META-ANALYSIS OF COGNITIVE-BEHAVIORAL TREATMENTS FOR SOCIAL PHOBIA

STEVEN TAYLOR
Department of Psychiatry, University of British Columbia

Summary — A meta-analysis was conducted using 42 treatment-outcome trials for social phobia. Six conditions were compared: Waiting-list control, placebo, EXP (within-session exposure and homework exposure), CT (cognitive restructuring without exposure exercises), CT+EXP, and SST (social skills training). All interventions, including placebo, had larger effect sizes than that of the waiting-list control, and the interventions did not differ in drop-out proportions. Only CT+EXP yielded a significantly larger effect size than placebo. Effects of treatments tended to increase during the follow-up period. These results support the use of cognitive-behavioral treatments for social phobia, especially the use of CT+EXP. Copyright © 1996 Elsevier Science Ltd

Social phobia is characterized by marked and persistent fear of social or performance situations in which embarrassment may occur (American Psychiatric Association [APA], 1994). Examples include fear of eating in public, fear of public speaking, and fear of interacting with strangers. In clinical settings, the majority of Social phobics report fears of more than one type of social situation (APA, 1994). Social phobia tends to develop early in life, with a lifetime prevalence of 2–4%, and a low rate of naturally-occurring recovery (Davidson, Hughes, George & Blazer, 1994; Schneier, Johnson, Hornig, Liebowitz & Weissman, 1992). Compared to people with no psychiatric condition, social phobics have increased rates of suicidal ideation, impaired social support, greater use of medical facilities, and impaired occupational and school performance (Davidson et al., 1994; Schneier et al., 1992).

Cognitive-behavioral therapies are commonly used to treat social phobia. Such treatments include (1) prolonged exposure to social stimuli (EXP), conducted within the therapy session and as homework assignments; (2) cognitive therapy (CT), which attempts to restructure maladaptive beliefs about social situations and the opinions of others; (3) combined CT and EXP; and (4) social skills training (SST), which includes assertiveness training and training in other interpersonal skills such as making appropriate eye contact, and commencing and maintaining conversations. Note that although SST entails EXP, it involves more than exposure to feared situations. SST involves training the person in new interpersonal skills, such as methods for improving assertiveness (e.g. the “broken record” technique, using “I” statements, and using active listening; for descriptions of these and other SST procedures, see Lange & Jakubowski, 1976; Liberman, DeRisi & Mueser, 1989; Wolpe, 1990). Training in these methods is not a part of EXP. Accordingly, SST and EXP will be considered separately in this article.
There have been several narrative reviews of the efficacy of these treatments (e.g. Donohue, van Hasselt & Hersen, 1994; Heimberg & Juster, 1995). Although these reviews advance our understanding of the treatment of social phobia, they do not quantify the relationship between the various treatment strategies and treatment outcome. Accordingly, the aim of the present study was to conduct a meta-analysis of cognitive-behavioral therapies. Three questions were addressed: (1) Are these treatments effective relative to a waiting list control and placebo? (2) Are there any benefits of adding CT to EXP? (3) Are gains maintained at followup? (Comparisons with drug treatments of social phobia were not conducted due to insufficient trials.)

Method

Inclusion and Exclusion Criteria

Studies consisted of published or unpublished English-language articles located from searches of Psychological Abstracts, Medline, Current Contents, conference programs, recent issues of journals, secondary sources (e.g. citations in book chapters or journal articles), and by contacting researchers in the area. Articles were included if they met the following criteria:

