Cognitive-Behavioral Treatment With Adult Alcohol and Illicit Drug Users: A Meta-Analyses of Randomized Controlled Trials*

MOLLY MAGILL, PH.D.,† AND LARA A. RAY, PH.D.†

Center for Alcohol and Addiction Studies, Brown University, Box G S121-4, Providence, Rhode Island 02912

ABSTRACT. Objective: This meta-analysis examined 53 controlled trials of cognitive-behavioral treatment (CBT) for adults diagnosed with alcohol- or illicit-drug-use disorders. The aims were to provide an overall picture of CBT treatment efficacy and to identify client or treatment factors predictive of CBT effect magnitude. Method: The inverse variance weighted effect size (Hedges' g) was calculated for each study and pooled using fixed and random effects methods. Potential study-level moderators were assessed in subgroup analyses by primary drug, type of CBT, and type of comparison condition. In addition, seven client and treatment variables were examined in meta-regression analyses. Results: Across studies, CBT produced a small but statistically significant treatment effect (g = 0.154, p < .005). The pooled effect was somewhat lower at 6-9 months (g = 0.115, p < .005) and continued to diminish at 12-month follow-up (g = 0.096, p < .05). The effect of CBT was largest in marijuana studies (g = 0.513, p < .005) and in studies with a no-treatment control as the comparison condition (g = 0.796, p < .005). Meta-regression analyses indicated that the percentage of female participants was positively associated and the number of treatment sessions was negatively associated with effect size. Conclusions: The findings demonstrate the utility of CBT across a large and diverse sample of studies and under rigorous conditions for establishing efficacy. CBT effects were strongest with marijuana users, when CBT was compared with no treatment, and may be larger with women than with men and when delivered in a brief format. (J. Stud. Alcohol Drugs 70: 516-527, 2009)

Cognitive-behavioral treatment (CBT) models are among the most extensively evaluated interventions for alcohol- or illicit-drug-use disorders. Based primarily on Marlatt and Gordon's (1985; Marlatt and Donovan, 2005) model of relapse prevention, these treatments target cognitive, affective, and situational triggers for substance use and provide skills training specific to coping alternatives. CBT treatment for alcohol or illicit drug use often includes the following strategies: (1) identifying intrapersonal and interpersonal triggers for relapse, (2) coping-skills training, (3) drug-refusal skills training, (4) functional analysis of substance use, and (5) increasing nonuse-related activities. These models have been manualized (e.g., Carroll, 1998; Kadden et al., 1992; Monti et al., 1989) and adapted for implementation in a variety of clinical settings. Further, CBT interventions have been tested in Stage III research to examine their utility in the “real-world” context, possible adaptations, and cost-effectiveness (National Institute on Drug Abuse, 1992; cited from Carroll and Onkin, 2005). As argued by Carroll and Rounsaville (2007), the addictions field would benefit from greater and continued attention devoted to the dissemination of research-based substance-use treatments. Review data to guide such efforts, however, are needed.

CBT interventions for substance-use disorders have generally received empirical support, yet their effectiveness as a whole has not been subjected to recent systematic review. Meta-analysis is a promising method of research synthesis useful for large bodies of research that may show disparate results (Lipsey and Wilson, 2001). To date, qualitative reviews have concluded that CBT is more effective than no treatment but have shown mixed results regarding questions of implementation (Monti and Rohsenow, 2003), durability of effects or possible delayed effects (Carroll, 1996; Carroll and Onkin, 2005; Miller et al., 2005), and efficacy over other treatments (Longabaugh and Morgenstern, 1999).

A search of the meta-analytic literature yielded one previous review on cognitive-behavioral alcohol or illicit drug treatment. Irvin and colleagues (1999) built on Carroll's 1996 review of relapse prevention by examining 26 experimental and quasi-experimental studies across drugs of abuse, including nicotine. They reported a small overall effect size for substance use, a medium effect for psychosocial outcomes, and suggested that no bias resulted from inclusion of only published research. Effect sizes were largest in quasi-experimental studies, in studies measuring outcome immediately following treatment, and in studies with self-reported outcomes. The review suggested that relapse prevention was more effective for alcohol-use disorders than for other substance-use disorders and when delivered in combination with pharmacological intervention. The combination of pharma-

Received: June 19, 2008. Revision: February 24, 2009.

*This research was supported by a T32 (AA07459-22) grant awarded by the National Institute on Alcohol Abuse and Alcoholism.

†Correspondence may be sent to Molly Magill at the above address or via email at: molly_magill@brown.edu. Lara A. Ray is now with the Department of Psychology, University of California at Los Angeles, Los Angeles, California.
collogial and psychosocial treatments has received increased attention, particularly in the field of alcoholism. The efficacy of relapse-prevention pharmacotherapies in combination with CBT strategies warrants additional review.

