

Effectiveness of and Dropout From Outpatient Cognitive Behavioral Therapy for Adult Unipolar Depression: A Meta-Analysis of Nonrandomized Effectiveness Studies

Eva Hans
Johannes Gutenberg University Mainz

Wolfgang Hiller
Johannes Gutenberg University Mainz

Objective: The primary aim of this study was to assess the overall effectiveness of and dropout from individual and group outpatient cognitive behavioral therapy (CBT) for adults with a primary diagnosis of unipolar depressive disorder in routine clinical practice. **Method:** We conducted a random effects meta-analysis of 34 nonrandomized effectiveness studies on outpatient individual and group CBT for adult unipolar depressive disorder. Standardized mean gain effect sizes are reported for end-of-treatment and 6-month follow-up effects for depression severity, dysfunctional cognitions, general anxiety, psychological distress, and functional impairment. The mean dropout rate from CBT is reported. We benchmarked our results against high-quality randomized controlled trials (RCTs). **Results:** Outpatient CBT was effective in reducing depressive severity in completer ($d = 1.13$) and intention-to-treat (ITT) samples ($d = 1.06$). Moderate to large posttreatment effect sizes ($d = 0.67$ – 0.88) were found for secondary outcomes. The weighted mean dropout rate was 24.63%. Posttreatment gains for depression were maintained at 6 months after completion of therapy. Effect sizes for depression were inferior to those of benchmark RCTs. **Conclusions:** Although clinical practice patients show lesser improvements in depressive symptoms than RCT patients, individual and group outpatient CBT can be effectively transported to routine clinical practice. The considerable treatment dropout rate, especially in individual CBT, must be improved. The small number of available studies and low quality of some reports stress the need for high-quality effectiveness studies.

Keywords: meta-analysis, effectiveness, cognitive behavioral therapy, unipolar depression, treatment outcome

Supplemental materials: <http://dx.doi.org/10.1037/a0031080.supp>

Several systematic reviews and meta-analyses of randomized controlled trials (RCTs) have shown cognitive behavioral therapy (CBT) to be effective for adult depression (Butler, Chapman, Forman, & Beck, 2006; Cuijpers, van Straten, Warmerdam, & Smits, 2008; Gloaguen, Cottraux, Cucherat, & Blackburn, 1998). Less is known about the differential benefits from individual and group therapy. Results of a comprehensive meta-analysis suggest that individual (mainly CBT) treatment may be more effective than group treatment and may also be associated with lower dropout rates in the treatment of depression (Cuijpers, van Straten, & Warmerdam, 2008). Moreover, there has repeatedly been found a psychotherapy dose–response relationship such that increasing the number of sessions provided is related to better treatment outcomes (Cuijpers, van

Straten, Schuurmans, et al., 2010; Cuijpers, van Straten, Warmerdam, & Smits, 2008).

RCTs are important to determine the efficacy of a treatment because they ensure high internal validity. However, results of RCTs do not provide information about how well the treatment tested under ideal research conditions can be transported to routine practice settings (Seligman, 1995; Westen & Morrison, 2001). Recruited patients who agree to randomization might be unrepresentative of patients actively seeking treatment in everyday practice (Buchkremer & Klingberg, 2001; Fydrich & Schneider, 2007; Seligman, 1995). For example, up to 76% of patients who presented for treatment of depression in clinical practice would have been excluded from RCTs (Stirman, DeRubeis, Crits-Christoph, & Brody, 2003; Stirman, DeRubeis, Crits-Christoph, & Rothman, 2005; Westen & Morrison, 2001) due to acute suicidality, subclinical diagnosis or insufficient severity of depression, comorbid psychiatric disorders, or adjunct treatments including antidepressants. Some researchers have suggested that the higher effect sizes for RCTs, in comparison to routine practice, may result from the exclusion of difficult-to-treat patients (Seligman, 1995; Westen & Morrison, 2001). Even so, when a routine practice sample of depressed adult outpatients was restricted according to the exclusion cri-

Eva Hans, Department of Clinical Psychology, Johannes Gutenberg University Mainz, Mainz, Germany; Wolfgang Hiller, Department of Clinical Psychology, Johannes Gutenberg University Mainz.

Correspondence concerning this article should be addressed to Eva Hans, Department of Clinical Psychology, Johannes Gutenberg University Mainz, Wallstrasse 3, 55122 Mainz, Germany. E-mail: hans@uni-mainz.de

teria commonly applied in RCTs, this was not found to have a significant influence on treatment effects (Schindler, Hiller, & Witthöft, 2011).

Apart from possibly negligible differences between patients, there remain considerable differences with regard to treatment settings. When providing treatment within the scope of an RCT, therapists are trained to adhere to a treatment manual and are extensively supervised and monitored. In contrast, in routine clinical practice, therapists report they never (47%) or rarely (22%) use treatment manuals (Addis & Krasnow, 2000). Treatment is also individually tailored, continuously adapted, and open-ended, usually resulting in a greater number of treatment sessions in comparison to RCTs.

Given concerns relating to the external validity of RCTs, Seligman and others (Chambless & Hollon, 1998; Chambless & Ollendick, 2001; Seligman, 1995) have emphasized the importance of clinical trials that are representative of routine clinical practice conditions, to complement results from RCTs. Seligman (1995) differentiated efficacy from effectiveness studies. Unlike effectiveness studies, he defined efficacy studies as RCTs with strict adherence to a treatment manual and a fixed number of sessions, along with highly selected patient samples to ensure high internal validity.

In the absence of a standard definition of an effectiveness study, criteria have been formulated to assess clinical representativeness on the efficacy–effectiveness continuum using the following six common criteria (Shadish, Matt, Navarro, & Phillips, 2000; Stewart & Chambless, 2009): (a) nonuniversity setting, (b) patient referral, (c) professional therapists, (d) flexible structure, (e) no monitoring of treatment implementation, and (f) no therapist training for study purposes. We agree with Stewart and Chambless (2009) that Shadish and colleagues' (2000) additional criteria of heterogeneous presenting problems, unlimited use of treatment procedures, and a flexible number of sessions are not clinically relevant. Most clinical effectiveness trials investigate a specific treatment for a particular disorder. Furthermore, treatment in everyday practice is restricted in duration by insurance coverage or financial constraints. Three additional criteria created by Stewart and Chambless best distinguished effectiveness from efficacy studies: no randomization, clinically representative rather than highly selected patients, and allowance of medication. In particular, randomized allocation of patients to an active versus another active or control condition is often not feasible in routine clinical practice for ethical and practical reasons (Seligman, 1995). Along with others (van Ingen, Freiheit, & Vye, 2009), we also consider randomization not to be representative of clinical practice.

The growing interest in the transportability of CBT to routine practice settings is also reflected in an increasing number of published effectiveness studies. Recently, two meta-analyses of effectiveness studies have shown the transportability of CBT for anxiety disorders to routine clinical settings. Stewart and Chambless (2009) conducted a meta-analysis of 56 effectiveness studies and found that CBT was highly effective in reducing anxiety disorder-specific symptoms ($d = 0.83$ – 2.59) as well as symptoms of both general anxiety ($d = 1.02$) and depression ($d = 0.73$ – 1.01). Van Ingen et al. (2009) included 11 effectiveness studies in their meta-analysis and reported mean pre–post effect sizes of $d = 1.35$ for anxiety and $d = 0.96$ for depression. These treatment gains were maintained during the 12-month follow-up. They also re-

ported an end-of-treatment mean dropout rate of 26.8% (range = 9.0–36.0). To see how treatment effects found in routine practice measure up to effects found in RCTs, a number of recent effectiveness studies for a variety of anxiety (Franklin, Abramowitz, Kozak, Levitt, & Foa, 2000; Hahlweg, Fiegenbaum, Frank, Schroeder, & von Witzleben, 2001; Lincoln et al., 2003; Stuart, Treat, & Wade, 2000; Wade, Treat, & Stuart, 1998), eating (Tuschen-Caffier, Pook, & Frank, 2001), and depressive disorders (Gibbons et al., 2010; Merrill, Tolbert, & Wade, 2003; Persons, Bostrom, & Bertagnolli, 1999; Schindler & Hiller, 2010; Westbrook & Kirk, 2005, 2007) have applied a benchmarking strategy. Outcomes of CBT for depressed adults in routine clinical practice were comparable with (Gibbons et al., 2010; Merrill et al., 2003; Persons et al., 1999; Westbrook & Kirk, 2005, 2007) or were inferior to efficacy benchmarks (Schindler & Hiller, 2010).

Despite the high relevance to clinical practice, to our knowledge, there has not been any previous meta-analysis performed to investigate the benefits of CBT for adult depression in routine clinical practice. There is also a paucity of research on the differential effectiveness of individual and group CBT. Finally, we were interested in whether a dose–response relationship similar to the one shown in previous research could be shown in routine practice. Accordingly, the primary aim of this study was to assess the overall effectiveness of outpatient CBT for adults with a primary diagnosis of unipolar depressive disorder at both end of treatment and at 6-month follow-up. We also estimated the mean dropout rate. To investigate if group CBT can be delivered as effectively as individual CBT and if dropout rates are similar between treatment formats in routine practice, we conducted subgroup analyses. A meta-regression analysis was performed to examine whether there is a dose–response relationship between the number of sessions delivered and the effect of CBT on depression. We also benchmarked our results against high-quality RCTs.

