

Does Cognitive-Behavioral Therapy Cure Obsessive-Compulsive Disorder? A Meta-Analytic Evaluation of Clinical Significance

JONATHAN S. ABRAMOWITZ

The University of Memphis

Meta-analytic methods were employed to investigate the clinical significance of exposure therapy in the treatment of obsessive-compulsive disorder (OCD). Samples of patients treated in 16 outcome trials were compared with 9 normative samples to clarify how similar the symptoms of treated patients were to those without history of OCD. The Maudsley Obsessional-Compulsive Inventory (MOCI) was used as the basis for comparison. As expected, substantial improvements during the course of treatment were observed. Additionally, at posttest, the average patient was functioning at a level more similar to the general population than to individuals with untreated OCD. Follow-up scores indicated that clinically significant changes were stable for up to 5 months. Treated patients, however, did remain more symptomatic than members of the general population. The results are discussed in terms of the advantages of routinely including tests of clinical significance in outcome research.

With an estimated lifetime prevalence rate of 2.5% in the United States (Karno, Golding, Sorenson, & Burnam, 1988) and similarly high rates around the world (Weissman et al., 1994), obsessive-compulsive disorder (OCD) is known to be one of the most common behavior disorders (Reiger, Narrow, & Raye, 1990). This anxiety disorder is characterized by two types of symptoms that may appear together or separately. *Obsessions* are intrusive thoughts, ideas, or images that are experienced as upsetting and anxiety-provoking. *Compulsions* are ritualistic behaviors that are performed intentionally and function to reduce the anxiety associated with obsessions. The severity of symptoms often fluctuates over the course of the disorder and can range from an occasional upsetting thought that becomes "stuck" in one's head to overwhelming guilt, anxiety, and urges to ritualize that significantly interfere with most areas of life.

I wish to thank Leslie Robinson for her helpful suggestions and comments during the preparation of this paper. Also, thanks to Steven Taylor for providing some of the unpublished data.

Correspondence regarding this research may be addressed to Jonathan S. Abramowitz, who is currently at Allegheny University of the Health Sciences, Eastern Pennsylvania Psychiatric Institute, Department of Psychiatry, 3200 Henry Ave., Philadelphia, PA 19129; e-mail: abramowitz@auhs.edu

Although once considered to be highly treatment resistant, contemporary cognitive-behavioral and pharmacological interventions have been developed to control much of the wide range of symptoms observed in OCD. In particular, a set of cognitive-behavioral procedures, exposure and ritual prevention (EX/RP), is generally acknowledged as the treatment of choice for this disorder (Expert Consensus Panel for Obsessive-Compulsive Disorder, 1997). Exposure involves systematic and prolonged confrontation with feared stimuli until anxiety levels decrease. Ritual prevention entails abstinence or postponement of compulsive rituals. In most cases, these two procedures are used in tandem with the occasional addition of cognitive restructuring techniques. Treatment manuals (Kozak & Foa, 1997) as well as self-help books (e.g., Foa & Wilson, 1991) that detail the use of these procedures are now widely available.

Results from OCD treatment research conducted around the world indicate that EX/RP is (a) significantly more effective than control treatments, and (b) associated with statistically significant improvement from pre- to posttest (see Stanley & Turner, 1995, and Steketee, 1993, for narrative summaries of the literature). Meta-analytic reviews of this literature have reported the magnitude of treatment effects in terms of a standard measure of effect size (d ; Cohen, 1977) that can be compared with outcomes of other treatments (Abramowitz, 1996, 1997; van Balkom et al., 1994; Christensen, Hadzi-Pavlovic, Andrews, & Mattick, 1987). Although each has employed slightly different meta-analytic methods, results from these reviews indicate that, on average, OCD patients treated with EX/RP are over one standard deviation better off at posttest than they were at the start of therapy—a large treatment effect as defined by Cohen.¹

While the beneficial effects of EX/RP are among the most consistent findings within the psychotherapy outcome literature, their relevance to the practical concerns of patients and therapists may be somewhat limited by the way in which the findings have been reported. For example, many studies use significance tests to compare means between treated and nontreated groups. However, the impact of a treatment cannot be determined by such tests because they indicate only whether differences in group means are reliable and not merely due to chance. Because significance tests do not address the magnitude of treatment effects, how much better off patients are as a result of therapy is not assessed. Meta-analysis has partly addressed this problem with the use of effect sizes. Effect sizes are a measure of how much treated patients have improved relative to pretest or a control group. However, most effect sizes measure only the relative difference between means and do not reveal information concerning posttest symptom severity relative to healthy individuals. In other words, even with treatments that have large effect sizes (e.g., $d > 0.8$; Cohen, 1977), it is possible for treated patients to still display

¹ Cohen (1977) asserted that in the social sciences, effect sizes of $d = .3$, $.5$, and $.8$ corresponded to small, typical, and large effects respectively.

moderate or severe OCD symptoms. Thus, an important practical issue not thoroughly addressed in the OCD treatment literature concerns the likelihood of meaningful change or recovery from symptoms. The traditional statistical criteria often used to judge the effectiveness of treatment may have little relevance to this area. Clearly, then, one area within this realm that requires attention is the assessment of *clinical significance* (Jacobson, Follette, & Revenstorf, 1984), the extent to which treatment meets the standards of efficacy desired by consumers, clinicians, and researchers alike. Or, as more than one OCD patient has probably asked his or her therapist, "Will exposure and ritual prevention cure OCD?"