Given the empirical and clinical proliferation of CBT, the absence of an updated meta-analysis on these approaches to alcohol or illicit drug treatment is surprising. Psychosocial addictions treatments, other than CBT, have received relatively more attention from meta-analytic inquiry. A review of this literature yielded four meta-analyses of brief motivational interventions (Burke et al., 2003; Dunn et al., 2001; Harvard et al., 2007; Moyer et al., 2002), three on methods based in contingency management (Griffith et al., 2000; Lussier et al., 2006; Prendergast et al., 2006), three studies on marital or family-based interventions (Edwards and Steinglass, 1995; Powers et al., 2008; Stanton and Shadish, 1997), and three on self-help approaches (Emrick et al., 1993; Kownacki and Shadish, 1999; Tonigan et al., 1996). Moreover, Butler et al. (2006) examined the current state of meta-analytic evaluation across 16 studies of CBT for psychiatric disorders and noted a need for meta-analysis specific to substance using populations.

The present meta-analysis provides a broad view of CBT efficacy for adults diagnosed with alcohol or illicit drug abuse or dependence. There are a number of promising CBT approaches available (McCraday, 2000) and their combination with pharmacological (Carroll and Onken, 2005) or additional psychosocial (Longabaugh and Morgenstern, 1999) treatments may hold greater promise than either type of treatment alone. This meta-analysis updates previous reviews and additionally includes research on combined cognitive-behavioral interventions. The objectives were as follows: (1) to provide a broad picture of CBT efficacy, (2) to clarify potential design characteristics that may inflate or diminish effect size and (3) to explore client or treatment factors as moderators of outcome, which can inform future dissemination efforts.

Method

Study inclusion

A number of criteria were used to select studies for this meta-analysis. First, studies had to be randomized controlled trials that used psychometrically established outcome measurement. Next, the treatment delivered was identified as cognitive-behavioral, relapse prevention, or coping-skills training. In addition, the CBT treatment could be either individual or group format and delivered alone or in combination with one or more treatments, including pharmacological treatment. The target population was adults (ages 18 and older) with a primary diagnosis of alcohol or illicit drug abuse or dependence as determined by the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (American Psychiatric Association, 1994). Finally, the studies were English language and published between 1980 and 2006 (inclusive).

Literature search

A literature search was conducted to identify eligible studies. First was a title, abstract, keyword, and subject search of treatment terms (e.g., cognitive-behavioral, relapse prevention, coping skills) and outcome targets (e.g., alcohol, cocaine, methamphetamine, stimulant, opiate, heroin, marijuana, cannabis, illicit drug, substances) in six databases (Campbell Collaboration, Cochrane Collaboration, PubMed, PsychINFO, Social Services Abstracts, and Social Work Abstracts). Second was a bibliographic search of qualitative and quantitative reviews on cognitive-behavioral or general substance-dependence treatment (i.e., Carroll, 1996, 1999; Donovan, 2003; Irvin et al., 1999; Longabaugh and Morgenstern, 1999; Marlatt and Witkiewitz, 2005; Miller et al., 2003; Monti and Rosehow, 2003; Morgenstern and Longabaugh, 2000; Prendergast et al., 2002). Third was a broad, all text or any field database search (PsychINFO, PubMed) to check for studies not identified by the previous methods. Finally, there was bibliographic review of articles derived at all search stages. Figure 1 provides a visual summary of study inclusion, which is consistent with QUOROM guidelines (Moher et al., 1999). The final meta-analytic sample comprised 59 research reports, describing 52 studies, and contributing 53 effect sizes, to result in an N of 9,308 individuals.

Effect size calculation

The standardized mean difference (Hedges’ g) was used to measure the relative effectiveness of CBT over comparison conditions for treating adult substance-use disorders. Conceptually, it is an estimate of treatment effect significance and magnitude expressed in standard deviation units. Hedges’ g has sound statistical properties in samples as small as 20 participants (Hedges, 1994) and includes a correction, f, for slight upward bias in estimated population effect (a distinguishing property from Cohen’s d). The formulae were:

\[ g_i = \frac{M_i - M_c}{s_p} \times \left(\frac{1}{f}\right) \]

where \( f = 1 - \frac{3}{(4 \times df - 1)} \)

and

\[ s_p = \sqrt{\frac{(n_i - 1)s_i^2 + (n_c - 1)s_c^2}{n_i + n_c - 2}} \]

\( M_i \) and \( M_c \) are the group means for the treatment and comparison, respectively; \( s_p \) is the pooled standard deviation; and \( s_i \) and \( s_c \) are the group standard deviations. In addition, effect sizes were inverse variance weighted before pooling, which afforded larger studies more influence on the pooled effect size (Hedges and Olkin, 1985).
An effect size was calculated for each study with the exception of the outpatient and aftercare arms of Project MATCH (1997), which contributed two effects to the pooled estimate \( (N = 53) \). Most were posttreatment between-group effect sizes, but 16 studies reported only follow-up data. For studies with follow-up outcomes \( (n = 34) \), effect sizes were additionally calculated at the first time-point (i.e., follow-up between-group effect size). The outcome indicator for effect size calculation was selected (Lipsey and Wilson, 2001; Wilson, 2000) in the following order: (1) biological measures,
(2) measures of use frequency, and (3) sample proportions. Effect sizes were reverse scored where necessary to ensure consistency in effect direction across studies (e.g., number of days drank). When outcome data were not reported in means and standard deviations, test statistics (e.g., t, F, r) were transformed into the standardized mean difference (see, e.g., Lipsey and Wilson, 2001). When dichotomous outcomes were presented (e.g., number of abstinent participants), odds ratios were calculated and similarly transformed using methods described by Chinn (2000). Finally, in studies that involved more than two comparison groups, effect sizes were calculated for the experimental group versus each comparison and then averaged to obtain a single effect per study. The potential impact of type of comparison condition on estimate magnitude was addressed in subgroup moderator analyses.

