

# Distinctions Without a Difference: Direct Comparisons of Psychotherapies for Alcohol Use Disorders

Zac E. Imel and Bruce E. Wampold  
University of Wisconsin—Madison

Scott D. Miller  
Institute for the Study of Therapeutic Change

Reg R. Fleming  
Youth and Family Addiction Services, Vancouver Island Health Authority

To estimate the relative efficacy of alcohol use disorder treatments, the authors meta-analyzed studies that directly compared 2 bona fide psychological treatments. The authors accommodated problems with the inclusion of multiple treatment comparisons by randomly assigning a positive/negative sign to the effect size derived from each comparison and then estimating the extent to which effect sizes were heterogeneous. The authors' primary hypothesis was that the variability in effect sizes of *bona fide* psychological treatments for alcohol use disorders that were directly compared would be zero. For both alcohol measures and measures of abstinence, analyses indicate that effects were homogeneously distributed about zero ( $F^2 = 10.61, 0.00$ , respectively), indicating that different treatment comparisons yielded a common effect size that was not significantly different from zero. Analyses also indicate that allegiance accounted for a significant portion of variability in differences between treatments. Implications for the treatment of alcohol use disorders as well as research on the mechanisms responsible for the benefit of treatment are discussed.

*Keywords:* psychotherapy, treatment outcome, alcohol use disorders, meta-analysis, allegiance

Narrative reviews and meta-analyses have indicated that psychological treatments for alcohol use disorders are effective in decreasing problem drinking. At the same time, the research is less clear regarding which, if any, of the many treatment approaches are most efficacious (Berglund et al., 2003; W. R. Miller, Andrews, Wilbourne, & Bennett, 1998; Morgenstern & McKay, 2007; Wilbourne & Miller, 2002). The standard for determining the relative efficacy of psychotherapies is the randomized clinical trial (RCT). In an RCT, patients are randomly assigned to at least two of several possible conditions, including: (a) wait-list/no-treatment, (b) treatment as usual and/or supportive/nonspecific controls, or (c) a bona fide psychological treatment, each allowing different conclusions about a particular treatment. Wampold et al. (1997) developed the construct of a *bona fide* psychotherapy to refer to a psychotherapy that was intended to be fully therapeutic as opposed to an intervention designed to control for some aspect of psychotherapy, such as attention from an empathic healer (see Wampold et al., 1997, and the discussion in the Method section for a more detailed discussion). The most well-known study comparing bona fide treatments for alcohol use disorders is likely Project

MATCH (Project MATCH Research Group, 1997). Briefly, this large RCT found little difference in drinking outcomes between motivational enhancement, 12-step facilitation, and relapse prevention. However, as the superiority (or lack thereof) of one treatment in any one study may be due to Type I or Type II error, and as narrative reviews of RCTs are particularly susceptible to researcher bias, meta-analysis has become particularly important in integrating the results of RCTs (Cooper & Hedges, 1994). Below, we have offered a brief review of meta-analyses—organized by type of comparison—that addressed the issue of relative efficacy in alcohol use disorders.

Several researchers have completed meta-analyses of RCTs comparing a bona fide treatment to wait-list or no-treatment conditions. For example, Irvin, Bowers, Dunn, and Wang (1999) found that relapse prevention programs were superior to wait-list or no-treatment, and Walters (2000) found that behavioral self-control training (BSCT) was superior to no-treatment. Surprisingly, however, a meta-analysis of Alcoholics Anonymous (AA) indicated that it was less effective than no-treatment (Kownacki & Shadish, 1999). A similar number of meta-analyses have aggregated the results of RCTs comparing a particular treatment to a discussion or nonspecific control. Irvin et al. (1999) found that relapse prevention was superior to discussion control conditions. In a more recent meta-analysis, Berglund et al. (2003) found that *specific treatment* was superior to nonspecific treatment, where specific treatment was defined as a treatment that (a) has a theoretical base, (b) is guided by a manual, and (c) employs therapists who receive training and supervision. Nonspecific treatment usually included some form of treatment as usual or supportive counseling.

---

Zac E. Imel and Bruce E. Wampold, Department of Counseling Psychology, University of Wisconsin—Madison; Scott D. Miller, Institute for the Study of Therapeutic Change, Chicago, Illinois; Reg R. Fleming, Youth and Family Addiction Services, Vancouver Island Health Authority, Victoria, British Columbia, Canada.

Correspondence concerning this article should be addressed to Zac E. Imel, Department of Counseling Psychology, 321 Education Building, University of Wisconsin—Madison, 1000 Bascom Mall, Madison, WI 53706. E-mail: zeimel@wisc.edu

In a recent quantitative review, Wilbourne and Miller (2002) summarized the body of evidence in support of each psychological treatment that has been empirically evaluated. The researchers summarized the evidence by assigning points to treatments on the basis of how much they have been studied, methodological quality, and outcome. Specifically, a well-controlled study with positive results for a particular treatment resulted in a large number of points being attributed to that treatment—the treatment with the most points wins. Using this metric, the psychological treatments with the most empirical support appeared to be motivational enhancement and social skills training. Those with the least empirical support included generic psychotherapy, mandated AA, confrontational counseling, and relaxation training.

The commonalities in these meta-analytic methodologies described above involve (a) comparisons of treatments to wait-list or no-treatment, and (b) comparisons to discussion controls. Inferences about the relative efficacy of treatments from these reviews are made from findings derived across studies. In such cases, the size of the effects for each treatment may be due in part to study level confounds, such as measure reactivity, sample severity, clinical site, and treatment team (Shadish & Sweeney, 1991; Wampold et al., 1997). For example, although Wilbourne and Miller's (2002) review was quite sophisticated, providing an important summation of the current state of the field, the apparent superiority or inferiority of certain treatments may be due to factors other than the treatment itself (e.g., how often a treatment is studied). A meta-analytic strategy that controls potential between-study confounds involves restricting meta-analyses to those trials that directly compare treatments within the same study (see Shadish & Sweeney, 1991).

There are several meta-analyses that have aggregated the results of studies directly comparing at least two treatments for alcohol use disorders. Walters (2000) also meta-analyzed direct comparisons of BSCT, finding that it was superior to "other" treatments. The "other" category included treatment as usual, education, coping skills, general counseling, self-monitoring, and aversion treatments. However, BSCT was not statistically superior to abstinence-based controls (Walters, 2000). In Kownacki and Shadish's (1999) meta-analysis noted above, the authors also analyzed five studies that directly compared AA and some alternative treatment in the same study, finding that AA was less effective than these alternative treatments. Irvin et al.'s (1999) meta-analysis indicated that relapse prevention was less effective than other active treatments. However, this meta-analysis only included one study of alcohol use disorder treatment and thus can not truly be interpreted as a meta-analysis. Berglund et al. (2003) meta-analyzed all published randomized controlled trials of the psychological treatment of alcohol use disorders. The authors reported that there appeared to be no differences between specific treatments. However, the authors did not analyze the studies using standard meta-analytic techniques, and they drew conclusions from a tally of significance tests. Out of 30 studies comparing specific treatments, 19 provided no evidence of differences between treatments.

