Defining Response in Clinical Trials for Obsessive-Compulsive Disorder: A Signal Detection Analysis of the Yale-Brown Obsessive Compulsive Scale

David F. Tolin, Ph.D.; Jonathan S. Abramowitz, Ph.D.; and Gretchen J. Diefenbach, Ph.D.

Objective: Many studies of the treatment of obsessive-compulsive disorder (OCD) have used percent reduction cutoffs on the Yale-Brown Obsessive Compulsive Scale (YBOCS) to classify patients as treatment responders. However, reduction criteria have varied from 20% to 50%, with studies of cognitive-behavioral therapy (CBT) using a more stringent criterion than studies of pharmacotherapy. The aim of this retrospective investigation was to determine optimal YBOCS reduction criteria for classifying patients as responders.

Method: Data from 87 adult clinic and research outpatients meeting DSM-IV-TR criteria for OCD according to structured interview were examined, comparing the percent YBOCS reduction from pretreatment to posttreatment with 2 "gold standard" criteria from the Clinical Global Impressions (CGI) scale: much or very much improved and mild illness or better. Signal detection analyses were used to determine the sensitivity, specificity, predictive value of a positive test, predictive value of a negative test, and efficiency of various YBOCS reduction cutoffs.

Results: A YBOCS reduction cutoff of 30% was optimal for predicting improvement on the CGI. The 20% cutoff used by many pharmacologic studies resulted in a high number of false positives, whereas the 50% cutoff used by most CBT studies resulted in a high number of false negatives. For predicting mild illness or better at posttreatment, a YBOCS reduction cutoff of 40% to 50% was optimal.

Conclusions: A YBOCS reduction criterion of 30% appears to be optimal for determining clinical improvement, whereas a 40% to 50% reduction criterion is appropriate for predicting mild illness at posttreatment. Future studies should employ a standard definition of treatment response in order to facilitate cross-study comparisons.

(J Clin Psychiatry 2005;66:1549–1557)

Received Dec. 20, 2004; accepted May 2, 2005; *From the Anxiety Disorders Center, The Institute of Living, Hartford, Conn. (Drs. Tolin and Diefenbach); the Department of Psychiatry, University of Connecticut School of Medicine, Farmington (Dr. Tolin); and the Department of Psychiatry and Psychology, Mayo Clinic, Rochester, Minn. (Dr. Abramowitz).*

This research was supported by grants from Hartford Hospital (Drs. Tolin and Diefenbach), the Donaghue Medical Research Foundation (Dr. Tolin), the Adam Zielinski Memorial Fund (Dr. Tolin), and the Obsessive-Compulsive Foundation (Dr. Abramowitz).

Drs. Tolin, Abramowitz, and Diefenbach report no additional financial or other affiliations relevant to the subject of this article.

Corresponding author and reprints: David F. Tolin, Ph.D., Anxiety Disorders Center, The Institute of Living, 200 Retreat Ave., Hartford, CT 06106 (e-mail: dtolin@harthosp.org).

The first-line treatments of choice for obsessive-compulsive disorder (OCD) are serotonin reuptake inhibitor (SRI) antidepressants and cognitive-behavioral therapy (CBT).[1] Numerous well-controlled studies attest to the efficacy of each of these forms of treatment.[2–13] Across studies of OCD, the widely accepted "gold standard" measure is the Yale-Brown Obsessive Compulsive Scale (YBOCS).[14,15]

Often, researchers report the efficacy of treatment in terms of the number of patients who are considered "treatment responders" and compare the proportion of responders in active versus comparison treatment groups. In many cases, patients are classified as responders when their scores on the YBOCS decrease by a certain percentage. For example, the Clomipramine Collaborative Study Group[13] designated patients as treatment responders when their YBOCS scores decreased by 35% or more from pretreatment.

One potential problem with the "treatment responder" designation is that different studies have used widely varying criteria in the determination of treatment response. This problem is illustrated in Table 1, which shows active treatment conditions in studies investigating either CBT or medications that used YBOCS percent reduction cutoffs to identify patients as treatment responders. Note that some studies have multiple entries in this table, indicating that these studies included multiple treatments (e.g., both a CBT arm and a medication arm) or examined different YBOCS percent reduction cutoffs. As

Table 1. Studies Using Percent Reduction Cutoffs on the YBOCS to Determine Treatment Response, the Percentage Used to Determine Response, the Kind of Treatment Employed, and the Percentage of Patients Identified as Responders

Percent YBOCS Reduction Used to Determine Response	Percentage of Responders (posttreatment)	Reference
CBT trials		
≥ 25	77	Cottraux et al,[16] 2001
≥ 25	70	Cottraux et al,[16] 2001
≥ 35	84	Warren and Thomas,[17] 2001
≥ 50	47	Cottraux et al,[16] 2001
≥ 50	47	Cottraux et al,[16] 2001
≥ 50	67	Simpson et al,[18] 1999
≥ 50	61	Kozak et al,[8] 2000
Medication trials		
≥ 20	50	Pigott et al,[7] 1990
≥ 20	20	Pigott et al,[7] 1990
≥ 20	56	Pato et al,[6] 1991
≥ 20	67	Pato et al,[6] 1991
≥ 25	69	Kampman et al,[19] 2002
≥ 25	40	Weiss et al,[20] 1999
≥ 25	48	Montgomery et al,[4] 1993
≥ 30	30	Weiss et al,[20] 1999
≥ 35	35	Tollefson et al,[3] 1994
≥ 35	84	Clomipramine Collaborative Study Group,[13] 1991
≥ 35	86	DeVeaugh-Geiss et al,[21] 1990
≥ 35	65	McDougle et al,[22] 1994
≥ 50	39	Kozak et al,[8] 2000