**Moderator coding procedure**

There were 10 variables examined as potential moderators of CBT effectiveness in subgroup and meta-regression analyses. These variables were coded by the first and second authors with a 25% random selection of studies double coded to establish rater agreement (n = 13). A minimum threshold of α = .70 for continuous codes (e.g., sample mean age) and 70% agreement for categorical codes (e.g., treatment delivery type) was required for a variable to be included in the analyses. Alpha values were in the excellent range (α = .97-.1.0). Percentage agreement for categorical variables ranged from 77% to 92%, with the exception of nonsignificant group differences at baseline (64% agreement), which was not included in the analyses. Variable coding guidelines were outlined in a codebook available, on request, from the first author.

**Data analysis**

**Model of inference and heterogeneity.** In calculating combined effect sizes, alcohol- and illicit-drug-use outcomes were considered fixed effects. Specifically, it was assumed that CBT effect sizes represented a single population or a distribution of populations with between-study heterogeneity that could be explained by known moderators. The significance of the Cochrane $Q$ test for heterogeneity determined whether this model of inference was tenable, that is, was it valid to combine these studies? If the null hypothesis was retained ($p > .05$), the studies were considered homogeneous and fixed effects values were the appropriate estimators. If rejected, a priori moderators were examined, and if the $Q$ value remained significant, random effects values were considered the better estimates. Hedges and Vevea (1998) describe this model of inference as conditionally random. In addition, $I^2$ values were provided as descriptors of proportion of between-study variability (Higgins and Thompson, 2002) along with the pooled effect values.

**Sensitivity analysis and publication bias.** Three types of sensitivity analyses were conducted. Heterogeneity and moderator analyses were the two primary methods for examining effect size validity and stability. However, trimmed estimates, with high weight or outlier studies removed (Baujat et al., 2002), were additionally provided with moderator subgroup data. Together, the $Q$ test, fixed and random effects values as well as trimmed effect sizes by moderator provided a thorough view of effect size stability across pooling methods.

To test for possible publication bias, two tests were conducted. First, the relationship between study precision and effect size was assessed using a rank-order correlation test (Begg and Mazumdar, 1994). In the rank-order test, small sample/less precise studies are assumed to be published only when they show large effects, resulting in a significant and negative correlation when publication bias is present in the meta-analytic sample. Next, the more commonly used fail-safe $N$ (Rosenthal, 1991), an estimate of a hypothetical number of null studies required to change the observed effect size to an insignificant value, was calculated.

**Moderator analyses.** For moderator analyses, three variables were examined in pooled subgroups: (1) primary outcome (alcohol, marijuana, cocaine/stimulant/opiate, or polydrug), (2) treatment type (cognitive-behavioral, cognitive-behavioral combined with pharmacotherapy, or cognitive-behavioral combined with another psychosocial treatment), and (3) comparison type (no-treatment or waitlist control, passive or usual service comparison, theoretically active comparison, and no cognitive-behavioral adjunct comparison). This latter subgroup included studies of cognitive-behavioral intervention added to another psychosocial treatment where the comparison was that treatment alone.

Given that Irvin and colleagues (1999) found larger effects in studies with posttreatment and self-report measurement, effect size data were regressed on primary study client and treatment characteristics, and these two design variables as well as posttreatment attrition rate were examined as covariates. Client variables were demographic and diagnostic: age, percentage female participants, inclusion of co-occurring non-substance-related disorders, and alcohol or illicit drug outcome. The treatment variables were delivery (as standalone or as aftercare), format (individual or group), and length (number of sessions). Missing variable codes were mean imputed, and a predictor was removed from the analysis if imputed values reached 20% of total cases (Piggott, 1994). Analyses were conducted using Wilson’s (2005) METAREG for weighted least squares and maximum likelihood regression in SPSS Version 15 (SPSS Inc., Chicago, IL). Variables with significant regression coefficients were placed into a final predictive model and, if the model contained significant residual heterogeneity, maximum likelihood random effects analyses were conducted.
Results