Authors of treatment studies have addressed the issue of functional change using different types of procedures. Most typically, the percent of patients who are symptom-free at posttest is reported. This method, however, infers that one must be totally symptom-free to be considered successfully treated. Such a criterion might be appropriate for disorders involving panic attacks, self-injurious, and bulimic behavior. However, considering that occasional obsessional thoughts or compulsive urges occur normally in individuals without OCD (Gibbs, 1996), it is possible for patients with OCD to achieve functional improvement without being symptom-free. In other studies, clinical significance is defined in terms of the percentage of change in symptoms. For example, Steketee, Foa, and Grayson (1982) classified treatment completers as "much improved," "improved," or "treatment failures" based on symptom reductions of 70% or more, 31% to 69%, and 30% or less, respectively. While these criterion represent more achievable goals, it is possible that patients with severe OCD might improve to 31% yet remain functionally similar to the pretest state. A more general difficulty with such criteria is that they are arbitrarily determined. Further, these improvement criteria consider only the patient's level of functioning relative to pretest, and they overlook characteristics of the functional population to whom treated patients are being implicitly compared.

An empirically derived approach to clinical significance has been proposed by Jacobson and colleagues (Jacobson et al., 1984; Jacobson & Truax, 1991). This approach is based on the idea that as a function of therapy, patients progress from the dysfunctional distribution of symptoms toward the functional population distribution. Jacobson et al. proposed three criteria for operationalizing this phenomenon: (a) does the level of functioning at posttest fall outside the range of the dysfunctional population? (b) does the level of posttest functioning fall within the range of the functional, nonclinical population? and (c) at posttest, is the patient more likely to be a member of the functional, as opposed to the dysfunctional, population? The first of these criteria is limited in that it does not address how closely treated patients are functioning relative to the nonclinical population. Additionally, a criticism of both the first and second criteria is that the range of populations was defined arbitrarily as being two standard deviations from the means. The third criterion, however, implies a comparison between treatment results and

normative data obtained from nonclinical samples. It can be expressed as a cutoff point beyond which scores further in the direction of improvement are statistically closer to the mean of the normative, as opposed to the dysfunctional population. If posttest scores fall within the normative population distribution, clinically significant improvement is said to have occurred. Because this criterion is not arbitrary and must be statistically derived using normative and treatment outcome data, it has been thought preferable to the other two (Jacobson & Truax). While some researchers have recently used similar criteria to assess the clinical significance of treatments for various behavior disorders such as unipolar depression (Nietzel, Russell, Hemmings, & Gretter, 1987; Robinson, Berman, & Neimeyer, 1990), generalized anxiety disorder (Durham & Allan, 1993), and agoraphobia (de Beurs, van Dyck, van Balkom, Lange, & Koele, 1994), the present research describes the first attempt to examine whether clinically significant changes occur for individuals receiving EX/RP treatment for OCD.

Before making comparisons between clinical and nonclinical populations, characteristics of the nonclinical population with respect to OCD symptoms deserve attention. Gibbs (1996) has reviewed the research on this topic and concluded that manageable obsessions and compulsions are somewhat frequent in the general population. She explained that although these symptoms closely resemble those in diagnosable OCD, they may be of shorter duration, less severity, and ego-syntonic in their subclinical form. Rachman and de Silva (1978) suggested that although the content of intrusive thoughts is similar in those with and without OCD, only people with OCD experience excessive distress about these thoughts. The importance of these findings is that there appears to be no natural dividing line between OCD as a clinical presentation and obsessions and compulsions in the general population. There is rather a continuum of severity characterized by frequency, intensity, and duration of symptoms.

Meta-analysis can be employed to address issues regarding the clinical significance of treatments for a given population in much the same way that it can be used to answer questions regarding the effects of treatment relative to other treatments or controls. The difference is that in assessing clinical significance, treatment data are compared with norms from nonclinical samples as opposed to outcome data from competing treatments (Kendall, 1984). An advantage of meta-analysis over the individual outcome study is that it yields a quantitative summary of treatment effects collected across many outcome trials. Of particular concern in the OCD treatment literature is that EX/RP procedures have been implemented using varied methods of delivery. For example, Abramowitz (1996) found that outcome may be related to whether exposure is supervised by a therapist or conducted by the patient on their own. Meta-analysis can increase the external validity of outcome results because data are aggregated from across slightly different implementations of the same general exposure treatment procedure (Matt & Navarro, 1997).

In the present study, the aggregated results of treatment trials are compared with normative data in order to address the following questions: First, at post-test, are OCD patients who have received EX/RP still distinguishable from individuals in the general, non-OCD population? Second, do treated OCD patients differ at follow-up assessment from the general population? Third, using the approach of Jacobson et al. (1984; Jacobson & Traux, 1991), can the effects of EX/RP on OCD symptoms be considered clinically significant?

Method

Preliminary Considerations

Because assessing clinical significance entails making comparisons between treated patients and individuals in the general population, a particular outcome measure must be chosen as the basis for comparison. Such an instrument must have been used widely in both treatment and in normative studies. That is, general population means and standard deviations must be available for this measure. As this review concerns the treatment of OCD, the Maudsley Obsessional-Compulsive Inventory (MOCI; Hodgson & Rachman, 1977), a 30-item self-report, true-false scale was chosen as the basis for comparison because it was the only psychometrically adequate measure to meet both of these requirements. The MOCI is a valid instrument as evidenced by its correlations with other measures of OCD (Freund, Steketee, & Foa, 1987). Additionally, it possesses good test-retest reliability ($\tau = .84$; Kraaijkamp, Emmelkamp, & van den Hout, 1986) and is a useful screening tool (Sternberger & Burns, 1990). It has also been used extensively in research with clinical, sub-clinical, and nonclinical populations. One limitation of this instrument is that it is mainly a measure of compulsions—items pertaining to obsessional symptoms are underrepresented. In addition, the MOCI is most sensitive to washing and checking rituals and is much less sensitive to cognitive, counting, and hoarding rituals. Scores of the MOCI range from 0 (*no OCD symptoms*) to 30 (*severe OCD symptoms*) and generally reflect the amount of time consumed with ritualizing (Taylor, 1995). Hodgson and Rachman, the authors of this instrument, reported a mean of 18.86 ($SD = 4.92$) for a sample of individuals with OCD.