1. Subjects were diagnosed as having social phobia according to DSM-III, DSM-III-R, or DSM-IV criteria.
2. Used 5 or more subjects per trial.
3. Used CT, EXP, SST, placebo, or waiting list control. Trials that combined treatments (e.g. SST+CT, EXP+placebo) were excluded, with the exception of trials combining CT and EXP, and trials combining SST and EXP (EXP is an inherent feature of SST).
4. CT+EXP was in the form of CT integrated into exposure assignments, as is used in the standard application of CT+EXP (Beck & Emery, 1985). Some trials presented patients with a series of sessions of CT followed by a series of trials of EXP. Such trials were omitted because they do not reflect clinical practice.
5. Provided sufficient information to compute effect sizes (supplemented, when necessary, by obtaining unpublished data from the authors).
6. Used broad rather than narrow outcome criteria. Broad criteria were used because the majority of social phobics fear more than one social situation (APA, 1994), and so reliance on a narrow outcome measure (e.g. a measure of fear of public speaking) may not accurately reflect the efficacy of treatment. Broad measures, assessing a range of social fears, provide a more accurate assessment of the overall severity of social phobia.

A total of 25 studies were located, yielding 43 treatment-outcome trials. The open trial by Fava, Grandi and Canestrari (1981) was omitted because its effect size (for EXP therapy) was an outlier (>3 SDs from the mean). Thus, 24 studies (indicated * in the reference section) yielded 42 trials, which were included in the meta-analysis. Three trials were uncontrolled; the remainder were reported in studies that compared two or more conditions (e.g. CT vs EXP vs waiting list control).

Table 1 shows the number of trials in each of the 6 conditions. A combination of within-session exposure and homework exposure was used in all EXP trials and in most (11/12) CT+EXP trials. The remaining trial used homework exposure without using exposure exercises during the sessions. There were two types of placebo trials; 4 trials used pill placebo and 2 trials used attention placebo. The two types of placebo trials produced similar mean effect sizes (pill placebo = 0.524, attention placebo = 0.416), and so they were combined.
Two studies used crossover designs, where subjects completed a waiting period then EXP (Mersch, 1995) or placebo followed by EXP (Taylor, Woody, Koch, McLean, Paterson & Anderson, in press). Four trials were extracted from these studies (i.e. a waiting list trial, placebo trial and two EXP trials). All other trials in this meta-analysis each used different samples. The four trials from Mersch and Taylor et al. were included in order to increase the number of trials in their respective conditions (Table 1). The pattern of results did not change when the two EXP trials were omitted from the analyses (post-treatment EXP effect size = 0.828 when the trials were excluded, and 0.817 when the trials were included).

Statistical Procedures

Effect sizes were computed according to Cohen’s (1988) d statistic. For each trial the magnitude of change from pre- to posttreatment was defined as \((M_{pre} - M_{post})/SD_{pooled}\), where \(SD_{pooled} = \sqrt{(SD_{pre}^2 + SD_{post}^2)/2}\). The magnitude of change from posttreatment to follow-up was defined by replacing \(M_{post}\) with \(M_{followup}\) and \(SD_{post}\) with \(SD_{followup}\). For the outcome measures used in the present study, positive effect sizes represent improvements in social phobia (i.e. reductions in symptom severity), whereas negative effect sizes indicate a worsening of social phobia.

The procedures described by Hunter and Schmidt (1990, pp. 512–513) were used to evaluate the possibility of inflated effect sizes due to the “file drawer” problem (Rosenthal, 1979). The latter may occur if there is a publication bias, such that studies obtaining significant findings (e.g. large effect sizes) are published, whereas findings obtaining null results (e.g. zero effects) are unpublished. To determine the likelihood of this bias, the fail-safe \(N\) is computed (Orwin, 1983), which is the number of unpublished trials obtaining zero effect sizes that are required to reduce an obtained mean effect size to a trivial level. If a large number of unpublished trials are required, then it is unlikely that obtained effect is biased by the file drawer problem. The number of unpublished (or unobtained) trials obtaining zero effect sizes is defined by \(k[(d_k/d_0) - 1]\), where \(k\) = the number of obtained trials, \(d_k\) = mean obtained effect size, and \(d_0\) = the trivial value to which the obtained effect would be reduced.