Sample descriptives

The sample for this meta-analysis consisted of 53 randomized trials of CBT for adults diagnosed with substance-use disorders published between 1982 and 2006. Of these studies, 25% were published outside of the United States. The mean sample size was 179 participants (range: 20-1,656). The treatment targets were: alcohol (n = 23), cocaine/stimulants (n = 11), polydrug (n = 11), marijuana (n = 6), and opiates (n = 2). The samples’ mean (SD) age was 38 (5.7) years. Studies included a moderate percentage of typically under-represented groups such as women (mean = 29.2% [21.9%]) and ethnic or racial minorities (25.9% [27.8%]). The majority of studies enrolled only individuals with a diagnosis of alcohol or drug dependence (80.1%), whereas the remaining studies also allowed individuals with a diagnosis of abuse. Approximately 64% of studies allowed nonsubstance-related co-occurring diagnoses (exclusive of suicidal or homicidal ideation and active psychosis). A slight majority of interventions were delivered in an individual format (57.7%), there was a mean of 18 (11.9) (range: 1-48) sessions, and these treatments were evenly distributed regarding delivery as standalone (n = 31) or aftercare following other services (n = 22). Table 1 provides an overview of study design characteristics, effect sizes and corresponding confidence intervals (CIs). These studies were methodologically rigorous with acceptable attrition rates (mean = 19.3% [12.9%]), and high rates of both biologically validated outcomes (75% of studies) and manualized treatment delivery (98% of studies).

Main treatment effect

Overall, the pooled CBT effect was \( g = 0.144 \) (95% CI: 0.094, 0.194, \( p < .005; N = 53 \)), which represents a small effect size according to guidelines suggested by Cohen (1977). The test for homogeneity was rejected at the .005 level (\( Q = 128.85, 52 \mathrm{df} \)), and the \( F \) value showed that 60% of the variance in effect sizes could be explained by differences between studies (a moderate level of heterogeneity; Higgins et al., 2003). The random effects estimate was slightly higher (\( g = 0.154, 95\% \mathrm{CI}: 0.066, 0.242, p < .005 \)) than the fixed effects estimate but was consistent regarding overall interpretation of magnitude. Because of the observed heterogeneity, the random effects value was the better estimator of CBT effect with alcohol- or other drug-use disorders.

The sample of studies reporting follow-up outcomes for CBT showed an effect size that was slightly lower (\( g = 0.108, 95\% \mathrm{CI}: 0.051, 0.165, p < .005; n = 34 \)) than that reported for the overall posttreatment sample and that was heterogeneous (\( Q = 56.95, 33 \mathrm{df}, p < .05 \)). Follow-up time-points ranged from 1 to 12 months, and analyses conducted separately by time showed that a significant portion of heterogeneity could be accounted for by two trials reporting data at 1 (Kelly et al., 2000) and 3 (Tucker et al., 2004) months following treatment. Analyses with Kelly et al. (2000) and Tucker et al. (2004) removed showed a larger effect at 6-9 months (fixed \( g = 0.115, p < .005; n = 23 \)) that diminished at 12 months (fixed \( g = 0.096, p < .05; n = 9 \)). These analyses supported the assumption of homogeneity, suggesting that time of outcome assessment was an important predictor of between-study variance. Therefore, the subgroup values at 6-9 and 12 months represent the optimal estimates of CBT effect at follow-up.

Publication bias

Our results did not suggest the presence of publication bias in the sample of studies reviewed. The rank-order correlation (Begg and Mazumdar, 1994) showed a negative, but nonsignificant, relationship between precision and effect size (\( r = -.03, p > .10 \), one tailed). Moreover, the fail-safe \( N \) indicated that at an alpha level of .05, 340 unpublished null studies would be required to reduce the overall observed estimate to statistical nonsignificance, which lends additional support for an absence of publication bias in the present review.

Subgroup moderators

Table 2 presents results for subgroup moderators and sensitivity analyses. First, for CBT across alcohol or other drug outcomes, pooled effect sizes were small with the exception of CBT with marijuana use, which had a moderate and homogenous effect size (fixed \( g = 0.513, p < .005; n = 6 \)). Second, studies of CBT combined with additional psychosocial treatment showed a larger effect size (random \( g = 0.305, p < .005; n = 19 \)) than for CBT combined with pharmacological treatment (fixed \( g = 0.208, p < .005; n = 13 \)) and for CBT alone (random \( g = 0.172, p < .05; n = 21 \)). Finally, a large effect size was found for CBT in comparison to no treatment (random \( g = 0.796, p < .005; n = 6 \)). A consistent, small-sized, effect was found across other types of comparison condition (i.e., usual services, active treatment) with the exception of CBT adjunctive treatment. The effect of CBT as an adjunct to a psychosocial treatment compared with psychosocial treatment alone was negative and nonsignificant (random \( g = -0.054, \text{NS}; n = 7 \)).

Regression moderators

Of the three methodological covariates studied—(1) type of outcome assessment (self-report or biological), (2) time of follow-up assessment (posttreatment to 4 months or 6-12 months), and (3) posttreatment attrition rate—only outcome type and follow-up time were related to effect size (\( b = -.163, p < .005; b = -.129, p < .05 \), respectively). Table 3 summarizes findings for the seven client and treatment characteristics, effect sizes and corresponding confidence intervals (CIs). These studies were methodologically rigorous with acceptable attrition rates (mean = 19.3% [12.9%]), and high rates of both biologically validated outcomes (75% of studies) and manualized treatment delivery (98% of studies).