Method

Identification and Selection of Studies

Given the absence of any standard definition of the term effectiveness, we defined *effectiveness studies* as *nonrandomized* clinical trials representative of routine practice, as opposed to efficacy studies that are RCTs (Seligman, 1995). As the terms efficacy and effectiveness are inconsistently used by authors and unreliably indexed by electronic databases (Reeves, Deeks, Higgins, & Wells, 2011), we used a variety of (truncated) keywords identified from effectiveness studies and meta-analyses thereof (Stewart & Chambless, 2009; van Ingen et al., 2009) to locate nonrandomized clinical trials in routine clinical practice: *clinic setting, community mental health center, dissemination, effectiveness, generalizability, generalization, mental health field, naturalistic, nonrandomized, outpatient clinic, private practice, routine practice, service setting, transportability, and uncontrolled*. We developed a search strategy combining these keywords with subject headings relevant to unipolar depression and cognitive behavioral therapy (cf. Web Appendix A in the supplemental materials for exact search strings). The following electronic databases were searched from inception of each database to January 2012: MEDLINE via Ovid, PsycINFO, and PSYINDEXplus. We supplemented electronic searches with hand searches of three reviews including nonran-

domized clinical trials (Coelho, Canter, & Ernst, 2007; Cuijpers, 1998a; Oei & Dingle, 2008) and of the reference lists of all effectiveness studies located. To identify other published or unpublished trials, enquiries requesting potentially relevant effectiveness studies were sent to experts in the field.

Selection Criteria

Effectiveness studies were considered eligible if they examined the outcome of face-to-face CBT for depressed adult outpatients. CBT refers to cognitive, behavioral, or a combination of cognitive and behavioral therapy. As we were interested in the effects of CBT delivered in routine clinical practice, we excluded RCTs because randomization of patients was deemed unrepresentative of routine care. We also excluded studies in which less than half of the typical number of 12 CBT sessions was offered. In routine care, treatment is sometimes provided for depression and anxiety within the same program. Since our focus was on depression, we limited our meta-analysis to studies in which the majority of patients met the criteria for a primary diagnosis of major depressive disorder, minor depressive disorder, or dysthymic disorder. Our aim was to provide a benchmark for mental health services principally targeting the general adult population (aged 18–65 years). Studies exclusively targeting specific populations such as elderly or medically ill patients were therefore excluded. Finally, if authors provided insufficient quantitative data for effect size estimation, studies had to be excluded. We included both peer-reviewed and non-peer-reviewed studies (e.g., dissertations). No studies were excluded due to time or language restrictions.

Data Extraction

For each study included in the meta-analysis, we coded patient and intervention characteristics. To obtain some indication of the extent to which the effect sizes can be generalized and trusted, we coded and reported the clinical representativeness and methodological quality of the included studies. We used a standardized coding protocol (available from the authors).

Clinical representativeness was coded according to criteria adapted from Stewart and Chambless (2009) as well as Shadish et al. (2000): (a) referrals, (b) therapists, (c) structure, (d) monitoring, (e) therapist pretherapy training, (f) patients, and (g) allowance of medication. Their criteria for clinical representativeness were adapted in that the clinical representativeness items “referrals” (referred through usual clinical routes vs. some active recruiting), “structure” (flexible structure vs. strict adherence to a treatment manual), “monitoring” (no adherence monitoring other than routine supervision vs. extensive supervision and/or monitoring), and “training” (pretherapy training vs. no training) were dummy coded. Due to authors often not providing sufficient information to rate whether a mental health care setting was university affiliated, we excluded Stewart and Chambless’s criterion of a clinically representative setting. The criterion of randomization was not used because we considered the randomization of patients presenting for treatment to be unrepresentative of routine clinical practice (cf. inclusion criteria and Web Appendix B in the supplemental materials).

The methodological quality of included studies was assessed according to the guidelines for assessing methodological quality of

nonrandomized controlled studies of the Cochrane Consumers and Communication Review Group (CCCRG), which were modified for the present study. The original criteria of allocation concealment, blinding, and baseline comparability were not applicable to the design of before-and-after studies (Ryan et al., 2007). Additional criteria were developed based on a checklist of the German Academic Advisory Council for the quality assessment of clinical trials (Wissenschaftlicher Beirat Psychotherapie, 2010). Following the recommendations of the CCCRG, information on individual quality items is provided rather than an overall numerical quality score. If information was not available or was unclear, quality items were coded as “not available” (Ryan et al., 2007). The following resulting criteria were used: (a) low dropout rate (<30%), (b) use of intention-to-treat (ITT) analysis, (c) formal diagnostic investigation, (d) minimum within-study sample size of 30, and (e) minimum follow-up length of 6 months (cf. Web Appendix C).

The first author coded all the studies. The second author was trained in the use of the coding protocol and independently coded 20% of the studies. Interrater reliability for the coding of clinical representativeness and methodological quality of the studies was assessed using Cohen’s kappa score. The strength of interrater agreement on methodological quality was perfect ($\kappa = 1.00$). Except for one item (monitoring: $\kappa = .00$), interrater agreement on the criteria for clinical representativeness can be considered fair (patients: $\kappa = .25$) to perfect (all other items: $\kappa = 1.00$). Discrepancies in ratings were resolved by discussion until consensus was reached.

Effect Size Calculation and Statistical Procedures

The primary outcome was the end-of-treatment effect for depression severity, as variously assessed by self-report and clinician-rated depression-specific outcome measures (e.g., the Beck Depression Inventory [BDI]). Posttreatment improvement in depression severity was investigated separately for both completer and ITT samples. Secondary outcomes were dysfunctional cognitions, general anxiety, psychological distress, and functional impairment. Analyses on secondary outcomes were restricted to completer samples due to an insufficient number of studies providing ITT data.

Effect sizes for depressive severity from posttreatment to follow-up were based on completer samples due to a lack of ITT data ($k = 2$). A nonsignificant result would suggest that end-of-treatment gains were retained for 6 months. We restricted analyses to subsamples of patients with complete data sets at posttreatment and follow-up. Upon request, two of the authors contacted provided additional data (Matsunaga et al., 2010; Schindler & Hiller, 2010), whereas others indicated they no longer had access to the data (Gelhart & King, 2001; Peterson & Halstead, 1998). For the three available studies providing data at 12-month follow-up, single within-study effect sizes were reported (Piacentini et al., 2011; Schindler & Hiller, 2010; Scott, 2005). The minimum follow-up period was set at 6 months after the end of treatment. Four studies had only followed up patients from between 1 to 3 months (J. S. Brown et al., 2011; Kingston, Dooley, Bates, Lawlor, & Malone, 2007; Nisbet Wallis, 2002; Swan, Sorrell, MacVicar, Durham, & Matthews, 2004) and were therefore not considered.

This meta-analysis followed the procedures for within-group effect size calculation outlined by Lipsey and Wilson (2001). Becker's (1988) standardized mean gain effect size statistic, using the pooled standard deviation, was calculated from means and standard deviations following the recommendations for within-group analyses (Dunlap, Cortina, Vaslow, & Burke, 1996; Maier-Riehle & Zwingmann, 2000). If insufficient information was reported, data were requested from corresponding study authors. Given the small sample sizes of some studies, introducing potential upward bias, a small sample bias correction was applied to each effect size (Hedges & Olkin, 1985).

In considering that between-studies heterogeneity was to be expected, a random effects analysis was deemed most appropriate for pooling effect sizes. The inverse variance weighted overall effect size and the respective 95% confidence interval were calculated for each outcome across studies (Lipsey & Wilson, 2001) if sufficient data (five or more within-study samples) were available. As the correlation between pre- and posttest scores (required for the calculation of the standard errors and inverse variance weights) is not commonly reported, we assumed a value of zero for this correlation, which leads to the most conservative (large) estimate of the standard error of each effect size. A *z* test was used to evaluate whether the overall effect sizes were significantly different from zero. Effect sizes were pooled using Lipsey and Wilson's (2001) SPSS mean effect size macro. According to Cohen's (1988, p. 40) hypothetical rule of thumb, effect sizes of 0.2, 0.5, and 0.8 were considered small, moderate, and large, respectively. Lipsey's (1990, p. 56) empirically based rule of thumb compares with Cohen's guideline, with effect sizes of less or equal to 0.32 being considered small, 0.33 to 0.55 moderate, and 0.56 to 1.2 large.

Subsample effect sizes were calculated if results were reported separately for individual and group therapy (J. S. Brown et al., 2011; Craigie & Nathan, 2009; Scott, 2005) as these can be considered statistically independent (Lipsey & Wilson, 2001) and also because we were interested in differences across treatment formats. In order to reduce the risk of artificial reduction of heterogeneity, we included one weighted mean effect size per study if results were reported separately for nonoverlapping (e.g., diagnostic) subsamples (Forand, Evans, Haglin, & Fishman, 2011; Gelhart & King, 2001; Haaga, DeRubeis, Stewart, & Beck, 1991; Oei & Yeoh, 1999; Persons et al., 1999). Similarly, in case of multiple outcome measures, effect sizes were averaged together to yield one effect size per outcome for each within-study sample. Outliers that were more than 3 *SDs* from the mean of effect sizes for depression severity (McCullough, 1991; Nalini, Kumaraiah, & Subbakrishna, 1996), dysfunctional cognitions (Nalini et al., 1996), and psychological distress (Scott, 2005: individual format) were removed from respective analyses and overall mean effect sizes were recalculated based on the remaining effect sizes (Lipsey & Wilson, 2001).

Homogeneity Analysis

Heterogeneity was assessed using the *Q* test (Hedges & Olkin, 1985) for evidence of statistically significant heterogeneity. This statistical test has low power given small numbers of studies with small samples (Hardy & Thompson, 1998). Therefore, we additionally computed the test statistic I^2 (Higgins, Thompson, Deeks,

& Altman, 2003) to quantify the amount of heterogeneity. It gives values ranging between 0% and 100%, with larger values indicating greater heterogeneity. According to Higgins and Thompson's (2004) rule of thumb, an I^2 of 25%, 50%, and 75% is considered small, moderate, and large, respectively.

Dropout

Most studies defined completion as the termination of treatment in accordance with therapist advice. A few studies defined completers as those having received equal or more than a predefined minimum number of sessions (Kingston et al., 2007; Kwon & Oei, 2003; Marshall & Mazie, 1987; Scott, 2005). We defined and computed dropout rates as the proportion of patients who have attended at least one but fewer than the number of sessions necessary to complete treatment.