Abbreviations: CBT = cognitive-behavioral therapy, YBOCS = Yale-Brown Obsessive Compulsive Scale.

can be seen in Table 1, one study might call patients whose YBOCS scores decreased by 20% or more "treatment responders," whereas another study might reserve this designation for patients whose YBOCS scores decreased by at least 50%. Not surprisingly, then, it is difficult to compare response rates across studies or across types of treatment. The wide range of cutoff scores for defining treatment response highlights the need to clarify the requirements for determining that a patient has responded to treatment. The purpose of the present study was to examine, using signal detection analyses,[23] the adequacy of various YBOCS percent reduction cutoffs for defining treatment response.

We selected 2 gold standard measures of treatment response based on Clinical Global Impressions (CGI) scale[24] ratings. The CGI is a clinician-rated scale that yields 2 scores: the severity of illness and the degree of clinical improvement from baseline. There has been increased interest in the issue of symptom improvement versus remission in OCD treatment.[25,26] Improvement implies that the patient's symptoms have decreased by a meaningful amount from baseline, although the patient may remain quite symptomatic (e.g., symptom reduction from the severe range to the moderate range). Remission, by contrast, implies that the person is no longer symptomatic (pragmatically, most authors consider this criterion to be met if symptoms are no more than mild). CGI scores can be used to address both of these outcome criteria. Thus, if "treatment response" is interpreted to mean that patients have improved to a meaningful extent, the CGI-Improvement (CGI-I) rating would be used. Conversely, if "treatment response" is interpreted to mean that patients are in or near remission, the CGI-Severity of Illness (CGI-S) rating would be used. We examined optimal YBOCS percent reduction cutoffs for predicting each of these clinician ratings. In addition, we provide descriptive statistics to show average posttreatment functioning of patients whose YBOCS scores decreased by various amounts.

METHOD

Participants

We retrospectively analyzed treatment data from 87 adult outpatients with a primary DSM-IV-TR diagnosis of OCD who had participated in treatment outcome studies or had received open clinic treatment. Fifty-four patients (62.1%) were seen at the Institute of Living in Hartford, Conn.; 33 (37.9%) were seen at the Mayo Clinic in Rochester, Minn. Seventy-two patients (82.8%) received therapist-directed CBT treatment. Thirteen patients (14.9%) received a self-administered version of CBT, and 2 (2.3%) received self-administered CBT with brief therapist support.

Forty-five patients (51.7%) were men, and 42 (48.3%) were women. Race was not available for 2 patients; of those remaining, 83 (95.4%) were white. Mean (SD) age was 37.46 (11.70) years. Twenty-five patients (28.7%) also met DSM-IV-TR[27] criteria for major depressive disorder.

Materials

All interviews were conducted by Ph.D.-level clinical psychologists or postdoctoral fellows with experience in the assessment of OCD. Diagnostic status and comorbidity were assessed using the Anxiety Disorders Interview Schedule for DSM-IV (ADIS-IV; N = 54)[28] or the Mini-International Neuropsychiatric Interview (N = 33).[29] OCD symptom severity was assessed using the YBOCS.[14,15] Raters at both sites were trained in the YBOCS by rescoring audiotaped and videotaped interviews and resolving scoring discrepancies. Interrater reliability for the YBOCS at the Hartford site is excellent (r = 0.97). Global severity of illness and posttreatment improvement were assessed using the CGI,[24] an overall rating based on all available clinical information (rather than simply an arithmetic derivation of the YBOCS). Interrater reliability for the CGI at the Hartford site is good (r = 0.81).

In addition to the clinical interviews, patients completed a number of self-report measures. Severity of depressive symptoms was assessed using the Beck De-

pression Inventory-II (BDI-II).[30–32] General anxiety was assessed using the State-Trait Anxiety Inventory-Trait (STAI-T) version.[33] Finally, patients reported their degree of functional impairment on the Sheehan Disability Scale (SDS).[34]

Procedure

An independent evaluator not otherwise involved with the patients' treatment administered all clinical interviews at pretreatment and posttreatment. Prior to the interview, the patient completed all self-report measures at home.

All patients received variants of CBT that incorporated exposure and response prevention. Seventy-two patients (82.8%) received standard CBT (delivered in either individual or group format) according to published manuals.[35,36] Fifteen patients (17.2%) received a self-directed version of CBT[37] as part of a clinical trial. In addition, 54 patients (62.1%) were taking SRI medications at intake.