Studies

Treatment outcome research. The treatment studies considered for this review were those included in an earlier meta-analysis of the exposure-based treatment of OCD that focused on the effects of variations in exposure procedures (for inclusion criteria, see Abramowitz, 1996). Whether outcome was clinically significant was not addressed in that article. Studies were only selected for the present review if the complete MOCI was administered and reported as a measure of outcome at pretest, posttest, and/or follow-up.

Ten studies that reported on 16 separate groups treated with EX/RP met this criterion. The year of publication ranged from 1983 to 1995. All treat-

TABLE 1
CHARACTERISTICS, MOCI SCORES, AND RELIABLE CHANGE INDEX (RCI)
FOR EACH EX/RP TREATMENT GROUP

Treatment study	No. N Sess.		Total MOCI score			RCI ^a	
			Pretest		Posttest		Follow-up
			<i>M</i>	(<i>SD</i>)	<i>M</i>		(<i>SD</i>)
Emmelkamp & de Lange (1983) ^b	6	13	16.2		8.0		2.23 ^c
Emmelkamp & de Lange (1983) ^b	6	13	16.5		8.0		2.33 ^c
Emmelkamp et al. (1988)	9	14	15.6 (4.0)	12.6 (5.4)	11.4 (4.7)		1.31 (1.92)
Emmelkamp et al. (1989) ^b	7	10	14.8 (5.3)	6.3 (4.6)	8.1 (4.4)		2.83 (2.24)
Emmelkamp et al. (1989) ^b	7	10	14.8 (5.3)	6.3 (4.6)	8.1 (4.4)		2.83 (2.24)
Emmelkamp et al. (1990) ^b	25	8	16.7 (5.5)	13.2 (6.4)	13.0 (6.2)		1.12 (1.11)
Emmelkamp et al. (1990) ^b	25	8	16.7 (5.5)	13.2 (6.4)	13.0 (6.2)		1.12 (1.11)
Mehta (1990)	15	24	17.5 (5.3)	10.7 (3.2)	12.4 (4.1)		2.10 (1.48)
Mehta (1990)	15	33	18.2 (6.3)	8.0 (2.7)	7.2 (2.9)		2.92 (3.08)
Foa et al. (1985)	9	15	20.7 (8.5)	14.5 (6.4)	9.7 (6.3)		1.29 (2.33)
Foa et al. (1985)	10	15	17.1 (6.7)	12.6 (4.7)	10.0 (5.7)		1.22 (1.91)
van den Hout et al. (1988)	43	20	15.5 (4.0)	9.2 (4.8)			2.80
Keijsers et al. (1994)	40	18	13.7 (5.3)	10.7 (6.3)			0.98
Emmelkamp & Beens (1991)	10	12	15.8 (6.7)	9.6 (6.9)	5.9 (2.5)		1.64 (2.57)
Emmelkamp & Beens (1991)	11	12	15.9 (4.6)	11.7 (6.1)	10.8 (4.1)		1.59 (2.04)
Woody et al. (1995)	33	16	13.8 (5.2)	9.2 (5.5)			1.44

Notes. ^a Follow-up RCI values in parentheses. ^b Average MOCI scores were used to estimate outcome for both treatment groups within these studies. ^c RCI was calculated using the standard deviation of the normative sample.

ments took place in outpatient settings and on an individual basis. While the therapist supervised exposure in 8 of the trials (50%), exposure was performed by patients as homework in 3 instances (19%). In the remaining 5 trials (31%), exposure was performed both in-session and for homework. The vast majority of treatments involved in vivo exposure to anxiety-evoking stimuli (13 trials, 81%). One trial (6%) used imaginal exposure and 2 (13%) combined in vivo with imaginal methods. In 9 of the 16 treatment conditions (56%), complete response prevention was part of the treatment; and in 4 (25%) instances, friends or family of the individual with OCD assisted with treatment. Table 1 presents the sample size, number of sessions, and MOCI data for each treatment group included in the analysis.² Additionally, the descriptive characteristics of these groups are summarized in Table 2.

² In three separate investigations (Emmelkamp, de Haan, & Hoogduin, 1990; Emmelkamp & de Lange, 1983; Emmelkamp et al., 1989), two variations of exposure procedures were compared with one another and no significant differences in outcome were detected. To maximize power, the authors combined the data and reported outcome as averaged across both treatment groups. In each of these cases, I used the average MOCI score to estimate outcome for both treatment groups in the particular study.

TABLE 2
SUMMARY OF CHARACTERISTICS OF THE 16 GROUPS TREATED BY EX/RP

Characteristic	<i>M</i>	Range
Number of treatment completers	16.94	6-43
Mean age of participants (years)	33.87	30-35
Percent of male patients	37.54	17-67
Pretreatment OCD symptom duration (years)	7.60	3-14.8
Length of treatment (weeks)	8.38	3-16
Length of follow-up (weeks)	20.00	9-44
Total number of sessions	14.31	8-24
Percent of patients who discontinued treatment	9.83	0-28

Note. Each mean is based on at least 12 of the 16 treatment groups.