There are several limitations of these meta-analyses of direct comparisons in terms of determining the relative efficacy of alcohol use disorders treatments. First, the meta-analyses of direct comparisons noted above contained relatively few studies. For example, Irvin et al.'s (1999) meta-analysis of relapse prevention

programs only contained one study that directly compared relapse prevention to another active intervention. Second, the comparison conditions analyzed in these meta-analyses—such as "alternative" or "other"—introduce a classification problem. Specifically, these categories (e.g., treatment as usual, education, supportive controls, etc.) include a variety of conditions that were not bona fide/specific treatments. Although the absolute efficacy and/or the specific ingredients of a treatment are often estimated *vis-à-vis* supportive or general counseling controls that are not intended to be therapeutic, meta-analyses that include these comparisons may spuriously overestimate differences between treatments. Kownacki and Shadish's (1999) meta-analysis indicated that AA was less effective than alternative interventions but included studies that compared conventional AA meetings to "traditional treatment," a psychological oriented clinic, and inpatient treatment, which were not fully described. These comparisons are more adequately described as comparisons of conventional AA meetings to "usual care" and not an alternative bona fide psychotherapy (see Ditman, Crawford, Forgy, Moskowitz, & Macandrew, 1967; Keso & Salaspuro, 1990; Walsh et al., 1991). Specifically, these conditions did not offer a description of specific theoretical orientation to guide the intervention, and it was unclear what type of treatment a patient may have received. The comparison of an intervention to usual care certainly provides more information about relative efficacy than comparison to a wait-list; however, a limitation of this design is often a lack of clarity regarding what is contained in a particular usual care condition. Usual care may involve referral to community treatment where the researchers have little influence or knowledge about the actual treatment provided to the patient. Although some usual care conditions may be intended to be therapeutic, the content of the intervention may vary from study to study to such an extent that conclusions about relative efficacy are ambiguous.

Third, previous meta-analyses for the treatment of alcohol use disorders have categorized treatments into classes of comparisons (e.g., relapse prevention vs. 12-step facilitation, 12-step facilitation vs. other). Categorization requires pairwise comparisons of classes, creating a large number of statistical tests. However, in the alcohol use disorder treatment literature, there are relatively few comparisons between any two particular classes. For example, Berglund et al. (2003) did not conduct a true meta-analysis of direct comparisons because there was not a sufficient number of comparisons within any particular class. Furthermore, meta-analyzing a collection of between class pairwise comparisons ignores comparisons within classes of treatments (e.g., comparisons of types of cognitive treatments)—preventing an omnibus test of the hypotheses that outcomes differ between all directly compared psychotherapies (Shadish, Montgomery, Wilson, & Wilson, 1993; Wampold et al., 1997).

An additional confounding variable in psychotherapy RCTs is researcher allegiance. The allegiance construct emerged as researchers became perplexed by the conflicting results of different studies that compared the same two treatments. Although the construct of researcher allegiance is not well understood, it can be defined as some combination of researcher bias (Rosenthal, 1966) and variable therapist belief/performance across treatments (Thase, 1999). Researchers have observed that differences between treatments could be explained by the allegiance of the researchers (Luborsky et al., 1999; Luborsky, Singer, & Luborsky, 1975).

Luborsky et al. (1975) first described the allegiance effect in a review of behavior therapy studies, finding studies characterized by strong allegiance to behavior therapy were the ones that found the behavior therapy to be superior to alternative treatments. The finding that researcher allegiance is associated with treatment outcome has been replicated in a number of subsequent publications (Gaffan, Tsalousis, & Kemp-Wheeler, 1995; Luborsky et al., 1999; S. Miller, Wampold, & Varhely, 2008; Robinson, Berman, & Neimeyer, 1990). Luborsky et al. (1999) demonstrated that the combination of three allegiance rating methods accounted for 69% of the variance in treatment outcomes, noting that there were no articles published by the founder of a treatment that contradicted the authors' allegiance.

Given the large association of researcher allegiance to relative efficacy in studies that directly compare psychotherapies, it is important to examine the possibility that any observed differences between treatments are related to researcher allegiance. For example, when differences between treatments were found in Berglund et al.'s (2003) review, the experimental treatment was often superior to the alternative. To our knowledge, there have been no direct investigations of researcher allegiance in the treatment of alcohol use disorders.

### Rationale

Wampold et al. (1997) offered a method for addressing several problems with past meta-analyses that estimated relative efficacy by meta-analyzing direct comparisons of at least two bona fide psychotherapies. Wampold et al.'s meta-analysis revealed no evidence of differences between bona fide treatments directly compared within studies. However, one limitation of this finding was that this omnibus conclusion might not apply to the treatment of particular disorders, such as alcohol use disorders (Crits-Christoph, 1997).

The current meta-analysis provides the following contributions: (a) an analysis of all available studies that compared two or more bona fide treatments for an alcohol use disorder within the same study, thus controlling for the potential study level confounds noted earlier (e.g., sample idiosyncrasies, how often a treatment is studied, measure reactivity) that are encountered when comparing effect sizes across studies, (b) an examination of relative efficacy in a specific clinically dysfunctional population, (c) avoidance of classifying treatments into categories such that an omnibus test of relative efficacy can be conducted, and (d) inclusion of only bona fide psychotherapies so that any treatment differences cannot be attributed to the use of a treatment that was not intended to be therapeutic. Our hypotheses were that effect sizes for direct comparisons of bona fide psychosocial treatments for alcohol use disorders would not vary significantly from zero and that any differences among treatments would be accounted for by researcher allegiance.

### Method

#### *Selection of Studies*

To locate all the studies that compared bona fide psychological treatments for alcohol use disorders, we conducted a comprehensive literature search of the online databases PsycINFO and Medline utilizing the search terms listed in Table 1. Subsequently, we

Table 1  
*Search Terms Used to Conduct a Comprehensive Literature Search of the Online Databases*

No.	Search terms
1	Alcohol and randomized controlled trial
2	Alcoholism and randomized controlled trial
3	Alcohol and controlled trial
4	Alcoholism and controlled trial
5	Heavy drinking and intervention
6	Counseling and alcohol and controlled trial
7	Problem drinking and intervention
8	Controlled drinking and intervention
9	Intervention and problem drinking and problem drinker and controlled trial
10	Intervention and alcohol consumption and controlled trial
11	Early intervention and alcohol and controlled trial
12	Alcohol and intervention and controlled trial
13	Alcoholism and intervention and controlled trial

examined abstracts for relevance to the current study. Additionally, the reference lists of both qualitative and quantitative reviews of psychological treatments for alcohol use disorders were used to obtain further studies.

Two raters (doctoral students in counseling psychology) extracted the following information from original studies included in the meta-analysis: (a) means and standard deviations for alcohol-related outcome measures, (b) names of the treatments, and (c) *n*'s for each treatment condition.

To be included in this meta-analysis, a study must have (a) contained sufficient statistics to calculate effect sizes for measures targeted by the study authors (i.e., measures related to alcohol use) and (b) compared at least two bona fide psychological treatments for an alcohol use disorder. Measures of other psychological symptoms, such as depression and psychological well-being, were not analyzed in the meta-analysis. Studies in which patients were not randomly assigned, or those that addressed the prevention of alcohol use disorders, were excluded. Although patients were required to meet diagnostic criteria for some alcohol use disorder, studies were not excluded if patients carried multiple diagnoses.

To determine whether the psychotherapies included in a comparison were intended to be therapeutic (e.g., were bona fide), Wampold et al. (1997) developed criteria to assess the qualities of a therapy condition involved in a treatment trial. In the current study, we adapted the criteria developed by Wampold et al. for use in the context of alcohol use disorders. First, treatments must have involved multiple sessions in which the therapist developed a relationship with the client, and the treatment was tailored to the patient. Accordingly, tape recorded interventions, one-session feedback interventions, or sessions that consisted solely of the therapist reading from a script were excluded (e.g., relaxation training scripts that were read to clients in a standard format without variation were not included). Second, the treatment must have met at least two of the following four criteria: (a) referenced an established approach (e.g., cognitive-behavioral training), (b) included a description of the therapy that mentioned psychological processes or specific ingredients of some kind, (c) used a manual to direct the therapy, and/or (d) contained a description of components necessary for change to occur (i.e., active ingredients). Wampold et al.'s original criteria required that the therapist providing the treatment

held at least a master's degree. As a number of commonly used treatments in alcohol use disorders do not require therapists with graduate degrees, we did not require treatments be conducted by therapists with at least a master's degree.