Summary of Analyses

To examine the relationship between various YBOCS percent reductions and dichotomous gold standard outcome measures, we used receiver operating curve (ROC) analyses.[23] ROC analyses, derived from signal detection theory, allow for an evaluation of prespecified cutoff scores in terms of sensitivity (the probability that patients meeting the gold standard criteria will exceed the test cutoff), specificity (the probability that patients not meeting the gold standard criteria will not exceed the test cutoff), predictive value of a positive test (the probability that patients exceeding the test cutoff will meet the gold standard criteria), predictive value of a negative test (the probability that patients not exceeding the test cutoff will not meet the gold standard criteria), and efficiency (the probability that the test and the gold standard will agree). To these analyses, Kraemer[38] proposed utilizing weighted kappa coefficients that adjust for base rates in the sample and thus correct for chance agreement between the test and the gold standard. Kraemer's ROC statistics include quality of sensitivity (κ [1.0]), quality of specificity (κ [0.0]), and quality of efficiency (κ [0.5]). For each of these coefficients, a value of 0.00 indicates chance agreement between the test and the gold standard, and 1.00 indicates perfect agreement.

Regardless of the analyses used, there is a tradeoff between sensitivity and specificity: optimally sensitive cutoff scores are most useful for screening purposes, due to the low frequency of false negatives, whereas optimally specific cutoffs are most useful for making definitive diagnoses due to the low frequency of false positives.[38] For the purposes of the present study, we assigned the greatest value to optimally efficient cutoffs, because they represent the greatest clinical utility (i.e., the probability that a given YBOCS cutoff will actually correspond to a gold standard rating of improvement).

RESULTS

Identification of Average YBOCS Reduction Criteria in Previous Studies

Among the 20 patient groups depicted in Table 1, the mean (SD) YBOCS cutoff required to label a patient as a "responder" was 33.0% (11.4%), with criteria ranging from 20% to 50%. Using an independent samples t test, we found that the 7 CBT samples required a significantly greater YBOCS reduction for responder status than did the 13 medication samples ($t = 2.51$, $df = 18$, $p = .02$).

On average, authors of CBT studies labeled patients responders when their YBOCS scores decreased by 40.7% (12.1%); in contrast, authors of medication studies labeled patients responders when their YBOCS scores decreased by 28.8% (8.9%). The modal requirement for CBT studies was a 50% or greater YBOCS reduction. Medication studies showed a bimodal requirement of either a 20% or greater YBOCS reduction or a 40% or greater YBOCS reduction. The percent reduction cutoff did not correlate significantly with the percentage of participants judged to be treatment responders ($r = 0.01$; $p = .97$). However, the percent reduction cutoff did correlate positively and significantly with the year of the study's publication ($r = 0.49$; $p < .05$).

Descriptive Information About the Present Sample

The mean (SD) YBOCS score at pretreatment was 24.21 (5.19), in the severe range. At posttreatment, the mean (SD) YBOCS score was 12.97 (8.00), in the mild range, reflecting a mean 46.43% reduction in OCD severity. The pretreatment YBOCS score did not correlate significantly with percent YBOCS reduction ($r = –0.14$; $p = .19$). At pretreatment, the mean (SD) CGI-S score was 4.86 (1.05), indicating marked illness; at posttreatment, the rating on this scale was 3.29 (1.66), indicating mild illness.

YBOCS Prediction of Improvement Ratings

To clarify the relationship between YBOCS percent reductions and CGI improvement ratings, we used ROC analyses[38] in which YBOCS percent reductions, at 5% intervals, were used to predict the gold standard of a CGI rating of "much improved" or "very much improved." For this analysis, a CGI rating of "much improved" or "very much improved" was coded as 1; all other CGI ratings ("minimally improved," "no change," "minimally worse," "much worse," "very much worse") were coded as 0.

Of the 87 participants, 59 (67.8%) met CGI criteria for improvement, and 28 (32.2%) did not receive this rating. The correlation between YBOCS percent reduction and (dichotomous) CGI improvement rating was 0.79. Table 2 shows the quality ROC (QROC) analyses of YBOCS percent reductions, at 5% intervals, from 5% to 70%. For each reduction cutoff, the table shows the level of the test, sensitivity, specificity, predictive value of a positive test, pre-

© Copyright 2005 Physicians Postgraduate Press, Inc.

Table 2. QROC Analysis of the Prediction of Clinical Global Impressions (CGI) Scale Ratings (much improved or very much improved) Using Yale-Brown Obsessive Compulsive Scale (YBOCS) Percent Reductions

YBOCS Reduction (%)	Level of the Test[a]	Sensitivity[b]	Specificity[c]	Predictive Value of a Positive Test[d]	Predictive Value of a Negative Test[e]	Efficiency[f]	κ (1.0)[g]	κ (0.0)[h]	κ (0.5)[i]	χ^2	p Value
≥ 5	0.89	1.00	0.34	0.75	1.00	0.78	1.00	0.26	0.41	22.60	< .001
≥ 10	0.86	1.00	0.41	0.77	1.00	0.80	1.00	0.32	0.48	27.84	< .001
≥ 15	0.85	1.00	0.45	0.78	1.00	0.82	1.00	0.35	0.52	30.57	< .001
≥ 20	0.80	1.00	0.59	0.83	1.00	0.86	1.00	0.49	0.65	42.26	< .001
≥ 25	0.74	0.97	0.72	0.88	0.91	0.89	0.87	0.63	0.73	47.28	< .001
≥ 30	0.71	0.97	0.79	0.90	0.92	0.91	0.88	0.71	0.79	54.33	< .001
≥ 35	0.70	0.95	0.79	0.90	0.88	0.90	0.83	0.70	0.76	50.71	< .001
≥ 40	0.61	0.88	0.93	0.96	0.79	0.90	0.69	0.89	0.78	53.32	< .001
≥ 45	0.59	0.84	0.93	0.96	0.75	0.87	0.63	0.88	0.73	47.98	< .001
≥ 50	0.56	0.83	0.97	0.98	0.74	0.87	0.61	0.94	0.74	49.43	< .001
≥ 55	0.49	0.72	0.97	0.98	0.64	0.80	0.45	0.93	0.61	36.79	< .001
≥ 60	0.41	0.62	1.00	1.00	0.57	0.75	0.35	1.00	0.52	30.71	< .001
≥ 65	0.24	0.36	1.00	1.00	0.44	0.57	0.16	1.00	0.27	13.84	< .001
≥ 70	0.17	0.26	1.00	1.00	0.40	0.51	0.10	1.00	0.19	9.06	< .01