A potential threat to the external validity of results gleaned from these treatment groups is that a number of groups were obtained from the same study or research team. For example, over half of the groups were obtained from research conducted by Paul M. G. Emmelkamp and associates. Because all treatment delivered in Emmelkamp's research likely possessed similar characteristics (e.g., similar recruiting methods, treatment protocols, supervision style, therapists), it raises the question of whether to treat results from the same study or research team as independent observations. If treatment results from groups that received therapy in the same laboratory were not independent, then subsequent statistical tests that assume the independence of data would be subject to bias. To address this issue empirically, a one-way analysis of variance was conducted in which the research group was the independent variable ($N = 7$) and the group's mean change on the MOCI (pretest-posttest) was the dependent variable ($N = 16$). This analysis revealed that the variability of change scores between research groups was not significantly larger than the variability of change scores within research groups, $F(5, 15) = 0.71$, $p = .60$. Thus, results from groups treated by the same research team were assumed to be independent observations.

Normative studies. A literature search similar to that for treatment outcome research was conducted for studies in which the MOCI was administered to samples drawn from the general population. This search resulted in 6 studies in which data for 9 separate normative samples were reported. Three studies were conducted in the United States, two in England, and one in Italy. Six of the 9 samples were comprised of university students. The typical normative sample consisted of 239 participants and was 33.12% male. The mean age of the participants was 26.1 years. Table 3 presents the MOCI scores and participant characteristics for each of the 9 normative samples included in the review.

Of the 9 normative samples described in Table 1, only one (Sanavio, 1988) screened out participants with OCD. Thus, each of the other normative samples, most of which were drawn from unscreened university student pop-

TABLE 3
CHARACTERISTICS AND MOCI SCORES FOR EACH NORMATIVE GROUP INCLUDED IN THE REVIEW

Study (location) and Sample	<i>N</i>	Age (<i>M</i>)	Percent Male	MOCI	
				<i>M</i>	<i>SD</i>
Arena et al., 1986 (U.S.) VA hospital staff	27	40.3	18.52	5.80	4.30
Dent & Salkovskis, 1986 (U.K.) University students	65	20.1	25	6.32	4.92
Medical students	142	18.9	50	7.26	4.41
Non-students	36	28.6	22	5.86	3.51
Frost et al., 1994 (U.S.) Undergraduate students	239			9.00	8.80
Graduate students	173	29.8	43.5	5.30	3.80
Reynolds & Salkovskis, 1991 (U.K.) Students	169	17-50	12.8	4.96	5.21
Sanavio, 1988 (Italy) Nonclinical participants	720			6.58	5.13
Sternberger & Burns, 1990 (U.S.) Undergraduate students	579	19.0	60	7.58	4.28

ulations, may have contained some participants who displayed OCD symptoms. Due to this difference in screening procedures, the 9 normative samples may actually contain data from two distinct populations: the general population, and a completely non-OCD population. Moreover, one might expect the general population to have higher mean MOCI scores than the sample known to be free of OCD. This is an important issue since the appearance of clinically significant improvement is largely influenced by characteristics of the normative data used in comparisons to treated patients (Nietzel et al., 1987). Changes in both the mean and variance of normative data can produce more or less strict criteria for clinical significance. To address this issue, a *z* test was performed to determine whether the mean MOCI score from the OCD-free sample in Sanavio's (1988) report ($M = 6.58$, $SD = 5.13$) differed significantly from scores of the other 8 unscreened samples ($M = 6.51$, $SD = 1.35$). This analysis revealed that the Sanavio sample was not significantly different from the mean of the other samples ($z = .03$), thus it was decided to use all 9 normative samples in the comparison to treated patients.

Results

Tests of Statistical Significance

Table 4 presents the mean MOCI score for the normative studies as well as for the pretest, posttest, and follow-up assessments of the 16 groups of OCD patients that received treatment. Means were calculated by weighting

TABLE 4
 MAUDSLEY OBSESSIVE-COMPULSIVE INVENTORY (MOCI) SCORES FOR THE NORMATIVE
 SAMPLE AND OCD PATIENTS TREATED WITH EX/RP

Group	N of groups	MOCI ^a	
		M	SD
Normative samples	9	6.52	1.3
OCD patients			
Pretest	16	16.95	1.6
Posttest	16	10.50	2.1
Follow-up	11	10.66	2.4

Note. Means and standard deviations are based on weighted least squares analyses in which MOCI scores were weighted by sample size.

^a Total MOCI score.

each treatment group's MOCI score by the sample size. Across treatment trials, the mean posttest MOCI score was significantly lower than the mean pretest score, paired $t(15) = 9.44, p < .01$. Thus, statistically reliable improvement in obsessive-compulsive symptoms took place during treatment. The follow-up mean was also significantly different from that at pretest, paired $t(10) = 7.21, p < .01$, indicating that progress was maintained after treatment had ended.

This improvement notwithstanding, patients who received EX/RP treatment reported significantly higher MOCI scores at posttest compared to participants in the normative studies, $t(23) = 4.06, p < .01$. A similar result appeared at follow-up assessment. The follow-up MOCI scores of treated patients were significantly larger than scores for the nonclinical sample, $t(18) = 3.86, p < .01$. This indicates that even though symptoms improved significantly during the course of treatment, OCD patients who had completed treatment could still be distinguished from healthy, nonclinical individuals.

