S. G., & Smits, J. A. J. (2008). Cognitive–behavioral therapy for adult anxiety disorders: A meta-analysis of randomized placebo-controlled trials. *Journal of Clinical Psychology*, *69*, 621—632.

Hollon, S. D. (1999). Allegiance effects in treatment research: A commentary. *Clinical Psychology: Science and Practice*, *6*, 107—112.

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

Hunsley, J., & Di Giulio, G. (2002). Dodo Bird, phoenix, or urban legend? The question of psychotherapy equivalence. *Scientific Review of Mental Health Practice*, *1*, 11—22.

Jadad, A. R., Moore, R. A., Carroll, D., Jenkinson, C., Reynolds, D. J., Gavaghan, D. J., et al. (1996). Assessing the quality of reports of randomized clinical trials: Is blinding necessary? *Controlled Clinical Trials*, *17*, 1—12.

Leichsenring, F., & Rabung, S. (2008). Effectiveness of long-term psychodynamic psychotherapy: A meta-analysis. *Journal of the American Medical Association*, *300*, 1551—1565.

Levy, K. N., Meehan, K. B., Kelly, K. M., Reynoso, J. S., Weber, M., Clarkin, J. F., et al. (2006). Change in attachment patterns and reflective function in a randomized control trial of transference-focused psychotherapy for borderline personality disorder. *Journal of Consulting and Clinical Psychology*, *74*, 1027—1040.

Leykin, Y., & DeRubeis, R. J. (2009). Allegiance in psychotherapy outcome research: Separating association from bias. *Clinical Psychology: Science and Practice*, *16*, 54—65.

Luborsky, L., Singer, B., & Luborsky, L. (1975). Comparative studies of psychotherapies. Is it true that "everyone has won and all must have prizes"? *Archives of General Psychiatry*, *32*, 995—1008.

Luborsky, L., Diguer, L., Seligman, D. A., Rosenthal, R., Krause, E. D., Johnson, S., et al. (1999). The researcher's own therapy allegiances: A "wild card" in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice*, *6*, 95—106.

Lynch, D., Laws, K. R., & McKenna, P. J. (2009). Cognitive behavioural therapy for major psychiatric disorder: Does it really work? A meta-analytical review of well-controlled trials. *Psychological Medicine*, 1—16.

McIntosh, V. V., Jordan, J., Carter, F. A., Luty, S. E., McKenzie, J. M., Bulik, C. M., et al. (2005). Three psychotherapies for anorexia nervosa: A randomized, controlled trial. *American Journal of Psychiatry*, *162*, 741—747.

McLean, P. D., & Hakistan, A. R. (1979). Clinical depression: Comparative efficacy of outpatient treatments. *Journal of Consulting and Clinical Psychology*, *47*, 818—836.

Miller, L. C., Barrett, C. L., Hampe, E., & Noble, H. (1972). Comparison of reciprocal inhibition, psychotherapy, and waiting list control for phobic children. *Journal of Abnormal Psychology*, *79*, 269—279.

Paul, G. L. (1967). Strategy of outcome research in psychotherapy. *Journal of Consulting Psychology*, *31*, 109—118.

Rachman, S., & Wilson, G. T. (1980). *The effects of psychological therapy.* Oxford: Pergamon Press.

Rosenthal, R., & Rubin, D. B. (1982). A simple, general purpose display of magnitude of experimental effect. *Journal of Educational Psychology*, *74*, 166—169.

Rosenthal, R., & Rubin, D. B. (1986). Meta-analytic procedures for combining studies with multiple effect sizes. *Psychological Bulletin*, *99*, 400—406.

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* London: Sage Publications.

Rosenthal, R. (1991). *Meta-analytic procedures for social research.* London: Sage Publications.

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.

Shapiro, D. A., & Shapiro, D. (1982). Meta-analysis of comparative therapy outcome studies: A replication and refinement. *Psychological Bulletin, 92,* 581−604.

Shapiro, D. A., Barkham, M., Rees, A., Hardy, G. E., Reynolds, S., & Startup, M. (1994). Effects of treatment duration and severity of depression on the effectiveness of cognitive–behavioral and psychodynamic–interpersonal psychotherapy. *Journal of Consulting and Clinical Psychology, 62,* 522−534.

Shapiro, D. A., Rees, A., Barkham, M., Hardy, G., Reynolds, S., & Startup, M. (1995). Effects of treatment duration and severity of depression on the maintenance of gains after cognitive–behavioral and psychodynamic–interpersonal psychotherapy. *Journal of Consulting and Clinical Psychology, 63,* 378−387.

Shear, M. K., Houck, P., Greeno, C., & Masters, S. (2001). Emotion-focused psychotherapy for patients with panic disorder. *American Journal of Psychiatry, 158,* 1993−1998.

Sloane, R. B., Staples, F. R., Cristol, A. H., & Yorkston, N. J. (1975). Short-term analytically oriented psychotherapy versus behavior therapy. *American Journal of Psychiatry, 132,* 373−377.

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

Svartberg, M., Stiles, T. C., & Seltzer, M. H. (2004). Randomized, controlled trial of the effectiveness of short-term dynamic psychotherapy and cognitive therapy for cluster C personality disorders. *American Journal of Psychiatry, 161,* 810−817.

Thompson, L. W., Gallagher, D., & Breckenridge, J. S. (1987). Comparative effectiveness of psychotherapies for depressed elders. *Journal of Consulting and Clinical Psychology, 55,* 385−390.

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

Ward, E., King, M., Lloyd, M., Bower, P., Sibbald, B., Farrelly, S., et al. (2000). Randomised controlled trial of non-directive counselling, cognitive–behaviour therapy, and usual general practitioner care for patients with depression. I: Clinical effectiveness. *British Medical Journal, 321,* 1383−1388.

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

Wilfley, D. E., Agras, W. S., Telch, C. F., Rossiter, E. M., Schneider, J. A., Cole, A. G., et al. (1993). Group cognitive–behavioral therapy and group interpersonal psychotherapy for the nonpurging bulimic individual: A controlled comparison. *Journal of Consulting and Clinical Psychology, 61,* 296−305.

Wilfley, D. E., Welch, R. R., Stein, R. I., Spurrell, E. B., Cohen, L. R., Saelens, B. E., et al. (2002). A randomized comparison of group cognitive–behavioral therapy and group interpersonal psychotherapy for the treatment of overweight individuals with binge-eating disorder. *Archives of General Psychiatry, 59,* 713−721.

Wilson, G. T., & Rachman, S. J. (1983). Meta-analysis and the evaluation of psychotherapy outcome: Limitations and liabilities. *Journal of Consulting and Clinical Psychology, 51,* 54−64.

Woody, G. E., Luborsky, L., McLellan, A. T., O'Brien, C. P., Beck, A. T., Blaine, J., et al. (1983). Psychotherapy for opiate addicts. Does it help? *Archives of General Psychiatry, 40,* 639−645.