For the purpose of meta-analysis of dropout from CBT, we followed the procedures described by Lipsey and Wilson (2001) to compute logit transformed effect sizes for proportions. Lipsey and Wilson's SPSS mean effect size macro was used for pooling effect sizes, using random effects analysis. The inverse variance weighted mean logit effect size for proportions and its corresponding confidence interval were back-transformed into proportions to ease interpretation.

Moderator Analyses

To investigate statistical heterogeneity, a priori specified subgroup and meta-regression analyses were performed. We used mixed effects analogs to the analysis of variance (ANOVA) to compare the effectiveness of individual versus group CBT at reducing depressive severity, dysfunctional cognitions, general anxiety, psychological distress, and functional impairment at end of treatment, as available for completer and ITT samples. We also compared the dropout rates across treatment formats (individual vs. group).

Mixed effects weighted meta-regression was used to investigate the association between the number of sessions delivered and the end-of-treatment effect on depression (Higgins & Thompson, 2004; Thompson & Higgins, 2002). Analogs to the analysis of variance (ANOVA) and meta-regression analyses were performed using Lipsey and Wilson's (2001) analog to the ANOVA and maximum likelihood meta-regression SPSS macro, respectively.

Selection of Benchmark Studies

Benchmark RCTs meeting our inclusion criteria were retrieved from a recent meta-analytic database of 281 RCTs on psychotherapy for adult depression (Cuijpers, van Straten, van Oppen, Andersson, & Hollon, 2010) and selected based on quality criteria outlined by the providers of that database (Cuijpers, van Straten, Bohlmeijer, Hollon, & Andersson, 2010). To avoid bias due to differences in specificity and reactivity across outcome measures (Minami, Wampold, Serlin, Kircher, & Brown, 2007), we restricted benchmark comparisons to the BDI as it was by far the most widely used outcome measure. In case of multiple CBT treatment conditions by treatment format, we calculated a weighted mean within-group effect size per RCT.

Power Calculations

Analogs to the ANOVA tend to have low statistical power under the mixed effects model. To determine whether we had sufficient power to detect small (0.32), medium (0.55), and large (1.20) moderator effect sizes (Lipsey, 1990, p. 56), we conducted retrospective power analyses according to Hedges and Pigott (2004) using the effect size estimates specified above and the observed (mixed model) within-group variances for each group.

Assessment of Reporting Bias

To identify potential small-study effects, funnel plots were visually inspected for evidence of asymmetry. In the absence of bias, the scatterplot of effect size against precision (inverse of standard error) should look like a symmetrical inverted funnel. Egger's linear regression method (Egger, Davey Smith, Schneider, & Minder, 1997; Sterne & Egger, 2006) was used to statistically test for funnel plot asymmetry. A significant deviation of the intercept of the (unweighted) regression of standard normal deviates on their precision from zero indicates the presence of bias. It is important to note, however, that unless the bias is substantial and more than 20 studies of varying sample sizes are included in the meta-analysis, the sensitivity of this test will be low (Sterne, Gavaghan, & Egger, 2000). Additionally, Orwin's (1983) Fail-safe N , the number of studies with null effect that would need to exist in order to reduce the significant overall effect size to 0.40, was determined. We chose this medium criterion effect size based on a wait-list effect size reported in a recent meta-analysis (Minami et al., 2007).

Results

Characteristics of Included Studies

A total of 30 effectiveness studies were retrieved from the electronic databases MEDLINE, PsycINFO, and PSYNDEXplus. Six additional studies were located through hand searching of reference lists of the included articles and review articles or were identified by experts. A flow diagram is provided to describe the inclusion process (see Web Figure 1 in the supplemental materials). The 34 studies (excluding two outlier studies) included in the meta-analysis are indicated by an asterisk in the reference section.

Sample Description

Effect sizes of CBT for reducing depressive severity were derived from all 34 studies and included a total of 1,880 patients who completed ($k = 31$) and 1,629 patients ($k = 15$) who were intended to be treated with CBT (see Web Tables 1 and 2 in the supplemental materials). The majority of the completer (weighted $M = 68.61\%$, $SD = 9.57$) and the ITT samples (weighted $M = 66.37\%$, $SD = 10.21$) were women, with weighted mean ages of 38.58 ($SD = 4.84$) and 37.38 years ($SD = 4.85$), respectively. Major depression was the most common disorder in both the completer (weighted $M = 79.71\%$, $SD = 18.70$) and intent-to-treat samples (weighted $M = 92.66\%$, $SD = 9.02$). If reported, comorbidity of Axis I (weighted $M = 38.56\%$, $SD = 19.08$ and weighted $M = 57.10\%$, $SD = 10.03$) and Axis II disorders (weighted $M =$

17.15%, $SD = 14.28$ and weighted $M = 34.53\%$, $SD = 20.87$) was common among the completer and ITT samples, respectively. The majority of completers used antidepressant medication (weighted $M = 58.44\%$, $SD = 23.56$); of the ITT sample, every other patient was medicated with an antidepressant (weighted $M = 49.91\%$, $SD = 20.63$).

Treatment

Completers were provided an average number of 21.71 individual ($Min = 10.64$, $Max = 39.80$, $SD = 11.87$) or 11.18 group therapy sessions ($Min = 6.00$, $Max = 24.00$, $SD = 3.76$). In two studies not differentiating between treatment formats, patients were offered up to 20 sessions individual CBT, group CBT, or a combination of both (Organista, Munoz, & Gonzalez, 1994: $M = 15.50$; Piacentini et al., 2011: range = 10–15). In studies reporting ITT analyses, patients received an average of 18.99 ($Min = 8.36$, $Max = 34.80$, $SD = 11.32$) individual or 8.78 ($Min = 7.13$, $Max = 12.00$, $SD = 2.18$) group therapy sessions. In three studies, patients received up to 20 individual sessions, group sessions, or a combination of both (Organista et al., 1994: $M = 10.00$; Piacentini et al., 2011: range = 10–15; Scott, 2005: $M = 9.13$). Individual and group sessions lasted 50 to 60 and 90 to 120 min, respectively.

Clinical Representativeness of Included Studies

Most studies indicated that their patients had been referred through usual clinical routes (65.63%) rather than additionally or exclusively solicited for participation (34.38%). In almost all studies (91.18%), common exclusion criteria for admission to routine outpatient treatment were applied. Treatment was predominantly provided by practicing therapists and therapists in training (93.10%; research therapists: 6.90%) of whom the majority (79.14%) were not specifically trained for study purposes. In all studies evaluating the effectiveness of individual therapy, treatment manuals were flexibly used, whereas all group therapy studies strictly manualized their treatments. Most studies did not formally assess therapist adherence (76.47%). In some studies, however, monitoring of treatment was implemented through either one or both extensive supervision and formal adherence checks (23.53%). In all studies, patients were allowed to use psychiatric medication.

Methodological Quality of Included Studies

Structured or semistructured validated diagnostic interviews or a diagnostic checklist was administered in only 15 out of 34 studies (44.12%). In the remaining studies, no diagnostic instrument was specified or diagnoses were based on clinical judgment. Effect sizes of 26 studies (76.47%) were derived from sample sizes greater than or equal to 30. Dropout data could be extracted from 23 studies (67.65%), less than half of which (47.83%) reported low dropout rates (<30%). More than a third of studies (41.18%) provided results based on ITT analyses. In 10 out of 34 studies (29.41%), patients were followed up for 6 months or more post-treatment; however, in two of the studies, insufficient information was provided and was unavailable upon request for the estimation of post-follow-up effect sizes (see Web Tables 1 and 2 in the supplemental materials).

End-of-Treatment Improvement in Primary and Secondary Outcomes

We found large pre–post effect sizes for improvement in depressive severity for completer ($d = 1.13$, 95% CI [1.02, 1.24]; $Q[30] = 54.44$, $p < .01$; see Figure 1 & Table 1) and ITT samples ($d = 1.06$, 95% CI [0.94, 1.18]; $Q(14) = 28.26$, $p < .05$; see Figure 2 & Table 1), respectively. There was moderate heterogeneity for completer and ITT analyses, respectively (see Table 1). Eight studies reported both completer and ITT analyses. As was to be expected, an additional analysis of these eight studies showed that the effect size for depression based on completer ($d = 1.34$, 95% CI [1.15, 1.52]) was significantly larger relative to the corresponding effect size based on ITT samples ($d = 1.04$, 95% CI [0.87, 1.21]; $Q_B[1] = 5.39$, $p < .05$).

There is a paucity of ITT data for secondary outcome measures. For completer analyses, moderate to large posttreatment effect sizes ($d = 0.67$ – 0.88) were found for patients' improvement in dysfunctional cognitions, general anxiety, psychological distress, and functional impairment. All effect sizes reached statistical significance ($p < .001$; see Table 1). There was low heterogeneity for dysfunctional cognitions and psychological distress. Heterogeneity was found to be low to moderate for general anxiety and functional impairment (see Table 1).

Overall, 23 studies reported posttreatment dropout rates restricted to those treated for depression. The weighted mean drop-

out rate was 24.63% (95% CI [17.45, 33.58]) with a wide range from 0% to 68%.

Maintenance of Treatment Gains

Few data were reported at 6 ($k = 7$) and 12 months ($k = 3$) posttreatment. When considering studies providing data at posttreatment and the respective follow-up period, posttreatment gains were maintained or even improved at 6 ($d = 0.02$, 95% CI [–0.19, 0.23]; $Q[6] = 7.31$, $p = .29$, $I^2 = 17.93\%$) and 12 months after completion of therapy (Matsunaga et al., 2010: $d = 0.72$, 95% CI [0.07, 1.38]; Piacentini et al., 2011: $d = 0.25$, 95% CI [–0.14, 0.64]; Schindler & Hiller, 2010: $d = 0.02$, 95% CI [–0.33, 0.36]), as indicated by the nonsignificant and positive post-follow-up effect sizes.