Estimation of Clinical Significance

To examine whether improvement during treatment was clinically significant, patients' posttest scores on the MOCI were plotted on the distribution of MOCI scores from the normative samples. To generate an estimate of the general population's MOCI distribution, the standard deviation from the normative samples was pooled from the 9 samples using the following formula:

$$\sigma = \sqrt{\frac{(N_1 - 1)(SD_1^2) + (N_2 - 1)(SD_2^2) \dots}{(N_1 - 1) + (N_2 - 1) \dots}}, \quad (1)$$

where N_n is a sample's size and SD_n is its standard deviation. The general population's standard deviation was found to be $\sigma = 5.3$. The pretest MOCI standard deviations from the 16 treatment groups were also pooled using the

above formula to determine the distribution of scores for untreated OCD patients. For this clinical sample, the standard deviation was $\sigma = 5.4$. (The MOCI means for these groups appear in Table 4.)

Next, the pretest, posttest, and follow-up mean MOCI scores for the treated patients (from Table 4) were converted to z -scores so that they could be expressed relative to the general population's MOCI distribution. Z -scores were calculated using the formula:

$$z = \frac{M_{ix} - \mu}{\sigma}, \quad (2)$$

where M_{ix} is the treated group's pretest, posttest, or follow-up MOCI score, μ is the mean MOCI score of the general population ($\mu = 6.52$), and σ is the standard deviation of the general population distribution ($\sigma = 5.3$). In Figure 1, the mean pretest, posttest, and follow-up MOCI scores are expressed as z -scores in terms of standard deviations from the mean of the general population. As can be seen, patients at pretest reported significantly more obsessive-compulsive symptoms on the MOCI than did the nonclinical participants in the general population. After treatment by EX/RP, however, patients improved to within one standard deviation of the mean of the general population. At follow-up, MOCI scores were still within one standard deviation of the mean.

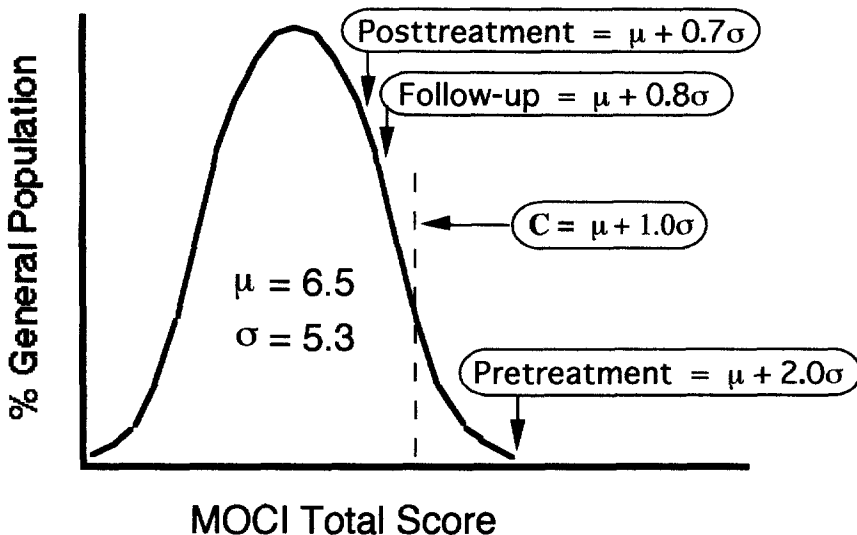


FIG. 1. Average pretest, posttest, and follow-up MOCI scores relative to the distribution of normative scores with reference to the cutoff for clinical significance ("C"). (The frequency distribution of normative scores is a hypothetical approximation.) Note. MOCI = Maudsley Obsessional-Compulsive Inventory.

In Figure 1, C (the dashed line) represents the cutoff below which MOCI scores are more likely to be drawn from the general population as opposed to the dysfunctional population's distribution of scores. C was calculated using the following formula, which was adapted from Jacobson et al. (1984; p. 341):

$$C = \frac{(M_{\text{ocd}})(SD_{\text{ocd}}) + (\sigma)(\mu)}{SD_{\text{ocd}} + \sigma}, \quad (3)$$

where M_{ocd} and SD_{ocd} are the pretest MOCI mean and pooled standard deviation from the 16 OCD samples, and μ and σ are the general population's MOCI mean and pooled standard deviation. The MOCI score that would represent the cutoff between the dysfunctional and general population distributions of MOCI scores was $C = 11.58$. In order to express this cutoff relative to the general population's MOCI distribution, it was converted to a z -score using formula 2 and substituting C for M_{ix} . The z -score for the cutoff was found to be 1.0. As is depicted in Figure 1, OCD symptoms improved 0.3 standard deviations past C at posttest and remained 0.2 standard deviations better at follow-up. This indicates that clinically significant improvement occurred for the average OCD patient as found in the outcome trials. Posttest and follow-up MOCI scores of those treated with EX/RP were somewhat more characteristic of the general population's MOCI distribution than of the distribution of scores of untreated OCD patients.

Reliable Change Index

The results presented thus far describe the overall functional level of patients after receiving therapy. However, the mere crossing of a cutoff point (C) does not necessarily indicate that the improvement made during treatment was statistically reliable in all cases. For example, apparent progress from pre- to posttest could be the result of a flawed outcome measure. Jacobson et al. (1984), Jacobson and Truax (1991), and Christensen and Mendoza (1986) have suggested the calculation of a reliable change index (RCI) to further explore this possibility. These researchers developed the RCI to examine whether reliable change occurred on a patient-by-patient basis in therapy outcome studies. However, it can be adapted for meta-analysis by calculating reliable change on a group-by-group basis, and by using group means as opposed to individual scores. RCI is calculated for each treatment group as follows:

$$RC = \frac{M_{\text{pre}} - M_{\text{post}}}{S_{\text{diff}}}, \quad (4)$$

where M_{pre} and M_{post} are a group's pre- and posttest MOCI scores and S_{diff} is the standard error of the difference between the two means. S_{diff} describes the spread of the distribution of change scores that would be expected if no actual change had occurred (Jacobson & Truax). It may be calculated as:

$$S_{\text{diff}} = \sqrt{2(S_E)^2}, \quad (5)$$

where S_E is the standard error of the measurement. S_E may be calculated using the test-retest reliability (r_{xx}) of the outcome measure (.84 for the MOCI), and the standard deviation of the group's pretest mean score (SD_{pre}) as follows:

$$S_E = SD_{\text{pre}}\sqrt{1 - r_{xx}}. \quad (6)$$

The far right column in Table 1 displays the RCI for each treatment group. With an alpha level of $p < .05$, an RCI of 1.96 or greater indicates that it is unlikely that the posttest score does not signify real change. As can be seen in Table 1, only 7 of the 16 (43.8%) posttest means and 6 of the 12 (50%) follow-up means met this criteria for reliable change.

Discussion

Although an entire body of literature affirms that EX/RP is successful in reducing obsessive-compulsive symptoms, the extent to which treated patients have functional improvements relative to the general population has not previously been explored. The goal of the present investigation was to determine whether improvement in the EX/RP treatment literature was clinically significant as defined by the nonarbitrary and empirically derived criteria constructed by Jacobson and colleagues (Jacobson et al., 1984; Jacobson & Truax, 1991). Meta-analytic procedures were employed to aggregate the results of treatment trials and compare symptoms in treated OCD patients with samples from the general population. Because the MOCI has been the dependent measure most often used to assess OCD symptoms in treatment studies, as well as in nonclinical samples, this instrument was chosen as the basis for comparison between the two populations.

From pre- to posttest, groups' mean MOCI scores progressed from the clinical distribution to the general population distribution. Further, after an average of 5 months following the end of treatment, the criteria for clinically significant improvement was still satisfied. Importantly, however, there is evidence that posttest MOCI scores only reflect reliable improvement in about half of the treatment groups. Thus, problems with this instrument may lead to invalid conclusions about treatment response. On the basis of these findings, it appears that while some patients who undergo cognitive-behavior treatment by EX/RP do experience functional, or clinically significant, improvements in their OCD symptoms, there are still a number of others who do not.

Does cognitive-behavioral therapy cure OCD? The present analyses indicate that the answer is no. Despite the demonstrated effects of EX/RP, treated patients on the average still reported higher levels of OCD symptoms than were found in the general population. However, given that subclinical obsessions and compulsions occur normally for a substantial proportion of the healthy population (Gibbs, 1996), it may be unrealistic to expect that treat-

ments such as EX/RP could result in complete alleviation of these symptoms. An additional consideration is that the MOCI does not directly assess the level of interference experienced as a result of these symptoms. Thus, post-exposure therapy, patients may still report having some obsessions or compulsions, yet experience minimal distress from them. This is in accord with Rachman and de Silva's (1978) suggestion that although clinical and subclinical obsessional symptoms are phenomenologically similar, people who experience excessive amounts of anxiety and functional distress from intrusive and unpleasant thoughts develop OCD, while those who are less concerned with such symptoms do not.

Without a group of studies that contained MOCI scores for the general population, a comparison between treated patients and normative samples could not have been made. Thus, this research highlights the importance of normative data for measures frequently used to assess treatment outcome. However, as is discussed above, the strictness of the criteria for judging clinical significance is directly related to the mean and standard deviation of the normative data used in comparisons to treated patients. Thus, it is important to acknowledge whether any important demographic differences across normative samples exist that might influence subsequent findings. In the present study, mean MOCI scores in the general population were not higher or more variable than scores obtained from samples that did not meet criteria for OCD. Thus, neither type of sample would have provided a more stringent criterion for clinical significance. However, in future research of this kind, where differences between types of normative samples could be identified, it would be advantageous to compare posttest data with different types of norms separately.

A limitation of this review is that the selection of studies was nonrandom and was based on the use of an individual outcome measure. This could lead to the over- or underrepresentation of work from particular investigators who consistently use, or do not use, the MOCI in their outcome research. The large number of studies by Emmelkamp that were reviewed are an artifact of this selection criterion. If the work of investigators who employed the MOCI is for some reason not representative of the OCD treatment outcome literature in general, then the present results would suffer similar generalizability difficulties.

A related shortcoming is that the evaluation of clinical significance was based only on one assessment instrument. Only about half of the posttest MOCI scores were found to be indicative of reliable change using the RCI. Taylor (1995) has pointed out that although the MOCI has been used in many treatment studies, it is primarily a symptom checklist and does not fully measure some of the important peripheral aspects of OCD (e.g., interference and resistance to compulsions). Additionally, because the MOCI is mainly comprised of items that assess washing and checking rituals, patients with other, equally incapacitating symptoms (e.g., hoarding and mental rituals, or severe obsessional thoughts), may obtain lower scores (Taylor). Although many

researchers use multiple dependent measures in OCD treatment studies (e.g., Yale-Brown Obsessive-Compulsive Inventory, Padua Inventory, and ratings of fear severity), the MOCI has been the most consistently employed outcome instrument in this literature. The continued development and consistent use of measures with better psychometric properties might allow for more reliable and valid assessments of clinically significant improvement in OCD patients in the future.