On the basis of Wampold et al.'s (1997) criteria, we excluded several types of treatment designs. First, any treatments that were explicitly designed to control for nonspecific aspects of other therapies (i.e., sham or control therapies) were excluded. Specifically, these treatments were not fully intended to be therapeutic but were designed to control for some aspect of a treatment that is purportedly critical to treatment success. Similarly, any treatment that was altered or restricted to reduce overlap with other therapies was excluded. Second, component (e.g., McCrady, Epstein, & Hirsch, 1999), dismantling, or parametric (e.g., Morgenstern, Blanchard, Morgan, Labouvie, & Hayaki, 2001) studies that varied the presence or amount of a particular technique were excluded, as the purpose of these studies was to evaluate the effect of specific ingredients, not to estimate the relative efficacy of two psychological treatments with different theoretical bases. For example, studies similar to Fals-Stewart, Birchler, and Kelley (2006), in which behavioral couples therapy was added to individual based treatment, were excluded. Third, as the purpose of the current study was to estimate the relative efficacy of two bona fide psychological treatments and not medical interventions, we excluded treatments in which patients were randomly assigned to medication conditions or pill placebos (e.g., Anton et al., 2005; Combine Study Research Group, 2003). Fourth, we also excluded studies that compared similar treatments in different settings. The purpose of these studies was to examine the impact of particular treatment setting or modality (inpatient vs. outpatient treatment; Rychtarik et al., 2000), not the relative efficacy of different bona fide psychological treatments.

To select studies that directly compared bona fide treatments for alcohol use, two counseling psychology doctoral students (one of these graduate students later participated in the extraction of data for calculating effect sizes) blindly reviewed each study. Decisions were made on the basis of the Introduction and Method sections, and reviewers were blind to the results of the study. If both reviewers determined that the study included two bona fide treatments for alcohol use disorders, the study was included in the analysis. However, in case of disagreement, the study was discussed by the two raters in an attempt to resolve the disagreement. If agreement could not be reached, the study was evaluated by Bruce E. Wampold. If this third rating resulted in the decision that the study included at least two bona fide treatments, the study was retained (i.e., if not, the study was rejected). Therefore, to be included in our analysis, the study must have been deemed to include two bona fide treatments by two of three raters. Out of 126 comparisons coded, there were 18 disagreements that resulted in a discussion between the two primary coders. Of these, only 4 resulted in the consultation of Bruce E. Wampold.

### *Strategies to Test Relative Efficacy*

The inclusion of multiple treatments creates a number of problems that obscure a direct test of the hypothesis that there are no true differences between treatments intended to be therapeutic. Specifically, if we avoid classifying treatments into categories, the ordering of effects becomes arbitrary. If multiple treatments are included in the meta-analysis and there is no reference treatment, it is unclear how to

aggregate the results of a study wherein Treatment A is superior to Treatment B, and another study wherein Treatment B is inferior to Treatment C. To address this problem, we utilized the strategy described by Wampold et al. (1997). First, we randomly assigned the sign (+/-) to the effect size derived from any treatment comparison. This procedure necessarily resulted in an aggregate effect near zero, as the positive effects are balanced by the negative effects. Accordingly, the aggregate effect is not the appropriate statistic to gauge treatment differences. Rather, treatment equivalence was evaluated through an inspection of the distribution of effects. If there were no differences between treatments, then effect sizes should be homogeneously distributed about zero. If there were true treatment differences, there should be relatively many obtained effects far from the zero (i.e., more studies found difference among treatments than would have been expected if the true differences among treatments was zero), which would result in rejection of the null hypothesis of homogeneity (see Wampold et al., 1997).

Second, we assigned a positive sign to the obtained effect for the difference between treatments. Thus necessarily, the aggregate effect size would be positive even when the true difference among treatments is zero. For example, studies comparing the same two treatments with opposite findings would both be treated as positive, inflating the aggregate effect. Nevertheless, this aggregate forms a gross upper bound of the differences between alcohol use disorder treatments.

To rule out allegiance as a threat to any differences detected, we also coded each study for researcher allegiance. To do so, we used a coding protocol developed by S. Miller et al. (2008), which was based on previous reports addressing the impact of researcher allegiance (Gaffan et al., 1995; Luborsky et al., 1999). Two raters (different raters than those who coded effect sizes) inspected the Method and Introduction sections of each study (i.e., were blind to results) and rated each treatment on a 5-point scale. A rating of 0 indicated that no evidence of allegiance to treatment was available, and a rating of 4 indicated strong allegiance to treatment. If the raters differed by more than one point in the rating of any treatment, they discussed the discrepancy and then arrived at a consensus rating. If the rating still differed by more than one point after this discussion, the ratings were declared a "disagreement," and the average of the two ratings was used as the final rating. Of 67 independent ratings of treatments, 7 differed by more than one point, all of which were resolved by discussion. The intraclass correlation was  $\rho = .70$  ( $p > .001$ ), indicating the presence of adequate rater agreement on the particular allegiance score of a given treatment.

The final allegiance score for a particular comparison was the absolute value of the difference in allegiance ratings for the two respective treatments, providing an estimate of the degree to which allegiance was balanced in the comparison. For example, if both treatments were coded as zero, allegiance was considered balanced (even though treatment descriptions provided no evidence of allegiance, the difference between treatments was zero). Similarly, if both treatments in a comparison were coded as fours (i.e., high allegiance to both treatments), allegiance was still considered balanced.

### *Unit of Analysis and Calculation of Effect Size*

In the current analysis, the effect size was the standardized mean difference derived from a direct comparison of two psychological treatments. Consequently, studies that contained more than two

treatments provide more than one comparison. Thus, a study that contained three treatments (e.g., Treatments A, B, and C) resulted in three comparisons (viz., A vs. B, A vs. C, and B vs. C).

To determine the difference between alcohol use disorder treatments in a given comparison, we calculated an effect size for each comparison as well as an estimate of the variance. First, we calculated effect sizes for each alcohol-related outcome variable and then aggregated across these variables within a study (Hedges & Olkin, 1985). Only aggregating measures concerned with the use and abuse of alcohol likely biased the analyses toward finding differences between treatments, as aggregating across all outcome measures (e.g., quality of life, depression, etc.) may obscure true difference in the primary measures of interest. Specifically, the effect size  $g$  was calculated by calculating the difference between the posttreatment means for the each condition and dividing by the pooled standard deviation of both treatments, that is,

$$g = \frac{M_A - M_B}{s}, \quad (1)$$

where  $M_A$  and  $M_B$  were the means for Treatments A and B, respectively, and  $s$  was the pooled standard deviation.<sup>1</sup>

To correct for bias in  $g$ , we calculated  $d$ , which provided an unbiased estimate of the population effect size (Hedges & Olkin, 1985)

$$d = \left[ 1 - \frac{3}{4N - 9} \right] g, \quad (2)$$

where,  $N = n_A + n_B$ , the sum of the number of participants in Treatment A and in Treatment B. As a secondary analysis, we also calculated separate effect sizes for the studies that reported at least one estimate of posttreatment abstinence.