Moderator Analyses

Differences between the effectiveness of individual (completer: $d = 1.25$, 95% CI [1.05, 1.44]; ITT: $d = 1.11$, 95% CI [0.99, 1.23]) and group CBT in reducing depressive severity at termination of treatment (completer: $d = 1.06$, 95% CI [0.92, 1.20]; ITT: $d = 0.91$, 95% CI [0.71, 1.10]) fell short of significance for completer ($Q_B[1] = 2.48$, $p = .12$) and ITT analyses ($Q_B[1] = 2.97$, $p = .08$; see Table 1). The statistically nonsignificant results should be regarded as inconclusive, as post hoc power calculations indicated that the statistical power was insufficient to detect small

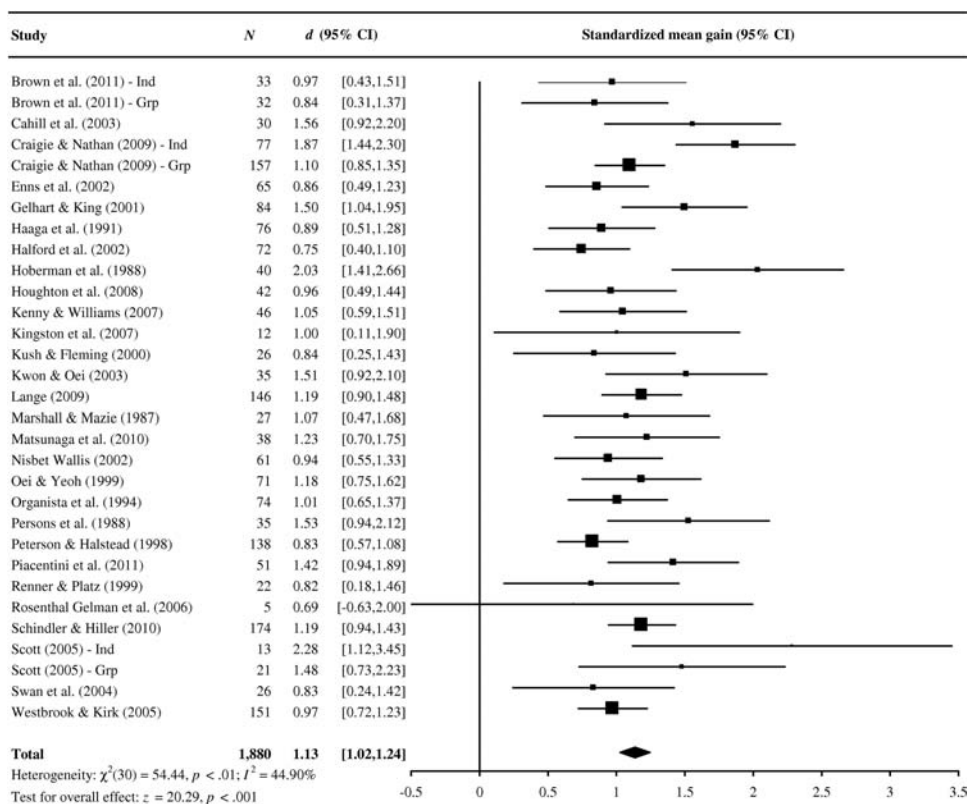


Figure 1. Forest plot of 31 completer effect sizes of cognitive behavioral therapy (CBT) for depression severity at posttreatment, using the random effects method. CI = confidence interval; Grp = group therapy; Ind = individual therapy.

Table 1
Pre-Post Effect Sizes by Outcome and Treatment Format

Variable	<i>k</i>	Statistics in subsamples					ANOVA analog		
		<i>d</i>	95% CI	<i>z</i>	<i>Q</i>	<i>I</i> ²	<i>Q</i> _B ^a	<i>Q</i> _w	<i>df</i>
Depressive severity									
Completer ^b	31	1.13	1.02, 1.24	20.29***	54.44**	44.90	2.48	29.21	27
Individual	9	1.25	1.05, 1.44	12.71***					
Group	20	1.06	0.92, 1.20	14.71***					
Intention-to-treat ^b	15	1.06	0.94, 1.18	17.10***	28.26*	50.46	2.97	9.01	10
Individual	8	1.11	0.99, 1.23	17.78***					
Group	4	0.91	0.71, 1.10	8.91***					
Dysfunctional cognitions									
Completer	7	0.79	0.54, 1.04	6.20***	7.82	23.27	—	—	
Individual	1	1.83	0.79, 2.87	3.44***					
Group	6	0.74	0.53, 0.96	6.77***					
General anxiety									
Completer	7	0.67	0.50, 0.84	7.88***	10.03	40.19	2.34	5.01	5
Individual	2	0.84	0.57, 1.11	6.18***					
Group	5	0.59	0.40, 0.77	6.20***					
Psychological distress									
Completer	8	0.88	0.75, 1.01	13.00***	5.76	0.00	0.73	5.03	6
Individual	3	0.92	0.76, 1.08	11.05***					
Group	5	0.80	0.57, 1.03	6.90***					
Functional impairment									
Completer	6	0.84	0.63, 1.04	8.12***	10.21	51.03	4.62*	3.97	4
Individual	2	1.05	0.81, 1.28	8.79***					
Group	4	0.71	0.51, 0.91	7.05***					

Note. *k* indicates the number of pooled effect sizes. Dashes indicate that data could not be computed due to insufficient sample sizes. CI = confidence interval; ANOVA = analysis of variance.

^a *df* for *Q*_B = 1. ^b The sum of the breakdown of effect sizes by treatment format does not add up to the total because a few studies provided combined treatment and were therefore not included in the comparison.

* *p* < .05. ** *p* < .01. *** *p* < .001.

(completer: $1-\beta = 0.11$; ITT: $1-\beta = 0.21$) or moderate effects (completer: $1-\beta = 0.23$; ITT: $1-\beta = 0.52$). Similarly, posttreatment gains with regard to general anxiety ($d = 0.84$, 95% CI [0.57, 1.11]) and psychological distress ($d = 0.92$, 95% CI [0.76, 1.08]) achieved in individual CBT were not statistically different from those achieved in group therapy (general anxiety: $d = 0.59$, 95% CI [0.40, 0.77]; $Q_B[1] = 2.34$, $p = .13$; psychological distress: $d = 0.80$, 95% CI [0.57, 1.03]; $Q_B[1] = 0.73$, $p = .39$). However, due to the small number of studies and the observed low statistical power for the nonsignificant findings (general anxiety: $1-\beta = 0.20$ and $1-\beta = 0.49$ for small and moderate effects; psychological distress: $1-\beta = 0.19$ and $1-\beta = 0.46$ for small and moderate effects), results should be considered inconclusive as to whether there is a small or moderate effect (see Table 1). Patients who attended individual CBT ($d = 1.05$, 95% CI [0.81, 1.28]) reported greater improvements in functioning than those who attended group CBT ($d = 0.71$, 95% CI [0.51, 0.91]; $Q_B[1] = 4.62$, $p < .05$; see Table 1).

Given potential confounding by study design and sample characteristics, we compared individual versus group CBT on the basis of the three studies providing both individual and group treatment. The difference between individual ($d = 1.58$, 95% CI [1.05, 2.12]) and group CBT ($d = 1.10$, 95% CI [0.63, 1.58]) regarding the treatment effect on depressive symptoms at posttreatment fell short of significance ($Q_B[1] = 1.74$, $p = .19$). Caution is given when interpreting this result due to the small number of studies.

We found the proportion of patients failing to complete treatment to be higher in individual ($M = 42.00\%$, 95% CI [26.32, 59.49], $Min = 24.02$, $Max = 68.00$, $k = 8$) than in group CBT ($M = 16.70\%$, 95% CI [10.85, 24.83], $Min = 2.00$, $Max = 59.21$, $k = 16$; $Q_B[1] = 8.46$, $p < .01$). Our meta-regression analysis did not reveal a significant relationship between the number of sessions delivered to completers and the effect of CBT on depression ($Q[1] = 0.06$, $p = .80$).

Benchmarking the Effectiveness of CBT for Depression

We selected high-quality benchmark RCTs for each treatment format against which we compared the effectiveness of CBT in reducing depressive severity as achieved in clinical practice. The pre-post efficacy effect sizes of completers-only and ITT analyses can be discerned from Table 2. The effect of individual CBT in clinical practice was inferior to effects in the benchmark RCTs despite clinical practice patients attending more individual sessions than those treated in the RCTs. Similarly, effects of group therapy were smaller in effectiveness versus efficacy studies. Among completers, the total number of attended group sessions was comparable between treatment settings, whereas patients intended to be treated in clinical practice attended fewer group sessions than those treated in research settings. Dropout rates varied greatly between RCTs, ranging from a dropout rate of 8.05% to 32.20% and 3.13% to 40.00% for

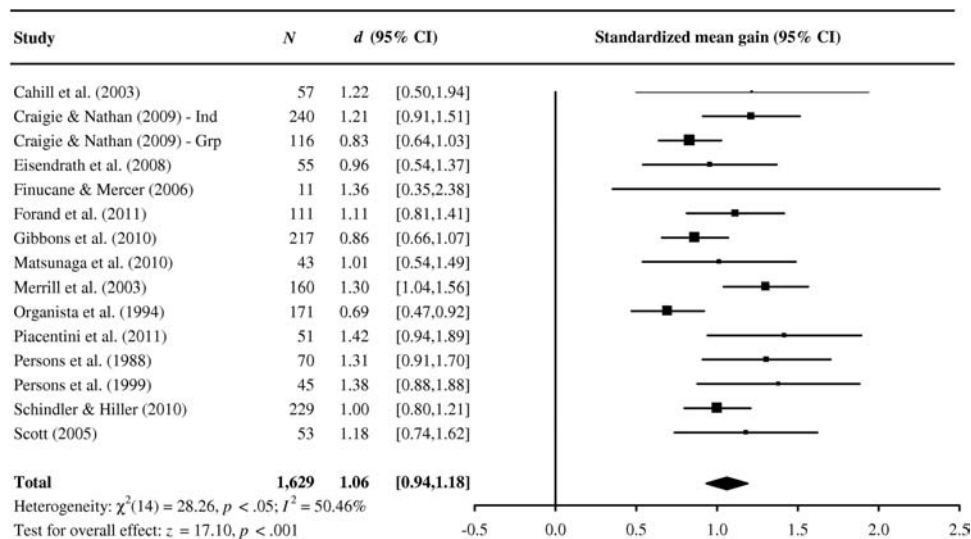


Figure 2. Forest plot of 15 intention-to-treat effect sizes of cognitive behavioral therapy (CBT) for depression severity at posttreatment, using the random effects method. CI = confidence interval; Grp = group therapy; Ind = individual therapy.

individual and group CBT, respectively. While the average dropout rate in individual CBT in clinical practice was higher compared to the benchmark RCTs, dropout rates in group CBT in clinical practice and research settings were comparable (see Table 2).