<table>
<thead>
<tr>
<th>Condition</th>
<th>% Dropouts</th>
<th>Effect size: Posttreatment</th>
<th>Effect size: 3-Month Follow-up*</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(M)</td>
<td>SD</td>
<td>(n)</td>
</tr>
<tr>
<td>Waiting list control</td>
<td>5.7</td>
<td>4.6</td>
<td>6</td>
</tr>
<tr>
<td>Placebo</td>
<td>14.7</td>
<td>7.5</td>
<td>6</td>
</tr>
<tr>
<td>EXP</td>
<td>16.4</td>
<td>7.4</td>
<td>6</td>
</tr>
<tr>
<td>CT</td>
<td>12.2</td>
<td>10.9</td>
<td>5</td>
</tr>
<tr>
<td>CT+EXP</td>
<td>18.0</td>
<td>11.0</td>
<td>12</td>
</tr>
<tr>
<td>SST</td>
<td>16.6</td>
<td>8.2</td>
<td>5</td>
</tr>
</tbody>
</table>

EXP = exposure therapy, CT = cognitive therapy, SST = social skills training, \(n\) = number of trials used in each analysis.

*Adjusted via linear interpolation so that all mean follow-ups are 3-months posttreatment.
Note that \( d_c \) cannot equal 0, because \( \frac{d_r}{d_c} \) would be undefined. Orwin (1983) suggested small effect size (i.e. 0.200) would qualify as a trivial value. However, such an effect size is considered non-trivial by some meta-analysts (Lipsey & Wilson, 1993). Accordingly, we will define a trivial effect size as 0.050.

Assessment

Previous meta-analyses have shown that interviewer-rated scales consistently yield larger effect sizes than self-report scales (Lambert, Hatch, Kingston & Edwards, 1986; Taylor, 1995). This may reflect the greater sensitivity of interviewer-rated scales. Alternatively, it may be an artifact reflecting the fact that observers may know whether they are conducting an assessment before or after the patient received an intervention, and so their ratings may be biased by expecting patients to have less severe symptoms at the end of the trial. Effect sizes are typically computed for each trial by averaging the effect sizes for each outcome measure (Hunter & Schmidt, 1990). Systematic bias in computing effect sizes can occur if some types of treatments are evaluated with interviewer-rated measures, while others are evaluated with self-report measures. For the trials included in the present study, only 2 used interviewer-rated scales (along with self-report measures). Accordingly, the effect sizes for interviewer-rated scales were not included in the present study.

Systematic biases in computing effect sizes also may occur if self-report measures differ in their sensitivities to treatment outcome. To illustrate, studies of one treatment (e.g. CT) may tend to use measure \( X \), while studies of another treatment (e.g. EXP) may tend to use measure \( Y \). If the sensitivity of \( X < Y \), then this will produce a biased estimate of the relative efficacy of CT and EXP. Analyses were conducted to compare the relative sensitivity of outcome measures, and to group them accordingly.

Six self-report outcome measures of global severity of social phobia were commonly used in the 42 trials. The Fear of Negative Evaluation scale (FNE: Watson & Friend, 1969) was included as one of these measures. Although some authors (e.g. Mattick & Peters, 1988) assumed the FNE is a cognitive measure, it is more appropriately regarded as a measure of social anxiety because many of its items pertain to the experience of distress or anxiety (e.g. “I become tense and jittery if I know someone is sizing me up”).

The relative sensitivity of the outcome measures was determined by conducting repeated measures \( t \)-tests between pairs of measures. Dependent measures were the effect sizes for each outcome measure. To illustrate, the relative sensitivity of the FNE and Social Phobia and Anxiety Scale (SPAI: Turner, Beidel, Dancu & Stanley, 1989) were compared by selecting trials that used these measures, and then comparing the mean effect sizes for the FNE and SPAI by a \( t \)-test. These tests were conducted on all trials except the waiting list control trials. To boost the sample size, 30 other trials were added, consisting of 11 trials of cognitive-behavioral treatment that did not meet inclusion/exclusion criteria (e.g. trials combining EXP with pill placebo) and 19 drug trials. Two groups of measures were identified, corresponding to measures yielding large effect sizes (group 1 measures) and those yielding smaller effect sizes (group 2 measures). Group 1 measures consisted of the social phobia subscale of the Fear Questionnaire (Marks & Mathews, 1979), the SPAI (Turner et al., 1989), and the Social Phobia Scale and the Social Interaction Anxiety Scale (Mattick, Peters & Clarke, 1989). Group 2 measures consisted of the FNE and the Social Anxiety and Distress Scale (Watson & Friend, 1969). Effect sizes for each trial were computed as the mean of effect sizes of group 1 measures, and mean of effect sizes of group
Treatment of Social Phobia