Each treatment comparison contributed one effect size to the meta-analysis. To arrive at a more precise estimate of the true difference between treatments compared within the same study, we aggregated all alcohol-related outcome measures collected directly after the completion of treatment (i.e., follow-up data were not considered), yielding a single effect size estimate for each comparison. Secondary measures of psychological well-being or physical health were not considered, as it was expected that the alcohol-related measures, those most often targeted by researchers in alcohol use disorder treatment studies, were those most likely to reveal differences between treatments. This procedure also accounted for the correlation between outcome measures (Hedges & Olkin, 1985, pp. 212–213). As the correlations between multiple outcome variables are rarely reported, we imputed an estimate of the correlation. In a study of the validity of measures of alcohol consumption, the average correlation of the measures was .40 (Grant, Tonigan, & Miller, 1995). Accordingly, a correlation of .40 was chosen to aggregate the effect sizes within comparisons.

### Statistical Analyses

As our meta-analysis involved the examination of effect size heterogeneity, we conducted a random effects meta-analysis, assuming that included studies were drawn from a larger population of studies (Hedges & Olkin, 1985) and thus allowing generalization to the population of studies rather than to the specific studies analyzed. Random effects meta-analysis can be considered a spe-

cial case of a multilevel model. Participants are nested within studies at the first level of the analysis. The studies themselves are treated as the second level, and variances are treated as known (Hox, 2002, pp. 139–149).

To evaluate the relative efficacy of alcohol use disorder treatments, we tested two multilevel models. The first was the unconditioned model in which effect sizes are not conditioned upon any study level variable. Specifically, at level one,

$$d_j = \delta_j + e_j, \quad (3)$$

where  $d_j$  is the outcome for an individual study,  $\delta_j$  is the true population effect size, and  $e_j$  is the variance of the errors (which in this case is known and provided by the estimate of the variability of the study effect size).

At level two,

$$\delta_j = \gamma_o + u_j, \quad (4)$$

where  $\delta_j$  is the estimate of true population effect size,  $\gamma_o$  is the grand mean of the effect sizes (i.e., the average effect), and  $u_j$  is the level two error. When randomly assigning a positive or negative sign (+/-) to each effect, the estimate,  $\gamma_o$  yields an effect that is necessarily close to zero; the test of treatment differences, as discussed above, is based on the variance component or the extent to which effects are homogeneously distributed about the grand mean  $\gamma_o$  (see Wampold et al., 1997). The test of variance component,  $\text{var}(u_j)$ , is provided by the  $H$  statistic, which estimates the extent to which effects deviate from the grand mean (Hedges & Olkin, 1985; Hox, 2002). The upper bound of treatment differences was also estimated with the unconditioned model.

As the  $H$  statistic only provides information in regards to the significance of heterogeneity, we conducted an additional test to quantify the extent of heterogeneity between studies. The  $I^2$  index quantifies the extent of heterogeneity by comparing the  $H$  value with its expected value if effects were homogenous, its degrees of freedom ( $df = k - 1$ ). If the  $H$  statistic is smaller than its degrees of freedom (i.e., the  $I^2$  is negative), then  $I^2$  is set to zero (i.e., there is no evidence of between-study heterogeneity). The  $I^2$  index can be interpreted as a percentage of heterogeneity, specifically, the part of total variance attributable to between-study variance (Higgins & Thompson, 2002). According to Higgins and Thompson (2002), an  $I^2$  of 25% is considered small, 50% is a medium-sized effect, and 75% is large.

The test of the effect of allegiance on treatment differences is provided by the conditioned model. In the conditioned model, the Level 1 model is similar to the earlier unconditioned. However, at Level 2 the equation is

$$\delta_j = \gamma_o + \gamma_I(\text{Allegiance}) + u_j, \quad (5)$$

where  $\gamma_o$  is the grand mean for studies with balanced allegiances (i.e., allegiance equals zero), and  $\gamma_I$  is the expected difference

<sup>1</sup> If researchers reported effects sizes in the form of proportions or odds ratios, we converted these estimates to the  $g$  statistic using the formula suggested by Chinn (2000). Specifically, Chinn demonstrated that a dichotomous outcome can be converted to a continuous effect size by computing the logit of the odds ratio effect size and dividing by 1.81. We estimated a variance of this effect size with the formula offered by Kleinbaum, Kupper, and Morgenstern (1982).

between treatments that differ in allegiance by one point. If researcher allegiance has a significant effect on treatment outcome, the fixed effect  $\gamma_i$  will be greater than zero, and  $\text{var}(u_j)$  will be significantly reduced relative to the unconditioned model.

Results

The literature search revealed 115 studies that appeared relevant to the current meta-analysis. Of these, 38 studies were coded as directly comparing at least two bona fide psychotherapies. Finally, 8 studies were excluded because data were insufficient to calculate an effect size. Consequently, 30 studies (47 effects) and 3,503 ( $M = 47.17$ , median = 19) patients were included in the meta-analysis. As described in the Method section, we estimated the relative efficacy of alcohol treatments with two models: (a) an unconditioned model, wherein effect sizes were not conditioned upon any study level variable, and (b) a conditioned model, wherein allegiance was entered as a predictor of variability in relative efficacy across studies. The results of the two models are presented in Table 2.

In the unconditioned model in which positive and negative signs were randomly assigned to each effect size, the grand mean,  $\lambda_o$ , was not significantly different from zero. However, as indicated earlier, this aggregate effect was necessarily close to zero and is not used to test treatment differences. Instead, the test of differences between compared treatments is the variability of effect sizes about the grand mean, estimated by the variance component. The variance component was .00 and corresponded to an  $H$  statistic of 51.46, which when compared to a chi-square distribution of 45 degrees of freedom was not significantly different from zero ( $p = .26$ ). The value of  $I^2$  was 10.61, which is considered small (Higgins & Thompson, 2002). Specifically, it appears that effect sizes were homogenously distributed about zero, and only 11% of the variability in the aggregate effect was due to between-study variability and not sampling error (see Table 2).

We also calculated an upper bound of treatment differences by aggregating the absolute value of each effect size. The upper bound of treatment differences, as provided by the grand mean, was quite small,  $\lambda_o = .11$ , but significantly greater than zero (see Table 3). A number needed to treat (NNT; see Kraemer & Kupfer, 2006) analysis indicated that even if the upper bound provided an

unbiased estimate of treatment differences, approximately 17 individuals would need to be treated with the superior treatment for one to experience benefit over the alternative treatment. Figure 1 provides the absolute value of the difference observed in each treatment comparison entered in the meta-analysis. Notice that the larger effects were those with generally larger standard errors.

Although effects were homogenously distributed about the grand mean, we also examined the possibility that allegiance might have an effect on relative efficacy. To do so, we conditioned effect sizes on the basis of an allegiance rating for each study. In the conditioned model, the grand mean remained not significantly different from zero, however, the effect of allegiance was significant ( $\lambda_i = .09$ ,  $p = .006$ ). This indicated that as the allegiance rating became more unbalanced in a given comparison (i.e., the researchers had a greater allegiance to one treatment than the other), the differences between treatments also increased. For example, the predicted effect size difference between some Treatment A with an allegiance rating of 4 and some Treatment B with an allegiance rating of 1 would be 0.29. In addition, an inspection of the  $I^2$  index indicated that conditioning effects on allegiance reduced the variability in effects by 100% (from 22% to 0%; see Table 2). An inspection of Figure 1 also illustrates the impact of allegiance on the predicted difference between treatment conditions by correcting the effect sizes for allegiance.