Important issues when considering the results of this review with regard to general clinical practice should also be noted. In many treatment studies, participants are selected carefully on the basis of whether or not they satisfy specific inclusion criteria, such as meeting *DSM* criteria for OCD and having no additional diagnoses. Thus, the conclusions of such studies are often based on the "perfect" patient who may not be representative of the OCD population at large. That is, many people who seek treatment for OCD suffer comorbid conditions such as depression, bipolar illness, psychotic features, personality disorders, and additional anxiety disorders (Steketee, 1993). Because the presence of comorbid psychopathology has been thought to hinder treatment by exposure procedures (Steketee), studies based on participants with diagnoses of only OCD may overestimate the effectiveness of treatment. Similarly, group means based only on treatment completers may overestimate the aggregated effects of treatment since they do not include scores from those who discontinued therapy due to lack of response. In this review, close to 10% of the participants in the aggregated sample discontinued treatment during the trial. Thus, this source of bias should be considered.

Overall, the results of this meta-analysis are supportive of the use of EX/RP for reducing obsessive-compulsive symptoms, particularly washing and checking rituals. It appears that treatment gains may be clinically significant for many who suffer such symptoms in the absence of additional serious psychopathology. That is, after treatment, many research patients function at a level that is more similar to the general population than to the OCD population. EX/RP, however, does not appear to *cure* OCD in that, at posttest, patients remain distinguishable from the average member of the population at large as measured by the MOCI. Further, there probably remain patients with OCD who do not respond well to EX/RP. This implies that despite general success, there is still room for improvement in the cognitive-behavioral treatment of this disorder. Future studies that are based on more heterogeneous OCD patient samples will further address the question of what predicts successful treatment. Especially as the health care landscape changes rapidly, there is also a need to make research results palatable for those with a more limited understanding of clinical science. The advantages of undertaking therapy can be better clarified by studies that report patients' posttreatment levels of functioning relative to nonclinical peers. Not only will such data help to further investigate the usefulness of therapies, but we will be working to disseminate worthy empirical findings and supporting the union of research and clinical practice.

References

Note. References marked with a single asterisk indicate treatment studies included in the meta-analysis. References marked with a double asterisk indicate sources for MOCI normative data.

- Abramowitz, J. S. (1996). Variants of exposure and response prevention in the treatment of obsessive-compulsive disorder: A meta-analysis. *Behavior Therapy, 27*, 583–600.
- Abramowitz, J. S. (1997). Effectiveness of psychological and pharmacological treatments for obsessive-compulsive disorder. A quantitative review. *Journal of Consulting and Clinical Psychology, 65*, 44–52.
- **Arena, J., Blanchard, E., Andrasik, F., & Appelbaum, K. (1986). Obsessions and compulsions in three kinds of headache sufferers: Analysis of the Maudsley Questionnaire. *Behaviour Research and Therapy, 24*, 127–132.
- Christensen, H., Hadzi-Pavlovic, D., Andrews, G., & Mattick, R. (1987). Behavior therapy and tricyclic medication in the treatment of obsessive-compulsive disorder: A quantitative review. *Journal of Consulting and Clinical Psychology, 55*, 701–711.
- Christensen, L., & Mendoza, J. L. (1986). A method of assessing change in a single subject: An alteration of the RC index. *Behavior Therapy, 17*, 305–308.
- Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (Rev. ed.). New York: Academic Press.
- de Beurs, E., van Dyck, R., van Balkom, A. J. L. M., Lange, A., & Koele, P. (1994). Assessing the clinical significance of outcome in agoraphobia research: A comparison of two approaches. *Behavior Therapy, 25*, 147–158.
- **Dent, H., & Salkovskis, P. (1986). Clinical measures of depression, anxiety, and obsessionality in non-clinical populations. *Behaviour Research and Therapy, 24*, 689–691.
- Durham, R. C., & Allen, T. (1993). Psychological treatment of generalized anxiety disorder: A review of clinical significance of results in outcome studies since 1980. *British Journal of Psychiatry, 163*, 19–26.
- *Emmelkamp, P. M. G., & Beens, G. (1991). Cognitive therapy with obsessive-compulsive disorder: A comparative evaluation. *Behaviour Research and Therapy, 29*, 293–300.
- *Emmelkamp, P. M. G., de Haan, E., & Hoogduin, C. A. L. (1990). Marital adjustment and obsessive-compulsive disorder. *British Journal of Psychiatry, 156*, 55–60.
- *Emmelkamp, P. M. G., & de Lange, I. (1983). Spouse involvement in the treatment of obsessive-compulsive patients. *Behavior Research and Therapy, 21*, 341–346.
- *Emmelkamp, P. M. G., van Linden van den Heuvell, C., Ruphan, M., & Sanderman, R. (1989). Home based treatment of obsessive-compulsive patients: Intersession interval and therapist involvement. *Behavior Research and Therapy, 27*, 89–93.
- *Emmelkamp, P. M. G., Visser, S., & Hoekstra, R. J. (1988). Cognitive therapy vs. exposure in vivo in the treatment of obsessive-compulsives. *Cognitive Therapy and Research, 12*, 103–114.
- Expert Consensus Panel for Obsessive-Compulsive Disorder. (1997). Treatment of obsessive-compulsive disorder. *Journal of Clinical Psychiatry, 58*(Suppl. 4), 3–28.
- *Foa, E. B., Steketee, G., & Grayson, J. B. (1985). Imaginal and in-vivo exposure: A comparison with obsessive-compulsive checkers. *Behavior Therapy, 16*, 292–302.
- Foa, E. B., & Wilson, R. (1991). *Stop obsessing: How to overcome your obsessions and compulsions*. New York: Bantam.
- Freund, B., Steketee, G. S., & Foa, E. B. (1987). Compulsive activity checklist (CAC): Psychometric analysis with obsessive-compulsive disorder. *Behavioral Assessment, 9*, 67–79.
- **Frost, R., Steketee, G., Cohn, L., & Griess, K. (1994). Personality traits in subclinical and non-obsessive-compulsive volunteers and their parents. *Behaviour Research and Therapy, 32*, 47–56.