Assessment of Reporting Bias

There was no evidence of small-study bias for completer effect sizes based on the visual inspection of the funnel plot and the Egger's test (intercept 1.11, 95% CI [-0.18, 2.40], $t = 1.75$, two-tailed $p = .09$) increasing our confidence that the effect of CBT for adult depression is not overestimated. However, there might have been a problem with small-study bias for ITT effect sizes. Less precise studies accumulate on the right-hand side of the funnel; that is, as studies are less precise, ITT effect sizes get bigger. Egger's test was statistically significant (intercept 2.09; 95% CI [-0.40, 3.78], $t = 2.68$, two-tailed $p < .05$), suggesting studies with smaller sample sizes reported higher effect sizes. However, according to Orwin's (1983) Fail-safe N , 55 and 23 unpublished studies with nonsignificant findings would be necessary in order to reduce the completer and ITT effect size, respectively, to the criterion effect size of 0.40. This moderate criterion effect size yields a conservative estimate of the Fail-safe N . Hence, given the small number of effectiveness studies identified, the relatively large Fail-safe N s suggest effect sizes were robust against the file drawer problem.

Discussion

The findings of this meta-analysis provide support for the effectiveness of outpatient CBT in routine clinical practice. Completers reported a substantial reduction in depression at end of treatment. The available evidence suggests that these treatment gains were maintained for at least 6 months after completion of therapy. Furthermore, the results show that complet-

ing CBT can also improve dysfunctional cognitions, general anxiety, psychological distress, and functional impairment, with moderate to large effect sizes. Despite favorable treatment effects for those completing treatment, dropout from treatment was considerable. Among patients who started CBT for depression, on average, every fourth person failed to complete therapy. This figure compares favorably to dropout rates from CBT for (mainly) anxiety disorders previously found in routine outpatient settings (Bados, Balaguer, & Saldaña, 2007; van Ingen et al., 2009). The high dropout rates from CBT for depression challenge treatment outcomes when only completers are analyzed. However, when ITT analyses were performed, the overall end-of-treatment effect for depression was still large, albeit smaller than for completers-only analyses.

Patients largely benefited from both individual and group CBT. Even though the average individually treated patient attended twice as many sessions as a patient who attended group-based CBT, we noted only a trend toward larger effects of individual CBT but failed to find any significant differences in the reduction of depression, general anxiety, and psychological distress between treatment formats. However, patients who completed the extended individual CBT achieved better functioning at end of treatment than those who completed group CBT. Similarly, exploratory results based on the three effectiveness studies directly comparing individual and group CBT revealed no significant difference between the effect of individual and group CBT on depression. Findings regarding the relative effectiveness of individual versus group CBT should be regarded as preliminary as they were based on a small number of studies and as the power was insufficient to detect small or moderate effects.

The individually treated patient, however, needs more sessions to complete treatment in routine practice than the one treated in group. There are several possible explanations for this, such as that group therapy may work better or faster than

Table 2

Benchmark Comparisons With Efficacy Studies: Pre-Post Effect Sizes for the Reduction of Depressive Severity by Treatment Format

Study	Format	Therapy type	Completer					Intention-to-treat					%Dropout
			<i>k</i>	<i>N</i>	<i>d</i>	95% CI	Sessions	<i>k</i>	<i>N</i>	<i>d</i>	95% CI	Sessions	
Effectiveness	Ind	CBT	9		1.26	1.05, 1.46	21.71	8	1.12	0.98, 1.26	18.99	42.00	
Dimidjian et al. (2006)	Ind	CT, BA	—	—	—	—	—	64	2.22	1.34, 3.11	≤24.00	10.71	
Elkin et al. (1989)	Ind	CBT	37	1.90	1.27, 2.53	16.20	59	1.44	0.99, 1.88	13.00 ^a	32.20		
Jacobson et al. (1996)	Ind	BA, AT, CBT	137	2.60	2.21, 2.99	12.00–20.00	149	2.46	2.10, 2.82	<12.00	8.05		
Jarrett et al. (1999)	Ind	CT	—	—	—	—	—	36	1.80	1.18, 2.42	17.40	13.89	
King et al. (2000)	Ind	CBT	117	1.66	1.33, 2.00	6.00–12.00	134	1.28	0.99, 1.56	≤12.00	12.69		
Effectiveness	Grp	CBT	18	1.06	0.92, 1.20	11.18	3	0.87	0.70, 1.04	8.78	16.70		
Allart-van Dam et al. (2003)	Grp	CBT	—	—	—	—	—	61	0.87	0.48, 1.25	9.39	17.74	
Beutler et al. (1991)	Grp	CT	—	—	—	—	—	21	1.29	0.57, 2.01	20.00	40.00	
R. A. Brown & Lewinsohn (1984)	Grp	CBT	25	1.28	0.62, 1.94	12.00	—	—	—	—	—	3.13	
Neimeyer et al. (2008)	Grp	CBT	35	1.19	0.64, 1.73	10.00	—	—	—	—	—	23.91	
Teri & Lewinsohn (1986)	Grp	CBT	47	1.93	1.37, 2.49	12.00	—	—	—	—	—	8.00	

Note. Dashes indicate data are not available. *k* indicates the number of pooled effect sizes. AT = automatic thoughts; BA = behavioral activation; CBT = cognitive behavioral therapy; CI = confidence interval; CT = cognitive therapy; %Dropout = proportion of patients who have attended at least one but fewer than the number of sessions necessary to complete treatment; Grp = group therapy; Ind = individual therapy.

^a Based on total sample.

individual therapy or perhaps therapists may not advise group therapy for the more severely disturbed patients. This may also help to understand the discrepant finding of an earlier meta-analysis of RCTs that individual CBT is more effective when directly compared to group CBT, as randomization ensures that similar patients are treated in either treatment format (Cuijpers, van Straten, & Warmerdam, 2008). More definitive evidence on the differential effectiveness of individual and group CBT is needed and is also highly relevant for clinical practice because group CBT might be a cost-effective, rapidly available alternative to individual CBT.

Especially striking is the comparison of dropout rates between individual and group CBT with dropout rates in routine practice settings being more than twice as high for individual relative to group CBT. This finding stands in contrast to a previous meta-analysis of RCTs suggesting that the dropout rate is lower in individual compared to (the same dose of) group therapy for depression (Cuijpers, van Straten, & Warmerdam, 2008). These inconsistent findings may in part be explained by the larger number of individual compared to group CBT sessions such that the dropout rate increases as the number of sessions provided increase. Please note, however, that comparisons between treatment formats within the same studies are much better suited than comparisons across studies to comparing differences between individual and group therapy. Contrary to previous research (Cuijpers, van Straten, Schuurmans, et al., 2010; Cuijpers, van Straten, Warmerdam, & Smits, 2008), there was no significant dose-response relationship between the number of sessions and the CBT effect on depression, which

may be due to the small number of studies available. Another possible reason is that a significant relationship could not be found because number of sessions is an aggregate statistic that is prone to obliterate differences in meta-regression (Thompson & Higgins, 2002). Thus, this question would be most appropriately addressed by a large-scale primary study.

When benchmarking our results against high-quality RCTs, we found that completers and ITT patients under clinical practice conditions achieved less improvement in depression than those in the RCT benchmarks at overall longer (individual CBT) or comparable (group CBT) treatment duration. However, we note that more heterogeneous samples of depressed patients entering routine practice increased the within-group variance, which may, in part, have contributed to the decreased effect sizes of CBT for depression being due to statistical reasons. Dropout rates from individual and group therapy in clinical practice were higher or similar to those reported in efficacy studies. Having more treatment sessions has been associated with a higher dropout rate in an earlier RCT of depressed elderly (Cuijpers, 1998b) and might be one factor contributing to the high rates of dropout from individual CBT in routine clinical practice. To summarize, we conclude that outpatient CBT can be effectively delivered in clinical practice settings, albeit somewhat less effectively than in research settings. Moreover, individually treated clinical practice patients were provided more CBT sessions but also dropped out at a higher rate compared to RCT patients.

There are several limitations that need to be acknowledged and addressed regarding the present meta-analysis. First, results should be regarded as preliminary due to the small number of effective-

ness studies available and the statistical imprecision of effect sizes due to heterogeneous samples in clinical practice, which also decreased the power to test for moderator effects using a random effects model.

Second, most studies reported only completer but no ITT data. Effect sizes based on completers are likely to be upwardly biased assuming that those patients systematically dropout who do not benefit or deteriorate from treatment. It is therefore important to consider dropout rates when interpreting effect sizes of completers-only analyses. Further, results according to ITT should routinely be reported.