Finally, we repeated the previous unconditioned model in a reduced set of 17 studies (25 effects,  $n = 2,746$ ,  $M = 68.22$ , median = 22) that reported posttreatment estimates of abstinence from alcohol. An inspection of Tables 4 and 5 indicates that these analyses were comparable to those of the unconditioned model derived from alcohol-related outcome measures. Specially, effect sizes appeared to be homogenously distributed about zero,  $I^2 = 0.00$  (see Table 4), and the upper bound was quite small (see Table 5).

Discussion

The primary aim of this meta-analysis was to examine whether there were any outcome differences between bona fide psychological treatments for alcohol use disorders. In addition, we evaluated the role of allegiance as a potential moderator of treatment differences. Determining the relative efficacy of treatments for alcohol use disorders has both pragmatic and theoretical implications.

Table 2  
Tests of Homogeneity for Unconditioned and Allegiance Conditioned Models (Random Signs)

Effect type	Coefficient	Variance component	SE	df	t ratio	$\chi^2$ (H statistic)	p	$I^2$
Unconditioned model								
Fixed effect								
Grand mean $\lambda_o$	.02		.02		0.82		.42	
Random effect								
True effect size $\delta_j$		.00		46		51.46	.27	10.61
Model conditioned on allegiance								
Fixed effect								
Intercept $\lambda_o$	.02		.02		0.77		.44	
Allegiance $\lambda_i$	.09		.04		2.92		.01	
Random effect								
True effect size $\delta_j$		.00		45		42.94	>.50	0.00

Table 3  
*Tests of Homogeneity for the Unconditioned Model (Upper Bound)*

Effect type	Coefficient	Variance component	SE	df	t ratio	$\chi^2$ (H statistic)	p	I <sup>2</sup>
Unconditioned model								
Fixed effect								
Grand mean $\lambda_o$	.11		.03		4.08		.00	
Random effect								
True effect size $\delta_j$		.00		46		36.19	>.50	0.00

Specifically, if there are differences between treatments in decreasing drinking behavior, then it would make sense to conduct more specific analyses to determine which particular treatments are most effective. Treatment differences may also indicate that developers have captured some insight into the pathology of alcohol use disorders in their treatment rationale and related interventions, ultimately leading to a better understanding of the disorder. As treatment rationales often have divergent theoretical bases (e.g., BSCT and 12-step facilitation), one might expect to find evidence of treatment differences in alcohol use disorders.

Our analyses provided no evidence of differences among treatments for alcohol use disorders. Specifically, effect sizes were homogeneously distributed about zero.  $I^2$  estimates were in the small range, and the upper bound was quite small. This pattern of results was mirrored in a reduced set of studies that reported at least one estimate of abstinence. Although the variability of effects about zero was small, we also found evidence of an allegiance effect. Specifically, our analyses indicated that as allegiance to compared treatments became more unbalanced, the expected difference between treatments increased, in favor of the treatment for which there was researcher allegiance. Allegiance accounted for most of the variability in treatment differences in alcohol measures. This effect is particularly noteworthy given that the unconditioned models revealed that there was little variability in effect sizes to explain.

This meta-analysis is consistent with and extends the preponderance of psychotherapy research, indicating that there remains little evidence to suggest any one type of therapy is inferior to any other. Wampold et al.'s (1997) meta-analytic finding that psychotherapies are equivalent was derived from a large body of studies, some of which utilized samples that were not clinically representative. In addition, the inclusion of large numbers of studies that were heterogeneous in terms of disorder treated could have obscured true differences between treatments for certain disorders (Crits-Christoph, 1997). DeRubeis, Brotman, and Gibbons (2005) have claimed that examining relative efficacy blind to type of disorder "is akin to asking whether insulin or an antibiotic is better, without knowing the condition for which these treatments are to be given . . . . Alternatively, researchers should begin with a problem and ask how treatments compare in their effectiveness for that problem" (p. 175). This meta-analysis extends the findings of Wampold et al. (1997) into a specific clinical population.

There are several limitations to the findings of this meta-analysis. First, our results cannot be generalized to all possible psychological treatments for alcohol disorders, only those directly compared in randomized controlled clinical trials in which patients were assigned to one of at least two bona fide psychological treatments. Conse-

quently, findings do not imply that "all" treatments are equally effective. For example, our results do not extend to the relative efficacy of inpatient and outpatient clinics, as the studies included in our analysis focused on the relative efficacy of different treatments models, not the effects of different treatment settings. Although our inclusion criteria of directly comparing two bona fide psychological treatments resulted in the exclusion of a number of studies and treatment conditions, the database of studies analyzed here contained a wide variety treatments, including 12-step facilitation, motivational enhancement therapy, BSCT, aversion therapy, relapse prevention, and psychodynamic treatment, among others. Consequently, the results of our analysis are not likely the result of including multiple comparisons of highly similar treatments. Finally, only published studies were collected for this meta-analysis. The file-drawer phenomenon can be a problem in meta-analysis (Wilson, 2000). However, the problem of missing unpublished studies is more likely to be a problem in narrative reviews in which inclusion criteria are most commonly undefined (Quintana & Minami, 2006). Moreover, given the costs associated with conducting an RCT of two bona fide psychotherapies, it is likely that quality trials will be published. Finally, there is evidence to suggest that measures of meta-analytic heterogeneity are underpowered, especially when the number of effects in the meta-analysis is small. However, our study was adequately powered to detect heterogeneity of effect sizes if it were present ( $k > 20$ ; Huedo-Medina, Sanchez-Meca, Marin-Martinez, & Botella, 2006). Accordingly, it is not likely that our results were due to Type II error.

The failure to discover evidence of outcome differences between a diverse array of treatments leads to some speculation about how it is that psychological treatments for alcohol use disorders lead to changes in drinking behavior and bears particularly on the medical or "technology model" of psychotherapy (Carroll & Rounsaville, 1990; Wampold, 2001). Central to a medical model is specificity, which stipulates that the benefits of a particular treatment are due to aspects of the treatment that are unique (i.e., not solely attributable to non-specific or common factors effects). In the treatment of alcohol problems, this requirement might be paralleled by the notion that recovery from an alcohol use disorder is a result of moving successfully through each of the 12 steps or learning new more adaptive coping skills that decrease the need for alcohol to cope with stressful situations. The medical model assumes when a treatment is effective, it is because the treatment developer has achieved an insight into the pathogenesis of the disorder and designed a treatment to address the particular deficit (DeRubeis et al., 2005). Moreover, if multiple treatments are effective, it may be that they each have achieved some unique insight into the disorder. The treatment relies on spiritual formulations of alcohol problems and does not place require psychological training for practitioners.