[a]The probability of exceeding the YBOCS reduction cutoff.
[b]The probability of exceeding the YBOCS reduction cutoff among those patients with a positive CGI-I rating.
[c]The probability of failing to exceed the YBOCS reduction cutoff among those patients with a negative CGI-I rating.
[d]The probability of having a positive CGI-I rating among those patients exceeding the YBOCS reduction cutoff.
[e]The probability of having a negative CGI-I rating among those patients failing to exceed the YBOCS reduction cutoff.
[f]The probability that the YBOCS reduction cutoff and the CGI-I rating agree.
[g]Quality index of sensitivity (incorporates the probability of a positive test in the sample).
[h]Quality index of specificity (incorporates the probability of exceeding the YBOCS reduction cutoff in the sample).
[i]Quality index of efficiency (a weighted average of the quality index of sensitivity and the quality index of specificity).
Abbreviations: CGI-I = CGI-Improvement scale, QROC = quality receiver operating curve.

dictive value of a negative test, and efficiency of the test. The table also provides quality indices of sensitivity, specificity, and efficiency (these indices take into account the proportion of participants who exceeded the YBOCS reduction cutoff in the sample).

As is the case in any signal detection analysis, the "optimal" level of a test depends on the aims of the prediction (e.g., in some cases, it may be beneficial to minimize false negatives, in others, to minimize false positives). For most clinical trials, the most important criterion is efficiency, or the probability that the YBOCS percentage cutoff will agree with the CGI rating of improvement. As can be seen in Table 2, a YBOCS reduction of 30% or greater had the highest efficiency score of 0.91, indicating a 91% probability of agreeing with the CGI rating. The predictive value of a positive test was 0.90, and the predictive value of a negative test was 0.92, indicating a low number of false positives or false negatives when using this cutoff score.

When the modal YBOCS percent reduction cutoffs used in previous trials (see above) were examined, the 50% or greater YBOCS reduction requirement for CBT trials had an 87% probability of agreeing with the CGI rating. The predictive value of a positive test was 0.98, but the predictive value of a negative test was only 0.74, indicating that a large number (26%) of true responders are likely to be missed by this cutoff. As described above, medication trials had a bimodal requirement of either a 40% or greater or a 20% or greater YBOCS reduction. A cutoff of a 40% or greater YBOCS reduction had a 90% chance of agreeing with the CGI rating and a predictive value of a positive test of 0.96. However, the predictive value of a negative test was only 0.79, indicating a false-negative rate of 21%. A cutoff of 20% or greater YBOCS reduction had an 86% chance of agreeing with the CGI rating. Here, the predictive value of a negative test was 1.00, indicating that all patients meeting or exceeding this YBOCS reduction cutoff were judged to be treatment responders. However, the predictive value of a positive test was only 0.83, indicating that 17% of patients meeting or exceeding this cutoff were not judged to be treatment responders (false positives).

YBOCS Prediction of Remission Ratings

We examined the use of various YBOCS percent reductions to predict the likelihood that a patient would be rated as in remission. Consistent with common practice, we did not require complete absence of symptoms for this rating; rather, patients received this rating if, at posttreatment, they received a CGI-S rating of "mildly ill" or less ("borderline ill," "normal, not at all ill"). Patients did not receive this rating if their global severity was rated above mild at posttreatment ("moderately ill," "markedly ill," "severely ill," "among the most extremely ill patients"). Of the 87 participants, 52 (59.8%) met CGI criteria for remission, and 35 (40.2%) did not. The correlation between YBOCS percent reduction and (dichotomous) CGI remission rating was 0.71.