- Gibbs, N. A. (1996). Nonclinical populations in research on obsessive-compulsive disorder: A critical review. *Clinical Psychology Review, 16*, 729-773.
- Hodgson, R. J., & Rachman, S. (1977). Obsessional-compulsive complaints. *Behaviour Research and Therapy, 15*, 389-395.
- Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15*, 336-352.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology, 59*, 12-19.
- Karno, M., Golding, J., Sorenson, S., & Burnam, M. A. (1988). The epidemiology of obsessive-compulsive disorder in five U.S. communities. *Archives of General Psychiatry, 45*, 1094-1099.
- *Keijsers, G. P. J., Hoogduin, C. A. L., & Schaap, C. P. D. R. (1994). Predictors of treatment outcome in the behavioral treatment of obsessive-compulsive disorder. *British Journal of Psychiatry, 165*, 781-786.
- Kozak, M. J., & Foa, E. B. (1997). *Mastery of obsessive-compulsive disorder: A cognitive-behavioral approach*. San Antonio, TX: The Psychological Corp.
- Kraaijkamp, H. J. M., Emmelkamp, P. M. G., & van den Hout, M. A. (1986). *The Maudsley obsessional-compulsive inventory: Reliability and validity*. Unpublished manuscript.
- Kendall, P. C. (1984). Behavioral assessment and methodology. In G. T. Wilson, C. M. Franks, K. D. Brownell, & P. C. Kendall (Eds.), *Annual review of behavior therapy* (Vol. 9, pp. 47-86). New York: Guilford Press.
- Matt, G. E., & Navarro, A. M. (1997). What meta-analyses have and have not taught us about psychotherapy effects: A review and future directions. *Clinical Psychology Review, 17*, 1-32.
- *Mehta, M. (1990). A comparative study of family-based and patient-based behavioural management in obsessive-compulsive disorder. *British Journal of Psychiatry, 157*, 133-135.
- Nietzel, M. T., Russell, R. L., Hemmings, K. A., & Gretter, M. L. (1987). Clinical significance of psychotherapy for unipolar depression: A meta-analytic approach to social comparison. *Journal of Consulting and Clinical Psychology, 55*, 156-161.
- Rachman, S. J., & de Silva, P. (1978). Abnormal and normal obsessions. *Behaviour Research and Therapy, 16*, 233-248.
- Reiger, D., Narrow, W. E., & Raye, D. S. (1990). The epidemiology of anxiety disorders: The epidemiologic catchment area experience. *Journal of Psychiatric Research, 24*(Suppl. 2), 3-14.
- **Reynolds, M., & Salkovskis, P. (1991). The relationship among guilt, dysphoria, anxiety and obsessions in a normal population—an attempted replication. *Behaviour Research and Therapy, 29*, 259-265.
- 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.
- Roper, G., Rachman, S., & Hodgson, R. (1973). An experiment on obsessional checking. *Behaviour Research and Therapy, 11*, 271-277.
- **Sanavio, E. (1988). Obsessions and compulsions: The Padua Inventory. *Behaviour Research and Therapy, 26*, 169-177.
- Stanley, M. A., & Turner, S. M. (1995). Current status of pharmacological and behavioral treatment of obsessive-compulsive disorder. *Behaviour Therapy, 26*, 163-186.
- Steketee, G. S. (1993). *Treatment of obsessive-compulsive disorder*. New York: Guilford Press.
- Steketee, G. S., Foa, E. B., & Grayson, J. (1982). Recent advances in the behavioral treatment of obsessive-compulsives. *Archives of General Psychiatry, 39*, 1365-1371.

- **Sternberger, L., & Burns, G. (1990). Compulsive Activity Checklist and the Maudsley Obsessional-Compulsive Inventory: Psychometric properties of two measures of obsessive-compulsive disorder. *Behavior Therapy, 21*, 117-127.
- Taylor, S. (1995). Assessment of obsessions and compulsions: Reliability, validity, and sensitivity to treatment effects. *Clinical Psychology Review, 15*, 261-296.
- van Balkom, A. J. L. M., van Oppen, P., Vermeulen, A. W. A., van Dyck, R., Nauta, M. C. E., & Vorst, H. C. M. (1994). A meta-analysis on the treatment of obsessive-compulsive disorder: A comparison of antidepressants, behavior, and cognitive therapy. *Clinical Psychology Review, 14*, 359-381.
- *van den Hout, M., Emmelkamp, P. M. G., Kraaykamp, H., & Grietz, E. (1988). Behavioral treatment of obsessive-compulsives: Inpatient vs. outpatient. *Behaviour Research and Therapy, 26*, 331-332.
- Weissman, M. M., Bland, R. C., Canino, G. J., Greenwald, S., Hwu, H., Lee, C. K., Newman, S. C., Oakley-Browne, M. A., Rubio-Stipec, M., Wickramarante, P. J., Wittchen, H. U., & Yeh, E. K. (1994). The cross national epidemiology of obsessive-compulsive disorder. *Journal of Clinical Psychiatry, 55*(Suppl. 3), 5-10.
- *Woody, S., Steketee, G., & Chambless, D. (1995). Reliability and validity of the Yale-Brown Obsessive-Compulsive Scale. *Behaviour Research and Therapy, 33*, 597-605.

RECEIVED: April 16, 1997

ACCEPTED: December 17, 1997