Third, we must consider threats to internal validity due to confounding variables. Only one included effectiveness study used a wait-list control group (Schindler & Hiller, 2010). The present meta-analysis was therefore based on within-group effect sizes that do not control for either the passage of time, social demand effects, regression to the mean, or other potential confounding variables that could have accounted for the improvement in depression and secondary outcomes. Previous meta-analyses have examined changes in depressive symptoms in adult outpatients randomized to (no treatment) waiting lists of psychotherapy studies. Over a waiting list period of 4 to 15 and 5 to 12 weeks, respectively, a weighted mean reduction in BDI scores of 15.7% (Posternak & Miller, 2001) and a mean within-group effect size of approximately 0.5 were found (Rutherford, Mori, Sneed, Pimontel, & Roose, 2012). However, depressive symptom improvement in adult depressed outpatients who had been on psychotherapy wait-lists of two recent effectiveness studies for up to 12 months has been found to be negligible to small (Barkham, Mullin, Leach, Stiles, & Lucock, 2007: $d = -0.01$ – 0.24 ; Schindler & Hiller, 2010: $d = -0.06$). It is important to notice that the duration of the waiting period compared or was even longer than the typical duration of treatment in routine clinical practice. Moreover, there may have been uncontrolled confounding by antidepressant medication and adjunct treatment even though two of the included studies have found no difference in treatment outcomes between patients with and without adjunct antidepressant treatment (Oei & Yeoh, 1999; Scott, 2005). It therefore seems unlikely that the changes would have occurred without CBT. Caution is given, however, when interpreting subgroup comparison and meta-regression results, as those are observational in nature and may be due to potential confounding by other study-level characteristics.

Fourth, for the ITT data, there was evidence of small-study effects based on the funnel plot and Egger's test. This may indicate publication or other reporting biases but may also be due to different reasons other than reporting bias such as true (clinical or methodological) heterogeneity between effectiveness studies, poorer methodological quality of smaller studies, or simply chance (Egger et al., 1997; Sterne, Egger, & Smith, 2001; Sterne et al., 2000, 2011). As bias may have resulted in an overestimation of ITT effect sizes, these estimates should therefore be regarded as preliminary. However, the significant number of effectiveness studies with null results needed to reduce the mean ITT effect size to trivial was larger than the number of included studies. The large Fail-safe N thus increases our confidence that depressed patients who start CBT in routine practice can be treated effectively.

Last, we encountered some methodological shortcomings. Formal diagnostic procedures to assess depression were described in

only less than half of the studies. The small sample sizes in one out of four studies can also be perceived as a limitation. A further limitation is that about 30% of the investigators failed to describe the extent of dropout. However, even if dropout is reported, ITT analysis is not commonly used. In addition to what has already been mentioned, it would be valuable if more patients were followed up after completion of CBT to test for the stability of treatment effects. Future investigators must try to overcome the methodological limitations of past effectiveness studies.

There is therefore a need for more high-quality effectiveness studies that meet minimal quality criteria such as the ones we have suggested. In particular, authors should indicate the number of patients who dropped out of treatment. Given the substantial rates of attrition, we emphasize the importance of analyzing data according to ITT. Moreover, to ensure a high quality of reporting, we strongly recommend using the reporting standards guide of the Transparent Reporting of Evaluations With Nonrandomized Designs (TREND) group (Des Jarlais, Lyles, & Crepaz, 2004). Authors should also appropriately index their reports as "effectiveness" studies to ease their identification.

To our knowledge, this is the first meta-analysis of nonrandomized effectiveness studies on outpatient CBT for adult depression. Despite methodological limitations inherent in the design of effectiveness studies, we demonstrated that outpatient individual and group CBT for depression is effectively transported to routine clinical practice. The high dropout from CBT needs to be reduced and more research is needed to investigate which characteristics of treatment provision can contribute to the reduction of dropout. Moreover, the small number of available studies and the poor methodological and reporting quality of some studies stress the need for more high-quality effectiveness studies.

References

- References marked with an asterisk indicate studies included in the meta-analysis. References marked with a plus sign indicate benchmark studies.
- Addis, M. E., & Krasnow, A. D. (2000). A national survey of practicing psychologists' attitudes toward psychotherapy treatment manuals. *Journal of Consulting and Clinical Psychology, 68*, 331–339. doi:10.1037/0022-006X.68.2.331
- [†]Allart-van Dam, E., Hosman, C. M. H., Hoogduin, C. A. L., & Schaap, C. P. D. R. (2003). The coping with depression course: Short-term outcomes and mediating effects of a randomized controlled trial in the treatment of subclinical depression. *Behavior Therapy, 34*, 381–396.
- Bados, A., Balaguer, G., & Saldaña, C. (2007). The efficacy of cognitive-behavioral therapy and the problem of drop-out. *Journal of Clinical Psychology, 63*, 585–592. doi:10.1002/jclp.20368
- Barkham, M., Mullin, T., Leach, C., Stiles, W. B., & Lucock, M. (2007). Stability of the CORE-OM and the BDI-I prior to therapy: Evidence from routine practice. *Psychology and Psychotherapy: Theory, Research and Practice, 80*, 269–278. doi:10.1348/147608306X148048
- Becker, B. J. (1988). Synthesizing standardized mean-change measures. *British Journal of Mathematical and Statistical Psychology, 41*, 257–278. doi:10.1111/j.2044-8317.1988.tb00901.x
- [†]Beutler, L. E., Engle, D., Mohr, D., Daldrup, R. J., Bergan, J., Meredith, K., & Merry, W. (1991). Predictors of differential response to cognitive, experiential, and self-directed psychotherapeutic procedures. *Journal of*

- Consulting and Clinical Psychology*, 59, 333–340. doi:10.1037/0022-006X.59.2.333
- *Brown, J. S., Sellwood, K., Beecham, J. K., Slade, M., Andiappan, M., Landau, S., . . . Smith, R. (2011). Outcome, costs and patient engagement for group and individual CBT for depression: A naturalistic clinical study. *Behavioural and Cognitive Psychotherapy*, 39, 355–358. doi:10.1017/S135246581000072X
- +Brown, R. A., & Lewinsohn, P. M. (1984). A psychoeducational approach to the treatment of depression: Comparison of group, individual, and minimal contact procedures. *Journal of Consulting and Clinical Psychology*, 52, 774–783. doi:10.1037/0022-006X.52.5.774
- Buchkremer, G., & Klingberg, S. (2001). Was ist fundierte Psychotherapie? Zur Diskussion um die Leitlinien für die Psychotherapieforschung [What is scientifically founded psychotherapy? The debate on guidelines for psychotherapy research]. *Der Nervenarzt*, 72, 20–30. doi:10.1007/s001150050708
- 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. doi:10.1016/j.cpr.2005.07.003
- *Cahill, J., Barkham, M., Hardy, G., Rees, A., Shapiro, D. A., Stiles, W. B., & Macaskill, N. (2003). Outcomes of patients completing and not completing cognitive therapy for depression. *British Journal of Clinical Psychology*, 42, 133–143. doi:10.1348/014466503321903553
- Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, 66, 7–18. doi:10.1037/0022-006X.66.1.7
- Chambless, D. L., & Ollendick, T. H. (2001). Empirically supported psychological interventions: Controversies and evidence. *Annual Review of Psychology*, 52, 685–716. doi:10.1146/annurev.psych.52.1.685
- Coelho, H. F., Canter, P. H., & Ernst, E. (2007). Mindfulness-based cognitive therapy: Evaluating current evidence and informing future research. *Journal of Consulting and Clinical Psychology*, 75, 1000–1005. doi:10.1037/0022-006X.75.6.1000
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- *Craigie, M. A., & Nathan, P. (2009). A nonrandomized effectiveness comparison of broad-spectrum group CBT to individual CBT for depressed outpatients in a community mental health setting. *Behavior Therapy*, 40, 302–314. doi:10.1016/j.beth.2008.08.002
- Cuijpers, P. (1998a). A psychoeducational approach to the treatment of depression: A meta-analysis of Lewinsohn's "Coping With Depression" course. *Behavior Therapy*, 29, 521–533. doi:10.1016/S0005-7894(98)80047-6
- Cuijpers, P. (1998b). Psychological outreach programmes for the depressed elderly: A meta-analysis of effects and dropout. *International Journal of Geriatric Psychiatry*, 13, 41–48. doi:10.1002/(SICI)1099-1166(199801)13:1<41::AID-GPS729>3.0.CO;2-B
- Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). The effects of psychotherapy for adult depression are overestimated: A meta-analysis of study quality and effect size. *Psychological Medicine*, 40, 211–223. doi:10.1017/S0033291709006114
- Cuijpers, P., van Straten, A., Schuurmans, J., van Oppen, P., Hollon, S. D., & Andersson, G. (2010). Psychotherapy for chronic major depression and dysthymia: A meta-analysis. *Clinical Psychology Review*, 30, 51–62. doi:10.1016/j.cpr.2009.09.003
- Cuijpers, P., van Straten, A., van Oppen, P., Andersson, G., & Hollon, S. D. (2010). *Characteristics of the 281 trials (in SPSS) and a description of the variables available for each study [Data file and code book]*. Retrieved from <http://www.psychotherapyrcts.org>
- Cuijpers, P., van Straten, A., & Warmerdam, L. (2008). Are individual and group treatments equally effective in the treatment of depression in adults? A meta-analysis. *The European Journal of Psychiatry*, 22, 38–51. doi:10.4321/S0213-61632008000100005
- Cuijpers, P., van Straten, A., Warmerdam, L., & Smits, N. (2008). Characteristics of effective psychological treatments of depression: A meta-regression analysis. *Psychotherapy Research*, 18, 225–236. doi:10.1080/10503300701442027
- Des Jarlais, D. C., Lyles, C., & Crepaz, N. (2004). Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: The TREND statement. *American Journal of Public Health*, 94, 361–366. doi:10.2105/AJPH.94.3.361
- +Dimidjian, S., Hollon, S. D., Dobson, K. S., Schmalzing, K. B., Kohlenberg, R. J., Addis, M. E., . . . Jacobson, N. S. (2006). Randomized trial of behavioral activation, cognitive therapy, and antidepressant medication in the acute treatment of adults with major depression. *Journal of Consulting and Clinical Psychology*, 74, 658–670. doi:10.1037/0022-006X.74.4.658
- Dunlap, W. P., Cortina, J. M., Vaslow, J. B., & Burke, M. J. (1996). Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, 1, 170–177. doi:10.1037/1082-989X.1.2.170
- Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629–634. doi:10.1136/bmj.315.7109.629
- *Eisendrath, S. J., Delucchi, K., Bitner, R., Fenimore, P., Smit, M., & McLane, M. (2008). Mindfulness-based cognitive therapy for treatment-resistant depression: A pilot study. *Psychotherapy and Psychosomatics*, 77, 319–320. doi:10.1159/000142525
- +Elkin, I., Shea, M. T., Watkins, J. T., Imber, S. D., Sotsky, S. M., Collins, J. F., . . . Parloff, M. B. (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. doi:10.1001/archpsyc.1989.01810110013002
- *Enns, M. W., Cox, B. J., & Pidlubny, S. R. (2002). Group cognitive behaviour therapy for residual depression: Effectiveness and predictors of response. *Cognitive Behaviour Therapy*, 31, 31–40. doi:10.1080/16506070252823643
- *Finucane, A., & Mercer, S. W. (2006). An exploratory mixed methods study of the acceptability and effectiveness of mindfulness-based cognitive therapy for patients with active depression and anxiety in primary care. *BMC Psychiatry*, 6, 14. doi:10.1186/1471-244X-6-14
- *Forand, N. R., Evans, S., Haglin, D., & Fishman, B. (2011). Cognitive behavioral therapy in practice: Treatment delivered by trainees at an outpatient clinic is clinically effective. *Behavior Therapy*, 42, 612–623. doi:10.1016/j.beth.2011.02.001
- Franklin, M. E., Abramowitz, J. S., Kozak, M. J., Levitt, J. T., & Foa, E. B. (2000). Effectiveness of exposure and ritual prevention for obsessive-compulsive disorder: Randomized compared with nonrandomized samples. *Journal of Consulting and Clinical Psychology*, 68, 594–602. doi:10.1037/0022-006X.68.4.594
- Fydrich, T., & Schneider, W. (2007). Evidenzbasierte Psychotherapie [Evidence-based psychotherapy]. *Psychotherapeut*, 52, 55–68. doi:10.1007/s00278-006-0522-x
- *Gelhart, R. P., & King, H. L. (2001). The influence of comorbid risk factors on the effectiveness of cognitive-behavioral treatment of depression. *Cognitive and Behavioral Practice*, 8, 18–28. doi:10.1016/S1077-7229(01)80039-0
- *Gibbons, C. J., Fournier, J. C., Stirman, S. W., DeRubeis, R. J., Crits-Christoph, P., & Beck, A. T. (2010). The clinical effectiveness of cognitive therapy for depression in an outpatient clinic. *Journal of Affective Disorders*, 125, 169–176. doi:10.1016/j.jad.2009.12.030
- Gloaguen, V., Cottraux, J., Cucherat, M., & Blackburn, I. M. (1998). A meta-analysis of the effects of cognitive therapy in depressed patients. *Journal of Affective Disorders*, 49, 59–72. doi:10.1016/S0165-0327(97)00199-7
- *Haaga, D. A., DeRubeis, R. J., Stewart, B. L., & Beck, A. T. (1991). Relationship of intelligence with cognitive therapy outcome. *Behaviour*