Main treatment effect

Overall, the pooled CBT effect was \( g = 0.144 \) (95% CI: 0.094, 0.194, \( p < .005; N = 53 \)), which represents a small effect size according to guidelines suggested by Cohen (1977). The test for homogeneity was rejected at the .005 level (\( Q = 128.85, 52 \mathrm{df} \)), and the \( F \) value showed that 60% of the variance in effect sizes could be explained by differences between studies (a moderate level of heterogeneity; Higgins et al., 2003). The random effects estimate was slightly higher (\( g = 0.154, 95\% \mathrm{CI}: 0.066, 0.242, p < .005 \)) than the fixed effects estimate but was consistent regarding overall interpretation of magnitude. Because of the observed heterogeneity, the random effects value was the better estimator of CBT effect with alcohol- or other drug-use disorders.

The sample of studies reporting follow-up outcomes for CBT showed an effect size that was slightly lower (\( g = 0.108, 95\% \mathrm{CI}: 0.051, 0.165, p < .005; n = 34 \)) than that reported for the overall posttreatment sample and that was heterogeneous (\( Q = 56.95, 33 \mathrm{df}, p < .05 \)). Follow-up time-points ranged from 1 to 12 months, and analyses conducted separately by time showed that a significant portion of heterogeneity could be accounted for by two trials reporting data at 1 (Kelly et al., 2000) and 3 (Tucker et al., 2004) months following treatment. Analyses with Kelly et al. (2000) and Tucker et al. (2004) removed showed a larger effect at 6-9 months (fixed \( g = 0.115, p < .005; n = 23 \)) that diminished at 12 months (fixed \( g = 0.096, p < .05; n = 9 \)). These analyses supported the assumption of homogeneity, suggesting that time of outcome assessment was an important predictor of between-study variance. Therefore, the subgroup values at 6-9 and 12 months represent the optimal estimates of CBT effect at follow-up.

Publication bias

Our results did not suggest the presence of publication bias in the sample of studies reviewed. The rank-order correlation (Begg and Mazumdar, 1994) showed a negative, but nonsignificant, relationship between precision and effect size (\( r = -.03, p > .10 \), one tailed). Moreover, the fail-safe \( N \) indicated that at an alpha level of .05, 340 unpublished null studies would be required to reduce the overall observed estimate to statistical nonsignificance, which lends additional support for an absence of publication bias in the present review.

Subgroup moderators

Table 2 presents results for subgroup moderators and sensitivity analyses. First, for CBT across alcohol or other drug outcomes, pooled effect sizes were small with the exception of CBT with marijuana use, which had a moderate and homogenous effect size (fixed \( g = 0.513, p < .005; n = 6 \)). Second, studies of CBT combined with additional psychosocial treatment showed a larger effect size (random \( g = 0.305, p < .005; n = 19 \)) than for CBT combined with pharmacological treatment (fixed \( g = 0.208, p < .005; n = 13 \)) and for CBT alone (random \( g = 0.172, p < .05; n = 21 \)). Finally, a large effect size was found for CBT in comparison to no treatment (random \( g = 0.796, p < .005; n = 6 \)). A consistent, small-sized, effect was found across other types of comparison condition (i.e., usual services, active treatment) with the exception of CBT adjunctive treatment. The effect of CBT as an adjunct to a psychosocial treatment compared with psychosocial treatment alone was negative and nonsignificant (random \( g = -0.054, \text{NS}; n = 7 \)).

Regression moderators

Of the three methodological covariates studied—(1) type of outcome assessment (self-report or biological), (2) time of follow-up assessment (posttreatment to 4 months or 6-12 months), and (3) posttreatment attrition rate—only outcome type and follow-up time were related to effect size (\( b = -.163, p < .005; b = -.129, p < .05 \), respectively). Table 3 summarizes findings for the seven client and treatment characteristics, effect sizes and corresponding confidence intervals (CIs). These studies were methodologically rigorous with acceptable attrition rates (mean = 19.3% [12.9%]), and high rates of both biologically validated outcomes (75% of studies) and manualized treatment delivery (98% of studies).
variables while controlling for assessment type and time. In the client model, the percentage of female participants had a positive association ($b = .005$, $p < .05$) and, in the treatment model, length of treatment had a negative association ($b = -.008$, $p < .005$) with effect size. The final fixed effects model accounted for 20.5% of the variance in CBT effect sizes, and these findings held in sensitivity analyses with two high-weight trials (Project MATCH, 1997; Anton et al., 2006) removed from the analysis. However, the $Q$ value for the residual was significant, indicating that the a priori moderators could not explain a fixed population effect. The final maximum likelihood model (with the random effects $\tau^2$ estimate of .04 added to individual inverse variance weights) resulted in an $R^2$ value of .169 ($p < .05$), thereby accounting for 16.9% of the variance in CBT effectiveness with adult alcohol- or illicit-drug-use disorders.