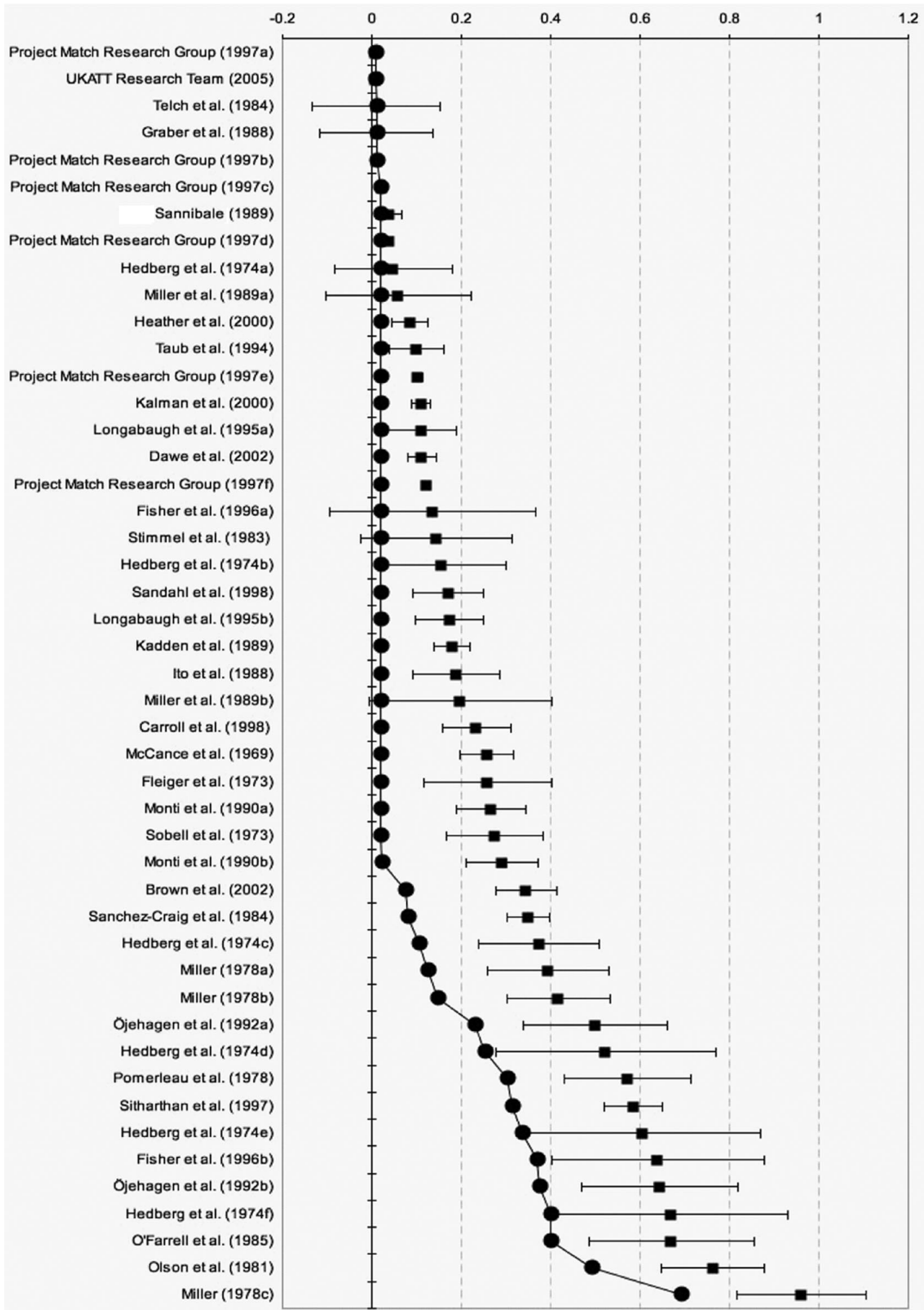


Figure 1. This figure provides the absolute value of the effect size (Cohen's *d*) for each treatment comparison. Square points indicate the actual effect size entered in the meta-analysis, and ovals indicate the absolute value of each effect size corrected for allegiance. Error bars provide a 95% confidence interval. We corrected for allegiance by calculating the predicted allegiance score for each comparison and multiplying by the allegiance coefficient ( $\lambda_1 = .09$ ; see Table 1). Finally, we subtracted this product from the observed effect size.



Table 4  
*Tests of Homogeneity for Unconditioned and Allegiance Conditioned Models (Random Signs, Abstinence Only)*

Effect type	Coefficient	Variance component	SE	df	t ratio	$\chi^2$ (H statistic)	p	I <sup>2</sup>
Unconditioned model								
Fixed effect								
Grand mean $\lambda_o$	.03		.04		0.711		.48	
Random effect								
True effect size $\delta_j$		.00		24		17.31	>.50	0.00
Model conditioned on allegiance								
Fixed effect								
Intercept $\lambda_o$	.01	.04			0.136		.89	
Allegiance $\lambda_f$	.08	.05			1.84		.08	
Random effect								
True effect size $\delta_j$		.00		23		13.91	>.50	0.00

Although a lack of evidence in regards to treatment differences cannot fundamentally resolve the polemic between the models of treatment, our findings suggest that additional head-to-head comparisons of bona fide psychological treatments are unlikely to provide further answers. In addition, our findings are consistent with: (a) mixed findings in regards to matching specific client diagnostic traits to treatment characteristics and (b) a failure to find consistent patterns of theory specific mediation and moderation of treatment effects (e.g., Berglund et al., 2003; Morgenstern & McKay, 2007; Project MATCH Research Group, 1997).

Morgenstern and McKay (2007) have argued that addictions treatment researchers have encountered the limitations of the technology model of psychotherapy and that the conception of non-specific mechanisms as semi-nuisance variables to be experimentally controlled has proven quite limited. Accordingly, it may be time for researchers to broaden the scope of inquiry to allow for an examination of other potential sources of treatment efficacy.

First, it may be useful to unyoke the study of potential therapeutic mechanisms described in a treatment rationale from those that may actually be operating in treatment. As an example, Owen et al. (2003) revealed that changes in self-efficacy predicted improvement in patients assigned to 12-step facilitation. Given that a philosophical tenet of 12-step models is admitting to powerlessness over alcohol, this finding suggests that the explicit rationale of a treatment may not be indicative of the psychological processes that occur in the treatment.

Second, there is a need for research that focuses on developing models of the complex interactions that occur between patients and

therapists (Morgenstern & McKay, 2007). Although the hypothesis of matching types of client to specific treatments has yielded equivocal results, recent matching research has abandoned a focus on specific treatment approaches for the study of how specific therapist behaviors interact with patient attributes (cf. Karno & Longabaugh, 2003). Additionally, treatment researchers typically ignore the therapist as a source of variability in clinical trials (Wampold & Serlin, 2000). This precludes an understanding of how effective therapists achieve desired outcomes that may be orthogonal from the theoretical approach. Moreover, as patients are often nested within therapists in clinical trials (patients see one therapist, and therapists see multiple patients), any correlation between a process measure and an outcome is necessarily a crude average of the between-therapist correlation and the within-therapist correlation. Disentangling these correlations is likely to provide more detailed information about the process of change in psychotherapy. For example, Baldwin, Wampold, and Imel (2007) reported a significant between-therapist correlation for alliance and outcome, but no significant within-therapist correlation. The more effective therapists were those that had consistently higher working alliance scores across patients.

It appears that despite several decades of comparative trials, evolving treatment philosophies and goals, as well as changes in treatment technology, there remains little evidence to suggest that bona fide psychological treatments differ in their effects. Consequently, research that looks beyond the therapeutic rationale as a guide to the psychological mechanisms responsible for change, to potentially more universal change factors may be increasingly beneficial.