- Research and Therapy*, 29, 277–281. doi:10.1016/0005-7967(91)90118-M
- Hahlweg, K., Fiegenbaum, W., Frank, M., Schroeder, B., & von Witzleben, I. (2001). Short- and long-term effectiveness of an empirically supported treatment for agoraphobia. *Journal of Consulting and Clinical Psychology*, 69, 375–382. doi:10.1037/0022-006X.69.3.375
- *Halford, W. K., Bernoth-Doolan, S., & Eadie, K. (2002). Schemata as moderators of clinical effectiveness of a comprehensive cognitive behavioral program for patients with depression or anxiety disorders. *Behavior Modification*, 26, 571–593. doi:10.1177/014544502236651
- Hardy, R. J., & Thompson, S. G. (1998). Detecting and describing heterogeneity in meta-analysis. *Statistics in Medicine*, 17, 841–856. doi:10.1002/(SICI)1097-0258(19980430)17:8<841::AID-SIM781>3.0.CO;2-D
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods*, 9, 426–445. doi:10.1037/1082-989X.9.4.426
- Higgins, J. P., & Thompson, S. G. (2004). Controlling the risk of spurious findings from meta-regression. *Statistics in Medicine*, 23, 1663–1682. doi:10.1002/sim.1752
- Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327, 557–560. doi:10.1136/bmj.327.7414.557
- *Hoberman, H. M., Lewinsohn, P. M., & Tilson, M. (1988). Group treatment of depression: Individual predictors of outcome. *Journal of Consulting and Clinical Psychology*, 56, 393–398. doi:10.1037/0022-006X.56.3.393
- *Houghton, S., Curran, J., & Saxon, D. (2008). An uncontrolled evaluation of group behavioural activation for depression. *Behavioural and Cognitive Psychotherapy*, 36, 235–239. doi:10.1017/S1352465808004207
- +Jacobson, N. S., Dobson, K. S., Truax, P. A., Addis, M. E., Koerner, K., Gollan, J. K., . . . Prince, S. E. (1996). A component analysis of cognitive-behavioral treatment for depression. *Journal of Consulting and Clinical Psychology*, 64, 295–304. doi:10.1037/0022-006X.64.2.295
- +Jarrett, R. B., Schaffer, M., McIntire, D., Witt-Browder, A., Kraft, D., & Risser, R. C. (1999). Treatment of atypical depression with cognitive therapy or phenelzine: A double-blind, placebo-controlled trial. *Archives of General Psychiatry*, 56, 431–437. doi:10.1001/archpsyc.56.5.431
- *Kenny, M. A., & Williams, J. M. G. (2007). Treatment-resistant depressed patients show a good response to mindfulness-based cognitive therapy. *Behaviour Research and Therapy*, 45, 617–625. doi:10.1016/j.brat.2006.04.008
- +King, M., Sibbald, B., Ward, E., Bower, P., Lloyd, M., Gabbay, M., & Byford, S. (2000). Randomised controlled trial of nondirective counselling, cognitive-behaviour therapy and usual general practitioner care in the management of depression as well as mixed anxiety and depression in primary care. *Health Technology Assessment*, 4, 1–83.
- *Kingston, T., Dooley, B., Bates, A., Lawlor, E., & Malone, K. (2007). Mindfulness-based cognitive therapy for residual depressive symptoms. *Psychology and Psychotherapy: Theory, Research and Practice*, 80, 193–203. doi:10.1348/147608306X116016
- *Kush, F. R., & Fleming, L. M. (2000). An innovative approach to short-term group cognitive therapy in the combined treatment of anxiety and depression. *Group Dynamics: Theory, Research, and Practice*, 4, 176–183. doi:10.1037/1089-2699.4.2.176
- *Kwon, S. M., & Oei, T. P. (2003). Cognitive change processes in a group cognitive behavior therapy of depression. *Journal of Behavior Therapy and Experimental Psychiatry*, 34, 73–85. doi:10.1016/S0005-7916(03)00021-1
- *Lange, I. (2009). Erfolg von kognitiver Verhaltenstherapie unter Routinebedingungen, im Prä-Post-Vergleich [Success of cognitive behavioral therapy in routine clinical practice using pre–post comparisons] (Unpublished master's thesis). Westfälische Wilhelms-Universität, Münster, Germany.
- Lincoln, T. M., Rief, W., Hahlweg, K., Frank, M., von Witzleben, I., Schroeder, B., & Fiegenbaum, W. (2003). Effectiveness of an empirically supported treatment for social phobia in the field. *Behaviour Research and Therapy*, 41, 1251–1269. doi:10.1016/S0005-7967(03)00038-X
- Lipsey, M. W. (1990). *Design sensitivity: Statistical power for experimental research*. Newbury Park, CA: Sage.
- Lipsey, M. W., & Wilson, D. B. (2001). Applied social research methods series: Vol. 49. *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Maier-Riehle, B., & Zwingmann, C. (2000). Effektstärkevarianten beim Eingruppen-Prä-Post-Design: Eine kritische Betrachtung [Effect strength variation in the single group pre–post study design: A critical review]. *Rehabilitation*, 39, 189–199. doi:10.1055/s-2000-12042
- *Marshall, T. K., & Mazie, A. S. (1987). A cognitive approach to treating depression. *Social Casework*, 68, 540–545.
- *Matsunaga, M., Okamoto, Y., Suzuki, S., Kinoshita, A., Yoshimura, S., Yoshino, A., . . . Yamawaki, S. (2010). Psychosocial functioning in patients with treatment-resistant depression after group cognitive behavioral therapy. *BMC Psychiatry*, 10, 22. doi:10.1186/1471-244X-10-22
- McCullough, J. P. (1991). Psychotherapy for dysthymia. A naturalistic study of ten patients. *Journal of Nervous and Mental Disease*, 179, 734–740. doi:10.1097/00005053-199112000-00004
- *Merrill, K. A., Tolbert, V. E., & Wade, W. A. (2003). Effectiveness of cognitive therapy for depression in a community mental health center: A benchmarking study. *Journal of Consulting and Clinical Psychology*, 71, 404–409. doi:10.1037/0022-006X.71.2.404
- Minami, T., Wampold, B. E., Serlin, R. C., Kircher, J. C., & Brown, G. S. (2007). Benchmarks for psychotherapy efficacy in adult major depression. *Journal of Consulting and Clinical Psychology*, 75, 232–243. doi:10.1037/0022-006X.75.2.232
- Nalini, N. R., Kumaraiah, V., & Subbakrishna, D. K. (1996). Cognitive behavioural therapy in the treatment of neurotic depression. *Nimhans Journal*, 14, 31–35.
- +Neimeyer, R. A., Kazantzis, N., Kassler, D. M., Baker, K. D., & Fletcher, R. (2008). Group cognitive behavioural therapy for depression outcomes predicted by willingness to engage in homework, compliance with homework, and cognitive restructuring skill acquisition. *Cognitive Behaviour Therapy*, 37, 199–215. doi:10.1080/16506070801981240
- *Nisbet Wallis, D. A. (2002). Depression, anxiety and self-esteem: A clinical field study. *Behaviour Change*, 19, 112–120. doi:10.1375/beck.19.2.112
- Oei, T. P., & Dingle, G. (2008). The effectiveness of group cognitive behavioural therapy for unipolar depressive disorders. *Journal of Affective Disorders*, 107, 5–21. doi:10.1016/j.jad.2007.07.018
- *Oei, T. P., & Yeoh, A. E. (1999). Pre-existing antidepressant medication and the outcome of group cognitive-behavioural therapy. *Australian and New Zealand Journal of Psychiatry*, 33, 70–76. doi:10.1046/j.1440-1614.1999.00520.x
- *Organista, K. C., Munoz, R. F., & Gonzalez, G. (1994). Cognitive-behavioral therapy for depression in low-income and minority medical outpatients: Description of a program and exploratory analyses. *Cognitive Therapy and Research*, 18, 241–259. doi:10.1007/BF02357778
- Orwin, R. G. (1983). A Fail-safe N for effect size in meta-analysis. *Journal of Educational Statistics*, 8, 157–159. doi:10.2307/1164923
- *Persons, J. B., Bostrom, A., & Bertagnolli, A. (1999). Results of randomized controlled trials of cognitive therapy for depression generalize to private practice. *Cognitive Therapy and Research*, 23, 535–548. doi:10.1023/A:1018724505659
- *Persons, J. B., Burns, D. D., & Perloff, J. M. (1988). Predictors of dropout and outcome in cognitive therapy for depression in a private-practice