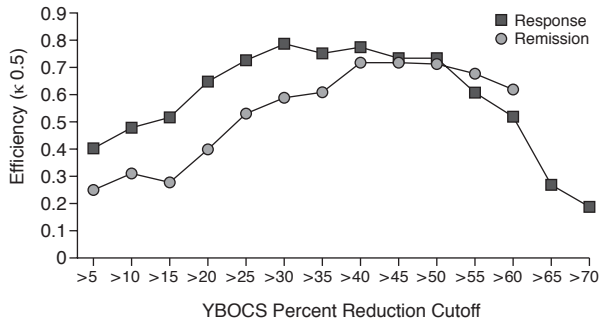
Table 3 shows the QROC analyses of YBOCS percent reductions, at 5% intervals, from 5% to 60%. Whereas (as

Table 3. QROC Analysis of the Prediction of Clinical Global Impressions (CGI) Scale Ratings of Remission (posttreatment mild illness or better) Using Yale-Brown Obsessive Compulsive Scale (YBOCS) Percent Reductions

YBOCS Reduction (%)	Level of the Test[a]	Sensitivity[b]	Specificity[c]	Predictive Value of a Positive Test[d]	Predictive Value of a Negative Test[e]	Efficiency[f]	κ (1.0)[g]	κ (0.0)[h]	κ (0.5)[i]	χ^2	p Value
≥ 5	0.89	0.98	0.24	0.64	0.90	0.67	0.83	0.15	0.25	10.65	< .010
≥ 10	0.86	0.98	0.30	0.66	0.92	0.69	0.86	0.19	0.31	14.04	< .001
≥ 15	0.85	0.96	0.30	0.65	0.85	0.68	0.73	0.18	0.28	11.34	< .001
≥ 20	0.81	0.96	0.41	0.69	0.88	0.73	0.80	0.26	0.40	18.45	< .001
≥ 25	0.73	0.94	0.57	0.75	0.88	0.78	0.78	0.41	0.53	27.98	< .001
≥ 30	0.70	0.94	0.62	0.77	0.88	0.81	0.80	0.46	0.59	32.63	< .001
≥ 35	0.69	0.94	0.65	0.79	0.89	0.82	0.81	0.49	0.61	35.08	< .001
≥ 40	0.60	0.90	0.81	0.87	0.86	0.86	0.75	0.69	0.72	45.48	< .001
≥ 45	0.58	0.88	0.84	0.88	0.84	0.86	0.72	0.72	0.72	45.64	< .001
≥ 50	0.56	0.86	0.86	0.90	0.82	0.86	0.69	0.76	0.72	46.00	< .001
≥ 55	0.49	0.78	0.92	0.93	0.76	0.84	0.58	0.83	0.68	42.44	< .001
≥ 60	0.41	0.69	0.97	0.97	0.69	0.81	0.47	0.93	0.62	38.55	< .001

[a]The probability of exceeding the YBOCS reduction cutoff.
[b]The probability of exceeding the YBOCS reduction cutoff among those patients meeting criteria for mild illness or better.
[c]The probability of failing to exceed the YBOCS reduction cutoff among those patients not meeting criteria for mild illness or better.
[d]The probability of meeting criteria for mild illness or better among those patients exceeding the YBOCS reduction cutoff.
[e]The probability of not meeting criteria for mild illness or better among those patients failing to exceed the YBOCS reduction cutoff.
[f]The probability that the YBOCS reduction cutoff and the mild illness or better rating agree.
[g]Quality index of sensitivity (incorporates the probability of a positive test in the sample).
[h]Quality index of specificity (incorporates the probability of exceeding the YBOCS reduction cutoff in the sample).
[i]Quality index of efficiency (a weighted average of the quality index of sensitivity and the quality index of specificity).
Abbreviation: QROC = quality receiver operating curve.

Figure 1. Quality Index of Efficiency (κ 0.5) for Different Yale-Brown Obsessive Compulsive Scale (YBOCS) Percent Reduction Cutoffs to Predict Response and Remission Status on the Clinical Global Impressions Scale



discussed above) a global rating of improvement was best predicted by a YBOCS reduction of 30% or greater, the more conservative rating of remission was predicted equally well by a YBOCS reduction of 40% or greater, 45% or greater, and 50% or greater (efficiency scores of 0.86 for each cutoff). For the lowest of these reductions (40% reduction; one of the modal cutoffs used in medication trials), the predictive value of a positive test was 0.87 (13% false-positive rate), and the predictive value of a negative test was 0.86 (14% false-negative rate). A YBOCS reduction cutoff of 45% or greater had a predictive value of a positive test of 0.88 (12% false-positive rate) and a predictive value of a negative test of 0.84 (16% false-negative rate). A cutoff of 50% or greater (the modal cutoff used in CBT trials) had a predictive value of a positive test of 0.90 (10% false-positive rate) and a predictive value of a negative test of 0.82 (16% false-negative rate). The other modal cutoff score used in medication trials, a YBOCS reduction of 20% or greater, had only a 73% chance of agreeing with the CGI rating of mild illness or better. The predictive value of a positive test was only 0.69 (31% false-positive rate), and the predictive value of a negative test was 0.88 (12% false-negative rate).

The quality indices of efficiency (κ 0.5) for various YBOCS percent reduction cutoffs to predict response and remission on the CGI are depicted graphically in Figure 1. As can be seen in the figure, the most efficient determinant of response is lower than that of remission (30% vs. 40%, respectively).

Descriptive Analysis of Functioning According to YBOCS Reduction

For descriptive purposes, we examined various aspects of psychosocial functioning for patients whose YBOCS scores decreased within specified cutoff ranges. Table 4 depicts patients whose YBOCS scores decreased from 0% to 100%, by 10% intervals. The outcome variables included the CGI-S and CGI-I ratings, posttreatment YBOCS score, SDS total score, BDI-II score, and STAI-T score. To these outcome variables, we added the percentage of patients meeting criteria for clinically significant change.[39,40]

Clinically significant change is defined by 2 criteria, both of which must be met. (1) A YBOCS score decrease

© Copyright 2005 Physicians Postgraduate Press, Inc.