### Discussion

Across a large, diverse, and rigorous sample of randomized trials, CBT for adult substance-use disorders demonstrated a small, but statistically significant, effect over comparison conditions. Meta-analyses of other alcohol or illicit drug treatments show effect sizes generally in the small to moderate range (e.g., Burke et al., 2003; Prendergast et al., 2002, 2006), and only a slightly larger effect was found in Irvin and colleagues’ (1999) meta-analysis of relapse prevention (four studies in this review overlap with those examined by Irvin et al.: Carroll et al., 1994; Hawkins et al., 1986; O’Malley et al., 1992; Sobell et al., 1995). Meta-analytic review is often late-stage evaluation research, and the data provided should therefore be of practical clinical and empirical value. To better illustrate the impact of CBT,
the $U_j$ index was calculated. This index transforms the effect size to a “success percentage,” or the percentage of treated participants that performed better than the median for the comparison group (Rosenthal and Rubin, 1982). The $U_j$ value for this meta-analysis indicated that 58% of patients receiving CBT fared better than patients in the comparison condition.

Across drugs of abuse, types of CBT treatment, and types of comparison condition, pooled effect sizes were small and fell primarily within a similar range. The exception was marijuana-use disorders, which had a moderate and homogeneous effect, and a $U_j$ value of 69%. Irvin and colleagues (1999) reported the highest effects for alcohol, but their review occurred when marijuana research was in its infancy.

Studies of CBT combined with an additional psychosocial treatment had a larger effect than either CBT combined with pharmacological treatment or CBT alone in both fixed and random effects estimates. Again, this is in contrast with Irvin et al. (1999), who reported greater effects for relapse prevention plus pharmacotherapy than for relapse-prevention only, but combined psychosocial treatments were not included in their meta-analytic sample. In general, large departures from their meta-analytic sample. In general, large departures from
or similar no-treatment-control. In light of these findings, it is important to note that no-treatment-control studies were a minority among those examined, and that effect size estimates were also positive and significant for both active and passive or treatment-as-usual comparisons.

Treatment effects for CBT diminished over time, with somewhat lower effects at 6- to 9-month follow-up and markedly diminished effects at 12 months. This finding is consistent with previous meta-analytic research on relapse prevention (Irvin et al., 1999) and does not support a delayed emergence of treatment effects suggested by Carroll (1996). The CBT focus on ongoing coping without the use of substances would possibly place it among treatments suited toward longer-term outcomes, and this may be why it is often used as aftercare. This review (as well as a previous review, Longabaugh and Morgenstern, 1999) has not found support for aftercare as a uniquely beneficial format for clinical delivery. Thus, future research will need to clarify the mechanisms of ongoing reduction in substance use (e.g., continued coping-skill use) in CBT to inform treatment adaptations that promote longer-term treatment gains.

Among the client factors examined, women appeared to benefit more from CBT than men. Dumaine (2003) also found a positive association between female participants and effect size in a meta-analysis of treatment with individuals with co-occurring substance and other mental health disorders. Dumaine’s findings were based on bivariate correlations, and this meta-analysis may be more conservative given that additional client and study design variables were accounted for in the analyses. Assessment of primary studies with female-only samples, however, highlights additional study characteristics to consider when interpreting this result.

Table 2. Main treatment effect by primary drug, type of CBT treatment, and type of comparison condition

<table>
<thead>
<tr>
<th>Variable</th>
<th>Alcohol</th>
<th>Marijuana</th>
<th>C/S/O</th>
<th>Polydrug</th>
<th>CBT</th>
<th>CBT + psychosoc.</th>
<th>CBT + pharm.</th>
<th>Vs active</th>
<th>Vs passive</th>
<th>Vs no treatment</th>
<th>Vs no adjunct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed effects</td>
<td>0.067&lt;.05</td>
<td>0.513&lt;.05</td>
<td>0.126&lt;.05</td>
<td>0.116</td>
<td>0.165&lt;.05</td>
<td>0.329&lt;.05</td>
<td>0.208&lt;.05</td>
<td>0.129&lt;.05</td>
<td>0.116&lt;.05</td>
<td>0.848&lt;.05</td>
<td>0.089&lt;.05</td>
</tr>
<tr>
<td>95% CI</td>
<td>-0.002, 0.136</td>
<td>0.375, 0.651</td>
<td>0.011, 0.242</td>
<td>-0.007, 0.239</td>
<td>0.085, 0.245</td>
<td>0.238, 0.421</td>
<td>0.070, 0.346</td>
<td>0.041, 0.217</td>
<td>0.052, 0.180</td>
<td>0.692, 1.010</td>
<td>-0.066, 0.244</td>
</tr>
<tr>
<td>Range</td>
<td>-0.670, 1.209</td>
<td>0.225, 0.824</td>
<td>-0.845, 0.626</td>
<td>-0.442, 0.642</td>
<td>-0.644, 0.626</td>
<td>-0.239, 1.210</td>
<td>-0.451, 0.867</td>
<td>-0.644, 0.626</td>
<td>-0.451, 0.867</td>
<td>0.288, 1.210</td>
<td>-0.845, 0.523</td>
</tr>
<tr>
<td>N</td>
<td>23</td>
<td>6</td>
<td>13</td>
<td>11</td>
<td>21</td>
<td>19</td>
<td>13</td>
<td>17</td>
<td>32</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Q (df)</td>
<td>34.20 (22)*</td>
<td>10.53 (5)</td>
<td>40.39 (12) §</td>
<td>10.96 (10)</td>
<td>37.80 (20)*</td>
<td>64.23 (18) §</td>
<td>18.53 (12)</td>
<td>20.09 (16)</td>
<td>34.10 (31)</td>
<td>18.66 (5) §</td>
<td>35.21 (6) §</td>
</tr>
<tr>
<td>τ</td>
<td>35.67*</td>
<td>52.53</td>
<td>70.29</td>
<td>8.72</td>
<td>47.09</td>
<td>71.97</td>
<td>35.25</td>
<td>20.38</td>
<td>31.26</td>
<td>73.21</td>
<td>82.96</td>
</tr>
</tbody>
</table>