2 measures. Effect sizes for group 1 measures were available for 91% of the 42 trials included in the meta-analyses, and effect sizes for group 2 measures were available for only 71% of trials. Because the group 1 measures were more sensitive and used in more trials, the meta-analyses in the present study were based on effect sizes derived from these measures. Group 2 measures were not used because they were less sensitive, used in fewer trials, and largely redundant with the group 1 measures.

Results

Preliminary Analyses

The six conditions did not differ in the duration (weeks) of their trials, \( F(5, 36) = 2.01, p > 0.1 \), grand mean = 9.3 weeks. The cognitive-behavioral treatments (i.e. EXP, CT, CT+EXP, and SST) did not differ in terms of the total amount of treatment, as indicated by the number of weeks of treatment multiplied by the mean weekly duration of each therapy session, \( F(3, 26) < 1 \). For the cognitive-behavioral treatments, 25 trials treated patients in groups and 4 trials treated patients individually. Effect sizes between the two formats did not differ significantly, \( t(26) = 1.19, p > 0.1 \). Accordingly, trials from group and individual therapies were combined.

For the analyses of the maintenance of treatment gains from posttreatment to follow-up, most studies reported followups of 3 months, and so when studies reported more than one follow-up assessment (e.g. 3 and 6 months), the follow-up closest to 3 months was used in order to increase the homogeneity of follow-up durations. Follow-up data were generally unavailable for waiting list controls and placebo conditions. A one-way ANOVA was used to compare the follow-up durations of EXP, CT, CT+EXP and SST. There was a trend toward significance, \( F(3, 23) = 2.81, p < 0.07 \), and so follow-up duration was used as a covariate.

Comparison of Treatments

Attrition. Table 1 shows the mean rates of dropout during the trials. Attrition across the 6 conditions was compared using a one-way ANOVA, where % dropout per trial was the dependent variable and treatment condition (6 conditions) was the independent variable. The conditions did not differ significantly in dropout rates, \( F(5, 36) = 1.74, p > 0.1 \). However, as Table 1 shows, there was a trend for the waiting list control to have the lowest dropout rate. Patients may have been less likely to drop out of the waiting list because they typically knew they would receive treatment after the waiting period, and waiting required little effort compared to the requirements of the treatments.

Effect sizes at posttreatment. Table 1 shows effect sizes at posttreatment. A one-way ANOVA was conducted, using posttreatment effect size as the dependent variable and treatment condition (6 conditions) was the independent variable. The overall test was significant, \( F(5, 32) = 11.23, p < 0.001 \), and so the conditions were compared by Tukey comparisons. As suggested by Table 1, the effect size of the waiting list control was significantly smaller than effect sizes of EXP, CT, CT+EXP and SST (\( ps < 0.05 \)). CT+EXP had a significantly larger effect size than that of placebo (\( p < 0.05 \)). The remaining pair-wise comparisons were not significant (\( ps > 0.05 \)).
Stability of effect sizes from posttreatment to follow-up. The stability of effect sizes was assessed by a two-way ANCOVA. The dependent variable was effect size and the covariate was follow-up duration. The independent variables were time of assessment (posttreatment vs follow-up) and treatment condition (EXP, CT, CT+EXP or SST). The treatment-by-time interaction was nonsignificant, $F(3, 21) = 1.64, p > 0.1$, and the time main effect was significant, $F(1, 21) = 5.40, p < 0.05$. Table 1 shows the mean and SD of the follow-up effect size for each condition, apart from the placebo and waiting list conditions. (Follow-ups generally were not conducted for the latter conditions.) To equate the treatments on follow-up duration, linear regression was used (performed within each treatment condition) to estimate the effect size at a follow-up duration of 3 months. This duration was selected because it was used in most (68%) trials reporting follow-up data. The table shows that the mean effect sizes tended to increase from posttreatment to follow-up.