Table 5  
*Tests of Homogeneity for the Unconditioned Model (Upper Bound, Abstinence Only)*

Effect type	Coefficient	Variance component	SE	df	t ratio	$\chi^2$ (H statistic)	p	I <sup>2</sup>
Unconditioned model								
Fixed effect								
Grand mean $\lambda_o$	.10		.04		2.53		.02	
Random effect								
True effect size $\delta_j$		.00		24		11.37	>.50	0.00

## References

- References marked with an asterisk indicate studies included in the meta-analysis.
- Anton, R. F., Moak, D. H., Latham, P., Waid, L. R., Myrick, H., Voronin, K., et al. (2005). Naltrexone combined with either cognitive behavioral or motivational enhancement therapy for alcohol dependence. *Journal of Clinical Psychopharmacology*, *25*, 349–357.
- Baldwin, S., Wampold, B. E., & Imel, Z. E. (2007). Untangling the alliance-outcome correlation: Exploring the relative importance of therapist and patient variability in the alliance. *Journal of Consulting and Clinical Psychology*, *75*, 842–852.
- Berglund, M., Thelander, S., Salaspuro, M., Franck, J., Andréasson, S., & Öjehagen, A. (2003). Treatment of alcohol abuse: An evidence-based review. *Alcoholism: Clinical and Experimental Research*, *27*, 1645–1656.
- \*Brown, T. G., Seraganian, P., Tremblay, J., & Annis, H. (2002). Process and outcome changes with relapse prevention versus 12-step aftercare programs for substance abusers. *Addiction*, *97*, 677–689.
- \*Carroll, K. M., Nich, C., Ball, S. A., McCance, E., & Rounsaville, B. J. (1998). Treatment of cocaine and alcohol dependence with psychotherapy and disulfiram. *Addiction*, *93*, 713–727.
- Carroll, K., & Rounsaville, B. (1990). Can a technology model of psychotherapy research be applied to cocaine abuse treatment? *NIDA Research Monograph*, *104*, 91–104.
- Chinn, S. (2000). A simple method for converting an odds ratio to effect size for use in meta-analysis. *Statistics in Medicine*, *19*, 3127–3131.
- Combine Study Research Group. (2003). Testing combined pharmacotherapies and behavioral interventions in alcohol dependence: Rationale and methods. *Alcoholism: Clinical and Experimental Research*, *27*, 1107–1122.
- Cooper, H., & Hedges, L. V. (Eds.). (1994). *The handbook of research synthesis*. New York: Russell Sage Foundation.
- 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.
- \*Dawe, S., Rees, V. W., Mattick, R., Sitharthan, T., & Heather, N. (2002). Efficacy of moderation-oriented cue exposure for problem drinkers: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *70*, 1045–1050.
- DeRubeis, R. J., Brotman, M. A., & Gibbons, C. J. (2005). A conceptual and methodological analysis of the nonspecifics argument. *Clinical Psychology: Science and Practice*, *12*, 174–183.
- Ditman, K. S., Crawford, G. G., Forgy, E. W., Moskowitz, H., & Macandrew, C. (1967). A controlled experiment on the use of court probation for drunk arrests. *American Journal of Psychiatry*, *124*, 160–167.
- Fals-Stewart, W., Birchler, G. R., & Kelley, M. L. (2006). Learning sobriety together: A randomized clinical trial examining behavior couples therapy with alcoholic female patients. *Journal of Counseling and Clinical Psychology*, *74*, 579–591.
- \*Fisher, M. S., & Bentley, K. J. (1996). Two group therapy models for clients with a dual diagnosis of substance abuse and personality disorder. *Psychiatric Services*, *47*, 1244–1250.
- \*Fleiger, D. L., & Zingle, H. W. (1973). Covert sensitization treatment with alcoholics. *Canadian Counsellor*, *7*, 269–277.
- 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.
- \*Graber, R. A., & Miller, W. R. (1988). Abstinence or controlled drinking goals for problem drinkers: A randomized clinical trial. *Psychology of Addictive Behaviors*, *2*, 20–33.
- Grant, K. A., Tonigan, J. S., & Miller, W. R. (1995). Comparison of three alcohol consumption measures: A concurrent validity study. *Journal of Studies on Alcohol*, *56*, 168–172.
- \*Heather, N., Brodie, J., Wale, S., Wilkinson, G., Luce, A., Webb, E., et al. (2000). A randomized controlled trial of moderation-oriented cue exposure. *Journal of Studies on Alcohol*, *61*, 561–570.
- \*Hedberg, A. G., & Campbell, L. (1974). A comparison of four behavioral treatments of alcoholism. *Journal of Behavior Therapy and Experimental Psychiatry*, *5*, 251–256.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine*, *21*, 1539–1558.
- Hox, J. (2002). *Multilevel analysis techniques and applications*. Hillsdale, NJ: Erlbaum.
- Huedo-Medina, T. B., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: *Q*-statistic or *I*<sup>2</sup> index? *Psychological Methods*, *11*, 193–206.
- Irvin, J. E., Bowers, C. A., Dunn, M. E., & Wang, M. C. (1999). Efficacy of relapse prevention: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, *67*, 563–570.
- \*Ito, J. R., Donovan, D. M., & Hall, J. J. (1988). Relapse prevention in alcohol aftercare: Effects on drinking outcome, change process, and aftercare attendance. *British Journal of Addiction*, *83*, 171–181.
- \*Kadden, R. M., Cooney, N. L., Getter, H., & Litt, M. D. (1989). Matching alcoholics to coping skills or interactional therapies: Post-treatment results. *Journal of Consulting and Clinical Psychology*, *57*, 698–704.
- \*Kalman, D., Longabaugh, R., Clifford, P. R., Beattie, M., & Maisto, S. A. (2000). Matching alcoholics to treatment: Failure to replicate finding of an earlier study. *Journal of Substance Abuse Treatment*, *19*, 183–187.
- Karno, M. P., & Longabaugh, R. (2003). Patient depressive symptoms and therapist focus on emotional material: A new look at Project MATCH. *Journal of Studies on Alcohol*, *64*, 607–615.
- Keso, L., & Salaspuro, M. (1990). Inpatient treatment of employed alcoholics: A randomized clinical trial on Hazelden-type and traditional treatment. *Alcohol Clinical Experimental Research*, *14*, 584–589.
- Kleinbaum, D. G., Kupper, L. L., & Morgenstern, H. (1982). *Epidemiologic research*. Belmont, CA: Lifetime Learning.
- Kownacki, R. J., & Shadish, W. R. (1999). Does alcoholics anonymous work? The results from a meta-analysis of controlled experiments. *Substance Use & Misuse*, *34*, 1897–1916.
- Kraemer, H. C., & Kupfer, D. J. (2006). Size of treatment effects and their importance to clinical research and practice. *Biological Psychiatry*, *59*, 990–996.
- \*Longabaugh, R., Wirtz, P. W., Beattie, M. C., Noel, N., & Stout, R. (1995). Matching treatment focus to patient social investment and support: 18-month follow-up results. *Journal of Consulting and Clinical Psychology*, *63*, 296–307.
- Luborsky, L., Diguier, 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.
- 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.
- \*McCance, C., & McCance, P. F. (1969). Alcoholism in north-east Scotland: Its treatment and outcome. *British Journal of Psychiatry*, *115*, 189–198.
- McCrary, B. S., Epstein, E. E., & Hirsch, L. S. (1999). Maintaining change after conjoint behavioral alcohol treatment for me: Outcomes at 6 months. *Addiction*, *94*, 1381–1396.
- Miller, S., Wampold, B. E., & Verhely, K. (2008). Direct comparisons of treatment modalities for youth disorders: A meta-analysis. *Psychotherapy Research*, *18*, 5–14.
- \*Miller, W. R. (1978). Behavioral treatment of problem drinkers: A comparative outcome study of three controlled drinking therapies. *Journal of Consulting and Clinical Psychology*, *46*, 74–86.