Notes: C/S/O = cocaine/stimulant/opiate; CBT = cognitive-behavioral treatment; psychosoc. = psychosocial; pharm. = pharmacological; CI = confidence interval. *The trimmed estimate with three outlying (Kelly et al., 2006; O’Malley et al., 1992; Sandahl et al., 2004) and two high weight (Anton et al., 2006; Project MATCH, 1997) trials removed was larger, significant, and homogeneous (g = 0.14, p <.05, Q >.05); the estimate with one high-weight study (MTRP, 2004) removed was 0.38 (p <.005, Q >.05); the trimmed estimate with two outlying trials (Carroll et al., 1991; Rawson et al., 2006) removed was slightly larger but remained heterogeneous (g = 0.19, p <.05). § the trimmed estimate with one outlying trial (Sandahl et al., 2004) removed was slightly higher (g = 0.18, p <.005, Q >.05) but remained heterogeneous; the trimmed estimate with two outlying studies (Kelly et al., 2001; MTRP, 2004) removed was lower and homogeneous (g = 0.20, p <.005, Q >.05); Anton et al. (2006) not included in analyses because it qualifies as both a psychosocial and pharmacological combined intervention; *positive effect comparisons include interpersonal psychotherapy (Carroll et al., 1991, 2004; Donovan and Ito, 1988), twelve-step facilitation (Brown et al., 2002; Maude-Griffin et al., 1998), motivational interviewing (Anton et al., 2005; Stephens et al., 2000), and contingency management (Budney et al., 2006); §the trimmed estimate with one outlying trial (Rawson et al., 2006) removed was larger, significant but remained heterogeneous (g = 0.22, p <.05, Q <.05). *p <.05; §p <.005.

Table 3. Study-level moderators of substance-use outcomes (n = 53)