Table 4. Descriptive Statistics of Posttreatment Measures for Various Yale-Brown Obsessive Compulsive Scale (YBOCS) Percent Reduction Ranges[a]

YBOCS Reduction, %	CGI-S, Mean (SD)	CGI-I (much improved or very much improved), N (%)	YBOCS, Mean (SD)	SDS, Mean (SD)	BDI-II, Mean (SD)	STAI-T, Mean (SD)	Clinically Significant Change, N (%)
0–9 (N = 8)	5.50 (1.41) (N = 8)	0 (0) (N = 8)	24.50 (6.28) (N = 8)	18.57 (6.83) (N = 7)	19.33 (15.63) (N = 3)	56.33 (9.29) (N = 3)	0 (0) (N = 8)
10–19 (N = 5)	4.20 (0.84) (N = 5)	0 (0) (N = 5)	19.40 (5.32) (N = 5)	10.00 (8.00) (N = 3)	7.00 (…) (N = 1)	40.00 (…) (N = 1)	0 (0) (N = 5)
20–29 (N = 8)	4.50 (1.31) (N = 8)	2 (25.0) (N = 8)	18.11 (4.46) (N = 8)	12.83 (5.38) (N = 6)	6.00 (2.16) (N = 4)	49.00 (10.23) (N = 4)	0 (0) (N = 8)
30–39 (N = 8)	4.00 (0.76) (N = 8)	5 (62.5) (N = 8)	16.12 (3.09) (N = 8)	13.83 (7.52) (N = 6)	14.50 (11.59) (N = 6)	42.00 (9.31) (N = 4)	2 (25.0) (N = 8)
40–49 (N = 4)	3.75 (0.96) (N = 4)	3 (75.0) (N = 4)	13.60 (2.88) (N = 4)	13.50 (8.35) (N = 4)	23.33 (12.90) (N = 3)	61.67 (7.51) (N = 3)	3 (60.0) (N = 4)
50–59 (N = 12)	3.08 (1.19) (N = 12)	12 (100) (N = 12)	11.40 (2.26) (N = 12)	12.30 (9.41) (N = 10)	7.00 (6.70) (N = 10)	44.50 (11.71) (N = 10)	11 (91.7) (N = 12)
60–69 (N = 21)	2.57 (0.81) (N = 21)	21 (100) (N = 21)	8.29 (1.42) (N = 21)	5.00 (3.86) (N = 14)	5.87 (6.42) (N = 15)	43.53 (10.95) (N = 15)	21 (100) (N = 21)
70–79 (N = 8)	1.50 (0.76) (N = 8)	8 (100) (N = 8)	6.12 (1.55) (N = 8)	4.67 (3.14) (N = 6)	2.14 (2.12) (N = 7)	33.14 (9.77) (N = 7)	8 (100) (N = 8)
80–89 (N = 4)	1.50 (0.58) (N = 4)	4 (100) (N = 4)	4.00 (0.82) (N = 4)	2.67 (0.58) (N = 3)	3.25 (3.30) (N = 4)	36.50 (3.87) (N = 4)	4 (100) (N = 4)
90–100 (N = 5)	1.00 (0.00) (N = 5)	5 (100) (N = 5)	1.00 (1.00) (N = 5)	2.33 (3.21) (N = 3)	2.50 (3.11) (N = 4)	28.50 (5.97) (N = 4)	5 (100) (N = 5)

[a]This table does not include the 4 patients whose YBOCS scores increased from pretreatment to posttreatment.
Abbreviations: BDI-II = Beck Depression Inventory-II, CGI-I = Clinical Global Impressions-Improvement scale, CGI-S = Clinical Global Impressions-Severity of Illness scale, SDS = Sheehan Disability Scale, STAI-T = State-Trait Anxiety Inventory-Trait version.
Symbol: … = not applicable.

greater than would be expected due to chance (i.e., the score must decrease by at least 1.96 times the standard deviation of that measure, taking into account the reliability of the measure itself). Using data from previous psychometric research on the YBOCS[41] showing an SD of 4.5 and a test-retest reliability of r = 0.88, reliable change on the YBOCS is defined as a decrease of at least 5 points. (2) A posttreatment YBOCS score that more closely resembles the normal range than the clinical range (e.g., at least 2 standard deviations below the clinical mean). Using the pretreatment mean and SD from the present sample, we determined that a score of 13 or below would be outside the pretreatment range. Therefore, patients met criteria for clinically significant change if their YBOCS scores decreased by at least 5 points and their posttreatment scores were 13 or below.

As can be seen in Table 4, at posttreatment, patients whose YBOCS scores decreased by 20% to 29% (20% was one of the most frequently used cutoffs in medication trials) were rated as between moderately and markedly ill on the CGI-S, and only 25% were rated "much improved" or "very much improved" on the CGI-I. On the YBOCS, their OCD symptoms were still within the moderate range. These patients rated themselves as moderately impaired on the SDS and appeared minimally depressed on the BDI-II, although their trait anxiety on the STAI-T was high. None of these patients met criteria for clinically significant change.