- Miller, W. R., Andrews, N. R., Wilbourne, P., & Bennett, M. E. (1998). A wealth of alternatives: Effective treatments for alcohol problems. In W. R. Miller & N. Heather (Eds.), *Treating addictive behaviors* (2nd ed.; pp. 203–216). New York: Plenum Press.
- \*Miller, W. R., & Dougher, M. J. (1989). Covert sensitization: Alternative treatment procedures for alcoholism. *Behavioural Psychotherapy*, *17*, 203–220.
- \*Monti, P. M., Abrams, D. B., Binkoff, J. A., & Zwick, W. R. (1990). Communication skills training, communication skills training with family and cognitive behavioral mood management training for alcoholics. *Journal of Studies on Alcohol*, *51*, 263–270.
- Morgenstern, J., Blanchard, K. A., Morgan, T. J., Labouvie, E., & Hayaki, J. (2001). Testing the effectiveness of cognitive-behavioral treatment for substance abuse in a community setting: Within treatment and post-treatment findings. *Journal of Consulting and Clinical Psychology*, *69*, 1007–1017.
- Morgenstern, J., & McKay, J. R. (2007). Rethinking the paradigms that inform behavioral treatment research for substance use disorders. *Addiction*, *102*, 1377–1389.
- \*O'Farrell, T. J., Cutter, H. S., & Floyd, F. J. (1985). Evaluating behavioral marital therapy for male alcoholics: Effects on marital adjustment and communication from before to after treatment. *Behavior Therapy*, *16*, 147–167.
- \*Öjehagen, A., Berglund, M., & Appel, C. (1992). A randomized study of long-term out-patient treatment in alcoholics: Psychiatric treatment versus multimodal behavioural therapy, during 1 versus 2 years of treatment. *Alcohol and Alcoholism*, *27*, 649–658.
- \*Olson, R. P., Ganley, R., Devine, V. T., & Dorsey, G. C. (1981). Long-term effects of behavioral versus insight-oriented therapy with inpatient alcoholics. *Journal of Consulting and Clinical Psychology*, *49*, 866–877.
- Owen, P. L., Slaymaker, V., Tonigan, J. S., McCrady, B. S., Epstein, E. E., Kaskutas, L. A., et al. (2003). Participation in alcoholics anonymous: Intended and unintended change mechanisms. *Alcoholism: Clinical and Experimental Research*, *27*, 524–532.
- \*Pomerleau, O., Pertschuk, M., Adkins, D., & Brady, J. P. (1978). A comparison of behavioral and traditional treatment for middle-income problem drinkers. *Journal of Behavioral Medicine*, *1*, 187–200.
- \*Project MATCH Research Group. (1997). Matching alcoholism treatments to client heterogeneity: Project MATCH post-treatment drinking outcomes. *Journal of Studies on Alcohol*, *58*, 7–29.
- Quintana, S. M., & Minami, T. (2006). Guidelines for meta-analyses of counseling psychology research. *Counseling Psychologist*, *34*, 839–877.
- Robinson, L. A., Berman, J. S., & Neimeyer, R. A. (1990). Psychotherapy for the treatment of depression: A comprehensive review of controlled outcome research. *Psychological Bulletin*, *108*, 30–49.
- Rosenthal, R. (1966). *Experimenter effects in behavioral research*. Appleton, WI: Century-Crofts.
- Rychtarik, R. G., Connors, G. J., Whitney, R. B., McGillicuddy, N. B., Fitterling, J. M., & Wirtz, P. W. (2000). Treatment settings for persons with alcoholism: Evidence for matching clients to inpatient versus outpatient care. *Journal of Consulting and Clinical Psychology*, *68*, 277–289.
- \*Sanchez-Craig, M., Annis, H. M., Bronet, A. R., & MacDonald, K. R. (1984). Random assignment to abstinence and controlled drinking: Evaluation of a cognitive-behavioral program for problem drinkers. *Journal of Consulting and Clinical Psychology*, *52*, 390–403.
- \*Sandahl, C., Herlitz, K., Ahlin, G., & Rönnerberg, S. (1998). Time-limited group psychotherapy for moderately alcohol dependent patients: A randomized controlled clinical trial. *Psychotherapy Research*, *8*, 361–378.
- \*Sannibale, C. (1989). A prospective study of treatment outcome with a group of male problem drinkers. *Journal of Studies on Alcohol*, *50*, 236–244.
- Shadish, W. R., Montgomery, L. M., Wilson, P., & Wilson, M. R. (1993). Effects of family and marital psychotherapies: A meta-analysis. *Journal of Consulting and Clinical Psychology*, *61*, 992–1002.
- Shadish, W. R., & Sweeney, R. B. (1991). Mediators and moderators in meta-analysis: There's a reason we don't let dodo birds tell us which psychotherapies should have prizes. *Journal of Consulting and Clinical Psychology*, *59*, 883–893.
- \*Sitharthan, T., Sitharthan, G., Hough, M. J., & Kavanagh, D. J. (1997). Cue exposure in moderation drinking: A comparison with cognitive-behavior therapy. *Journal of Consulting and Clinical Psychology*, *65*, 878–882.
- \*Sobell, M. B., & Sobell, L. C. (1973). Individualized behavior therapy for alcoholics. *Behavior Therapy*, *4*, 49–72.
- \*Stimmel, B., Cohen, M., Sturiano, V., Hanbury, R., Korts, D., & Jackson, G. (1983). Is treatment for alcoholism effective in persons on methadone maintenance? *The American Journal of Psychiatry*, *140*, 862–866.
- \*Taub, E., Steiner, S. S., Weingarten, E., & Walton, K. G. (1994). Effectiveness of broad spectrum approaches to relapse prevention in severe alcoholism: A long-term, randomized, controlled trial of transcendental meditation, EMG biofeedback and electronic neurotherapy. *Alcoholism Treatment Quarterly*, *11*, 187–220.
- \*Telch, M. J., Hannon, R., & Telch, C. F. (1984). A comparison of cessation strategies for the outpatient alcoholic. *Addictive Behaviors*, *9*, 103–109.
- Thase, M. E. (1999). What is the investigator allegiance effect and what should we do about it? *Clinical Psychology: Science and Practice*, *6*, 113–115.
- \*UKATT Research Team. (2005). Effectiveness of treatment for alcohol problems: Findings of the randomized UK Alcohol Treatment Trial (UKATT). *British Medical Journal*, *331*, 541–546.
- Walsh, D. C., Hingson, R. W., Merrigan, D. M., Levenson, S. M., Cupples, L. A., & Heeren, T. (1991). A randomized trial of treatment options for alcohol-abusing workers. *The New England Journal of Medicine*, *325*, 775–782.
- Walters, G. D. (2000). Behavioral self-control training for problem drinkers: A meta-analysis of randomized control studies. *Behavior Therapy*, *31*, 135–149.
- Wampold, B. E. (2001). *The great psychotherapy debate: Methods, models, and findings*. Hillsdale, NJ: Erlbaum.
- Wampold, B. E., Mondin, G. W., Moody, M., Stich, F., Benson, K., & Ahn, H. (1997). A meta-analysis of outcome studies comparing bona fide psychotherapies: Empirically, "all must have prizes". *Psychological Bulletin*, *122*, 203–215.
- Wampold, B. E., & Serlin, R. C. (2000). The consequences of ignoring a nested factor on measures of effect size in analysis of variance. *Psychological Methods*, *5*, 425–433.
- Wilbourne, P. L., & Miller, W. R. (2002). Treatment for alcoholism: Older and wiser? *Alcoholism Treatment Quarterly*, *20*, 41–59.
- Wilson, D. B. (2000). Meta-analyses in alcohol and other drug abuse treatment research. *Addiction*, *95*, S419–S439.

Received October 29, 2007

Revision received June 4, 2008

Accepted June 19, 2008 ■