<table>
<thead>
<tr>
<th>Model</th>
<th>β</th>
<th>b</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Client factors</td>
<td>Biological outcome</td>
<td>-.191</td>
<td>-.118</td>
<td>-.187</td>
</tr>
<tr>
<td>6- to 12-month outcome</td>
<td>-.170</td>
<td>-.102</td>
<td>-.163</td>
<td>.104</td>
</tr>
<tr>
<td>Mean age of participants</td>
<td>-.071</td>
<td>-.004</td>
<td>-.058</td>
<td>.560</td>
</tr>
<tr>
<td>Percentage female participants</td>
<td>.248</td>
<td>.005</td>
<td>2.66</td>
<td>.008</td>
</tr>
<tr>
<td>Co-occurring disorder inclusion</td>
<td>-.046</td>
<td>-.028</td>
<td>-.44</td>
<td>.659</td>
</tr>
<tr>
<td>Illicit drug outcome</td>
<td>.123</td>
<td>.073</td>
<td>0.98</td>
<td>.327</td>
</tr>
<tr>
<td>R² = .186</td>
<td>Qₑ (46) = 104.86²</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment factors</td>
<td>Biological outcome</td>
<td>-.129</td>
<td>-.079</td>
<td>-1.16</td>
</tr>
<tr>
<td>6- to 12-month follow-up outcome</td>
<td>-.228</td>
<td>-.134</td>
<td>-2.23</td>
<td>.026</td>
</tr>
<tr>
<td>Treatment delivery</td>
<td>.152</td>
<td>.096</td>
<td>1.61</td>
<td>.108</td>
</tr>
<tr>
<td>Treatment format</td>
<td>.017</td>
<td>.011</td>
<td>0.17</td>
<td>.866</td>
</tr>
<tr>
<td>Treatment length</td>
<td>-.312</td>
<td>-.008</td>
<td>-3.00</td>
<td>.003</td>
</tr>
<tr>
<td>R² = .193</td>
<td>Qₑ (47) = 103.52²</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final model–weighted least squares⁴</td>
<td>Biological outcome</td>
<td>-.083</td>
<td>-.050</td>
<td>-.078</td>
</tr>
<tr>
<td>6- to 12-month follow-up outcome</td>
<td>-.261</td>
<td>-.153</td>
<td>-2.57</td>
<td>.010</td>
</tr>
<tr>
<td>Percentage female participants</td>
<td>.200</td>
<td>.004</td>
<td>2.06</td>
<td>.039</td>
</tr>
<tr>
<td>Treatment length</td>
<td>-.252</td>
<td>-.006</td>
<td>-2.47</td>
<td>.014</td>
</tr>
<tr>
<td>R² = .205</td>
<td>Qₑ (4) = 26.25³</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qₑ (48) = 101.95³</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g = 0.144⁴</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final model–maximum likelihood</td>
<td>Biological outcome</td>
<td>-.048</td>
<td>-.037</td>
<td>-.304</td>
</tr>
<tr>
<td>6- to 12-month follow-up outcome</td>
<td>-.185</td>
<td>-.134</td>
<td>-1.44</td>
<td>.149</td>
</tr>
<tr>
<td>Percentage female participants</td>
<td>.226</td>
<td>.004</td>
<td>1.69</td>
<td>.089</td>
</tr>
<tr>
<td>Treatment length</td>
<td>-.259</td>
<td>-.007</td>
<td>-1.76</td>
<td>.077</td>
</tr>
<tr>
<td>R² = .169</td>
<td>Qₑ (4) = 10.90*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Qₑ (47) = 53.52</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>τ = .04</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g = 0.155³</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: Attribution rate nonsignificant as a methodological covariate. ¹Results held in the final weighted least squares model with the two high-weight trials (Project MATCH, 1997; Antón et al., 2006) removed.
* p <.05; § p <.005.
Specifically, study sample size (Gilbert et al., 2006; Kelly et al., 2000) and strength of comparison (Conrod et al., 2000) may be additional factors contributing to a positive association between proportion of female participants and CBT effect size.

This meta-analysis found no difference in effectiveness of CBT by format (group or individual), found little evidence for its value as an adjunctive treatment, and found support for a benefit of shorter duration interventions. Given absent differences by format, group CBT may be the most cost-effective option for clinical delivery. CBT as an adjunct to a psychosocial treatment may not yield improved outcomes beyond that treatment alone, but this finding may represent the minimal benefit of adding CBT to contingent reinforcement when the comparison is voucher incentives only (four of seven studies within this subgroup: Budney et al., 2006; Epstein et al., 2003; Rawson et al., 2002, 2006). The current review also suggests larger effect sizes with shorter duration CBT interventions. However, of the 10 studies with greater than 20 sessions, 7 compared CBT with at least a support group or treatment as usual (Kadden et al., 1989, 2001; McKay et al., 1997; Rawson et al., 2002, 2006; Rosenblum et al., 2005; Wetzel et al., 2004). The current research therefore supports a benefit from shorter CBT interventions, but whether this finding is also related to additional study characteristics, such as strength of comparison condition, requires further investigation.

A number of potential limitations are notable from the current review. Diagnostic tests did not suggest the presence of publication bias, but it is unknown whether inclusion of unpublished research would have substantively affected effect size magnitude. It is also unknown whether the decision to extract one outcome per study, rather than treating type of outcome measure as a modifying variable, would have resulted in different indices of CBT effect. CBT models may have been penalized by the minority of no-treatment comparisons or by the averaging of treatment arms given the impact of strength of comparison condition on effect magnitude. As demonstrated by Wampold (2001), direct evaluation of two psychosocial interventions rarely shows significant differences. Finally, potential collinearity among study characteristics such as gender and sample size or length of treatment and strength of comparison underscores the caution needed when interpreting study-level moderators in meta-analysis (see e.g., Lipsey, 2003; Wilson, 2000).

The current research demonstrates the overall effectiveness of CBT across adult alcohol- and other drug-use disorders. It may be particularly effective with marijuana-use disorders, with women, when combined with an additional psychosocial treatment, and when delivered in a brief format. This review also suggests that group CBT is as effective as CBT delivered as an individual treatment, and does not show that CBT is uniquely beneficial as aftercare or when delivered as an adjunctive treatment particularly in combination with contingency management. The noted findings provide provisional clinical guidelines and future directions for dissemination research.

Acknowledgments

This meta-analysis was originally conducted in partial fulfillment of the requirements for Dr. Molly Magill’s doctoral degree from Boston College Graduate School of Social Work. The authors thank Drs. Thomas O’Hare and Timothy Apodaca for their valuable guidance on this manuscript. They also thank Drs. Joseph Pedulla and Robert Dunigan for their committee participation.

References

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

AMERICAN PSYCHIATRIC ASSOCIATION. Diagnostic and Statistical Manual of Mental Disorders (DSM-IV), Washington, DC, 1994.


MAGILL AND RAY 527