Patients whose YBOCS scores decreased by 30% to 39% (a cutoff of 30% was the optimal score for predicting a rating of "much improved" or "very much improved") were rated as moderately ill at posttreatment, and 62.5% were rated "much improved" or "very much improved." OCD severity on the YBOCS was moderate among these patients, and they rated themselves as moderately impaired, mildly depressed, and highly anxious. Twenty-five percent of these patients met criteria for clinically significant change.

The other modal cutoff used in medication trials was 40% or greater (and this was one of the optimal cutoff scores for predicting mild illness or better). Patients whose YBOCS scores decreased by 40% to 49% received moderate illness ratings at posttreatment, although 75% were rated "much improved" or "very much improved." The mean YBOCS score was 14, in the mild range, among these patients, and they rated themselves as moderately impaired, moderately depressed, and highly anxious. Sixty percent met criteria for clinically significant change.

Finally, the modal YBOCS reduction cutoff in CBT trials was 50% or greater. Patients whose YBOCS scores decreased by 50% to 59% were rated as mildly ill at posttreatment, and all were rated "much improved" or "very much improved." Their OCD severity was in the mild range. These patients rated themselves as moderately impaired, minimally depressed, and highly anxious. Ninety-two percent met criteria for clinically significant change.

DISCUSSION

The observed difference in criteria for classifying patients as "treatment responders" highlights the need to develop a standardized criterion of OCD treatment response. As described previously, studies of medications have typically used a YBOCS reduction cutoff of either 20% or 40% to label patients as responders. By comparison, CBT studies have typically required YBOCS scores to decrease by 50% in order for patients to be classified as responders. One might speculate that the different responder criteria across studies reflect a difference in pretreatment severity. For example, it might be more difficult for a patient with a baseline YBOCS score of 38 to achieve a 30% reduction (a reduction of 12 points) than for one with a score of 20 (a reduction of only 6 points). However, the present data argue against this assumption: pretreatment YBOCS score was not significantly associated with percent YBOCS score reduction. In addition, the studies reviewed earlier in this article do not indicate a difference in pretreatment OCD severity between CBT and medication studies.

Another possibility is that a given YBOCS percent reduction, e.g., 20%, has different implications in terms of overall functional improvement for patients treated with pharmacotherapy versus CBT. For example, the antidepressant effects of selective serotonin reuptake inhibitor medications might be argued to increase the functional "meaning" of a 20% YBOCS reduction. In the present study, all patients were receiving some form of CBT, and medications were not studied systematically among these patients (although the majority of patients were taking SRI medications). Thus, a limitation of this study, along with the relatively small sample size (N = 87), is that it might be argued that the obtained results apply only to CBT studies, and not to pharmacology studies. This issue has not received adequate empirical study. In one study,[42] patients receiving medications (unspecified) while waiting for CBT did not show a greater decrease in depression than did patients receiving CBT without medications. In another study,[43] patients receiving 8 weeks of fluoxetine while awaiting CBT did not show a greater decrease in depression than did patients receiving either cognitive therapy or exposure therapy. The most recent comparative treatment study[44] assessed depression before and after treatment with CBT versus clomipramine; when published, these results will likely inform our understanding of whether antidepressant medications reduce comorbid depression to a greater extent than does CBT.

A more likely explanation for the discrepancy between responder criteria in CBT versus medication trials comes from our finding that the criterion correlated positively with the year of the study's publication. The average year of publication for the medication trials was 1994, and the average year for CBT trials was 2000. Thus, it may be that the bar has been raised in OCD treatment trials over time and a greater proportion of CBT trials have been published in recent years.

The present study illustrates how various YBOCS reduction cutoffs perform as indices of treatment response. If, in labeling a patient a "treatment responder," we mean that the patient's condition *improved* by a meaningful degree, a 30% or greater reduction on the YBOCS appears to be the optimal criterion. This designation results in a false-positive rate of only 10%, and a false-negative rate of only 8%. It should be noted, however, that improvement does not necessarily imply good posttreatment functioning. Patients whose YBOCS scores decreased by 39% or less were still rated as moderately ill at posttreatment. We note as well that most clinical trials require a YBOCS score of 16 or higher (signifying at least moderate illness) for inclusion; therefore, the majority of patients whose YBOCS scores decreased by 39% or less would still have met criteria to *enter* the study at posttreatment.

The 20% or greater reduction criterion used by many pharmacologic trials is likely too lenient to allow for adequate evaluation of the effects of the study medication. Although this cutoff eliminates false negatives (i.e., no actual treatment responders are missed), the false-positive rate is high (i.e., many patients are classified as "responders" when they did not, in fact, improve substantially). We note that in a recent large-scale study,[37] OCD patients receiving placebo medications showed an 11% decrease on the YBOCS; therefore, a cutoff of 20% may be insufficient to control for the nonspecific effects of treatment. Adding to this concern is the effect of repeated testing, which may in itself lead to reduced reports of symptom severity.[45–47] By contrast, the 50% or greater YBOCS reduction requirement for CBT had the opposite problem. Although very few actual treatment nonresponders would be labeled "responders" using this criterion (i.e., a low rate of false positives), a high number of actual treatment responders would be missed. Thus, the use of a 50% reduction cutoff may substantially underestimate the strength of an intervention.