Fail-safe N. Recall that fail-safe N is defined as number of unpublished null trials (i.e. those obtaining zero effect sizes) required to reduce an obtained mean effect size to a trivial magnitude (0.050). For the posttreatment data, the necessary number of unpublished null trials was 123 (EXP), 58 (CT), 223 (CT+EXP) and 48 (SST). These results suggest that there would need to be a large number of unpublished null trials to reduce obtained mean effect sizes to trivial levels. It seems unlikely that there would be so many unpublished trials finding zero effects for these treatments, and so we conclude that the findings were unlikely to have been biased by the “file-drawer” problem.

Discussion

To return to the questions raised at the beginning of this article, the results indicate the following. First, EXP, CT, CT+EXP and SST were effective relative to waiting list control and did not differ in dropout proportions. Second, only CT+EXP yielded a statistically greater effect size than placebo. CT, EXP and SST had effect sizes that were not significantly different from that of placebo. The effect size of CT+EXP was not significantly different from that of EXP, although there was a trend for CT+EXP to yield the largest effects (Table 1). Thus, there appears to be some benefit to adding CT to EXP.

Third, effect sizes generally tended to increase from posttreatment to follow-up. In the present analysis, most follow-up durations were 3 months. Heimberg, Salzman, Holt and Blendell (1993) recently reported that treatment gains from CT+EXP are maintained at 5-year follow-up, and Turner, Beidel and Cooley-Quille (1995) reported that gains from SST were maintained at 2-year follow-up. Although these findings suggest that cognitive-behavioral therapies continue to exert their effects during the follow-up period, they need to be interpreted with caution. Only Turner et al. reported that subjects generally did not receive further treatment during the follow-up period. It may be that apparent gains in other trials were due to the effects of additional treatment.

The treatments did not differ in dropout proportions; each was associated with a considerable rate of dropout (12.2–18.0%), suggesting that the utility of treatments could be improved by developing means of keeping patients in treatment. One strategy, found useful in a study by Sherman and Anderson (1987), is to begin therapy by asking patients to generate reasons for completing treatment. This is thought to make it easier for patients to think of reasons for remaining in therapy when they are deciding whether or not to drop out.
Treatment of Social Phobia

There was insufficient data to determine the relationship between type of social phobia (specific vs generalized) and treatment outcome, or whether the presence of comorbid avoidant personality has implications for treatment efficacy. A recent study by Hope, Herbert and White (1995b) found that type of social phobia and presence of avoidant personality disorder were unrelated to the degree of treatment response.

An important issue for further investigation concerns the efficacy of cognitive-behavioral therapies compared to pharmacotherapies. In my search for suitable trials for the present study, I located 19 trials of pharmacotherapy, using 7 different medications. Five trials combined medication with exposure instructions, and so they would have to be analyzed separately from trials using medication without exposure. Some trials conducted posttreatment assessments when the patients were still on medication, whereas other studies conducted the assessment after medication had been withdrawn. Unfortunately, there were too few trials per condition to include the pharmacotherapies in the meta-analysis. With the inevitable growth of research into the treatment of social phobia, such analyses should be possible within the next few years.

Acknowledgements — This research was supported in part by grant #M74-90 from the British Columbia Health Research Foundation. The author would like to thank Cheryl Shea Gelernter, Richard G. Heimberg, Debra A. Hope, Agnes Scholing and Samuel M. Turner for providing unpublished material. Correspondence concerning this article should be addressed to Steven Taylor, Department of Psychiatry, 2255 Wesbrook Mall, University of British Columbia, Vancouver, B.C., Canada V6T 2A1.

References