- setting. *Cognitive Therapy and Research*, 12, 557–575. doi:10.1007/BF01205010
- *Peterson, A. L., & Halstead, T. S. (1998). Group cognitive behavior therapy for depression in a community setting: A clinical replication series. *Behavior Therapy*, 29, 3–18. doi:10.1016/S0005-7894(98)80015-4
- *Piacentini, D., Mirabella, F., Leveni, D., Primerano, G., Cattaneo, M., Biffi, G., . . . Gigantesco, A. (2011). Efficacia di un intervento cognitivo-comportamentale manualizzato nella depressione post-partum [Effectiveness of a manualized cognitive behavioral intervention for postnatal depression]. *Rivista di Psichiatria*, 46, 187–194. doi:10.1708/889.9809
- Posternak, M. A., & Miller, I. (2001). Untreated short-term course of major depression: A meta-analysis of outcomes from studies using wait-list control groups. *Journal of Affective Disorders*, 66, 139–146. doi:10.1016/S0165-0327(00)00304-9
- Reeves, B. C., Deeks, J. J., Higgins, J. P. T., & Wells, G. A. (2011). Chapter 13: Including non-randomized studies. In J. P. T. Higgins & S. Green (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions (Version 5.1.0)*. Retrieved from <http://www.cochrane-handbook.org>
- *Renner, W., & Platz, T. (1999). Kognitive und symptombezogene Effekte standardisierter Verhaltenstherapie: Evaluation eines ambulanten Gruppenprogramms [Cognitive and symptomatic effects of standardized behavior therapy: Evaluation of an outpatient group program]. *Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie*, 47, 271–291.
- *Rosenthal Gelman, C., Lopez, M., & Foster, R. P. (2006). Evaluating the impact of a cognitive-behavioral intervention with depressed Latinas: A preliminary report. *Social Work in Mental Health*, 4, 1–16. doi:10.1300/J200v04n02_01
- Rutherford, B. R., Mori, S., Sneed, J. R., Pimontel, M. A., & Roose, S. P. (2012). Contribution of spontaneous improvement to placebo response in depression: A meta-analytic review. *Journal of Psychiatric Research*, 46, 697–702. doi:10.1016/j.jpsychires.2012.02.008
- Ryan, R., Hill, S., Broclain, D., Horey, D., Oliver, S., & Prictor, M., & the Cochrane Consumers and Communication Review Group. (2007). *Study quality guide*. Retrieved from <http://www.latrobe.edu.au/chcp/assets/downloads/StudyQualityGuide050307.pdf>
- *Schindler, A., & Hiller, W. (2010). Therapieeffekte und Responderaten bei unipolar depressiven Patienten einer verhaltenstherapeutischen Hochschulambulanz [Therapy effects and response rates of cognitive-behavioral treatment for unipolar depressive patients in an outpatient clinic]. *Zeitschrift für Klinische Psychologie und Psychotherapie: Forschung und Praxis*, 39, 107–115. doi:10.1026/1616-3443/a000019
- Schindler, A. C., Hiller, W., & Witthöft, M. (2011). Benchmarking of cognitive-behavioral therapy for depression in efficacy and effectiveness studies: How do exclusion criteria affect treatment outcome? *Psychotherapy Research*, 21, 644–657. doi:10.1080/10503307.2011.602750
- *Scott, R. H. (2005). Effectiveness of group versus individual psychotherapy for depression in a community counselling setting using the “mind over mood” cognitive therapy manual. *Dissertation Abstracts International: Section B. Sciences and Engineering*, 65(12-B), 6673.
- Seligman, M. E. (1995). The effectiveness of psychotherapy: The Consumer Reports study. *American Psychologist*, 50, 965–974. doi:10.1037/0003-066X.50.12.965
- 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. doi:10.1037/0033-2909.126.4.512
- Sterne, J. A. C., & Egger, M. (2006). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 99–110). Chichester, England: Wiley. doi:10.1002/0470870168.ch6
- Sterne, J. A. C., Egger, M., & Smith, G. D. (2001). Systematic reviews in health care: Investigating and dealing with publication and other biases in meta-analysis. *British Medical Journal*, 323, 101–105. doi:10.1136/bmj.323.7304.101
- Sterne, J. A. C., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology*, 53, 1119–1129. doi:10.1016/S0895-4356(00)00242-0
- Sterne, J. A. C., Sutton, A. J., Ioannidis, J. P., Terrin, N., Jones, D. R., Lau, J., . . . Higgins, J. P. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *British Medical Journal*, 343, d4002. doi:10.1136/bmj.d4002
- Stewart, R. E., & Chambless, D. L. (2009). Cognitive-behavioral therapy for adult anxiety disorders in clinical practice: A meta-analysis of effectiveness studies. *Journal of Consulting and Clinical Psychology*, 77, 595–606. doi:10.1037/a0016032
- Stirman, S. W., DeRubeis, R. J., Crits-Christoph, P., & Brody, P. E. (2003). Are samples in randomized controlled trials of psychotherapy representative of community outpatients? A new methodology and initial findings. *Journal of Consulting and Clinical Psychology*, 71, 963–972. doi:10.1037/0022-006X.71.6.963
- Stirman, S. W., DeRubeis, R. J., Crits-Christoph, P., & Rothman, A. (2005). Can the randomized controlled trial literature generalize to nonrandomized patients? *Journal of Consulting and Clinical Psychology*, 73, 127–135. doi:10.1037/0022-006X.73.1.127
- Stuart, G. L., Treat, T. A., & Wade, W. A. (2000). Effectiveness of an empirically based treatment for panic disorder delivered in a service clinic setting: 1-year follow-up. *Journal of Consulting and Clinical Psychology*, 68, 506–512. doi:10.1037/0022-006X.68.3.506
- *Swan, J., Sorrell, E., MacVicar, B., Durham, R., & Matthews, K. (2004). “Coping with depression”: An open study of the efficacy of a group psychoeducational intervention in chronic, treatment-refractory depression. *Journal of Affective Disorders*, 82, 125–129. doi:10.1016/j.jad.2003.09.002
- *Teri, L., & Lewinsohn, P. M. (1986). Individual and group treatment of unipolar depression: Comparison of treatment outcome and identification of predictors of successful treatment outcome. *Behavior Therapy*, 17, 215–228. doi:10.1016/S0005-7894(86)80052-1
- Thompson, S. G., & Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, 21, 1559–1573. doi:10.1002/sim.1187
- Tuschen-Caffier, B., Pook, M., & Frank, M. (2001). Evaluation of manual-based cognitive-behavioral therapy for bulimia nervosa in a service setting. *Behaviour Research and Therapy*, 39, 299–308. doi:10.1016/S0005-7967(00)00004-8
- van Ingen, D. J., Freiheit, S. R., & Vye, C. S. (2009). From the lab to the clinic: Effectiveness of cognitive-behavioral treatments for anxiety disorders. *Professional Psychology: Research and Practice*, 40, 69–74. doi:10.1037/a0013318
- Wade, W. A., Treat, T. A., & Stuart, G. L. (1998). Transporting an empirically supported treatment for panic disorder to a service clinic setting: A benchmarking strategy. *Journal of Consulting and Clinical Psychology*, 66, 231–239. doi:10.1037/0022-006X.66.2.231
- *Westbrook, D., & Kirk, J. (2005). The clinical effectiveness of cognitive behaviour therapy: Outcome for a large sample of adults treated in routine practice. *Behaviour Research and Therapy*, 43, 1243–1261. doi:10.1016/j.brat.2004.09.006
- *Westbrook, D., & Kirk, J. (2007). The clinical effectiveness of cognitive behaviour therapy: Outcome for a large sample of adults treated in routine practice [Erratum]. *Behaviour Research and Therapy*, 45, 1703–1704. doi:10.1016/j.brat.2006.11.008

Westen, D., & Morrison, K. (2001). A multidimensional meta-analysis of treatments for depression, panic, and generalized anxiety disorder: An empirical examination of the status of empirically supported therapies. *Journal of Consulting and Clinical Psychology, 69*, 875–899. doi: 10.1037/0022-006X.69.6.875

Wissenschaftlicher Beirat Psychotherapie. (2010). *Methodenpapier des Wissenschaftlichen Beirats Psychotherapie nach § 11 PsychThG: Verfahrensregeln zur Beurteilung der wissenschaftlichen Anerkennung von Methoden und Verfahren der Psychotherapie* [Method paper of the German Academic Advisory Council for Psychotherapy according to §

11 PsychThG: Procedural rules for the assessment of the scientific validation of psychotherapeutic methods and techniques] (Version 2.8). Retrieved from <http://www.wbpsychotherapie.de/downloads/Methodenpapier28.pdf>

Received July 21, 2011

Revision received June 7, 2012

Accepted October 9, 2012 ■