If, on the other hand, the "treatment responder" designation is to imply that the patient is in remission (i.e., posttreatment condition is rated mild or better), a criterion of 20% or greater is clearly inadequate for this purpose, with a 31% false-positive rate. Instead, a YBOCS reduction cutoff of 40% to 50% appears optimal. Because differences in signal detection were fairly minimal among cutoff scores in this range, researchers may wish to use the 40% reduction cutoff for this purpose. We note, however, that posttreatment CGI-S scores (Table 4) suggest that the difference between a 40% and 50% YBOCS reduction may represent the difference between moderate and mild posttreatment illness.

We note as well that YBOCS reduction criteria generally appear less stringent than does the criterion of clinically significant change.[39,40] For example, whereas 62.5%

of patients whose YBOCS scores decreased by 30% to 39% were rated "much improved" or "very much improved," only 25% met criteria for clinically significant change. The most likely reason for this finding is that clinically significant change is based on both reduction and absolute score at posttreatment (i.e., a patient's YBOCS score must decrease significantly and be in the mild range at posttreatment). Given the increasing emphasis on clinically significant change in outcome research,[47] it seems particularly important for outcome researchers to define a priori their criteria for meaningful change. It may well be that the best definition comes from multiple data sources, rather than from change scores on any single measure.[49]

Drug names: clomipramine (Anafranil and others), fluoxetine (Prozac and others).

REFERENCES

1. March JS, Frances A, Carpenter D, et al. The Expert Consensus Guideline Series: Treatment of Obsessive-Compulsive Disorder. J Clin Psychiatry 1997;58(suppl 4):1–70
2. DeVeaugh-Geiss J, Landau P, Katz R. Treatment of obsessive-compulsive disorder with clomipramine. Psychiatr Ann 1989;19:97–101
3. Tollefson GD, Rampey AH Jr, Potvin JH, et al. A multicenter investigation of fixed-dose fluoxetine in the treatment of obsessive-compulsive disorder. Arch Gen Psychiatry 1994;51:559–567
4. Montgomery SA, McIntyre A, Osterheider M, et al. A double-blind, placebo-controlled study of fluoxetine in patients with DSM-III-R obsessive-compulsive disorder. The Lilly European OCD Study Group. Eur Neuropsychopharmacol 1993;3:143–152
5. Montgomery SA, Kasper S, Stein DJ, et al. Citalopram 20 mg, 40 mg and 60 mg are all effective and well tolerated compared with placebo in obsessive-compulsive disorder. Int Clin Psychopharmacol 2001;16: 75–86
6. Pato MT, Pigott TA, Hill JL, et al. Controlled comparison of buspirone and clomipramine in obsessive-compulsive disorder. Am J Psychiatry 1991;148:127–129
7. Pigott TA, Pato MT, Bernstein SE, et al. Controlled comparisons of clomipramine and fluoxetine in the treatment of obsessive-compulsive disorder: behavioral and biological results. Arch Gen Psychiatry 1990;47: 926–932
8. Kozak MJ, Liebowitz MR, Foa EB. Cognitive behavior therapy and pharmacotherapy for obsessive-compulsive disorder: the NIMH-sponsored collaborative study. In: Maser JD, ed. Obsessive-Compulsive Disorder: Contemporary Issues in Treatment. Mahwah, NJ: Lawrence Erlbaum Associates; 2000:501–530
9. van Balkom AJ, de Haan E, van Oppen P, et al. Cognitive and behavioral therapies alone versus in combination with fluvoxamine in the treatment of obsessive compulsive disorder. J Nerv Ment Dis 1998;186:492–499
10. Fals-Stewart W, Marks AP, Schafer J. A comparison of behavioral group therapy and individual behavior therapy in treating obsessive-compulsive disorder. J Nerv Ment Dis 1993;181:189–193
11. Cottraux J, Mollard E, Bouvard M, et al. Exposure therapy, fluvoxamine, or combination treatment in obsessive-compulsive disorder: one-year follow-up. Psychiatry Res 1993;49:63–75
12. Lindsay M, Crino R, Andrews G. Controlled trial of exposure and response prevention in obsessive-compulsive disorder. Br J Psychiatry 1997;171:135–139
13. Clomipramine in the treatment of patients with obsessive-compulsive disorder. The Clomipramine Collaborative Study Group. Arch Gen Psychiatry 1991;48:730–738
14. Goodman WK, Price LH, Rasmussen SA, et al. The Yale-Brown Obsessive Compulsive Scale, 1: development, use, and reliability. Arch Gen Psychiatry 1989;46:1006–1011
15. Goodman WK, Price LH, Rasmussen SA, et al. The Yale-Brown Obsessive Compulsive Scale, 2: validity. Arch Gen Psychiatry 1989;46: 1012–1016
16. Cottraux J, Note I, Yao SN, et al. A randomized controlled trial of cognitive therapy versus intensive behavior therapy in obsessive compulsive disorder. Psychother Psychosom 2001;70:288–297
17. Warren R, Thomas JC. Cognitive-behavior therapy of obsessive-compulsive disorder in private practice: an effectiveness study. J Anxiety Disord 2001;15:277–285
18. Simpson HB, Gorfinkle KS, Liebowitz MR. Cognitive-behavioral therapy as an adjunct to serotonin reuptake inhibitors in obsessive-compulsive disorder: an open trial. J Clin Psychiatry 1999;60:584–590
19. Kampman M, Keijsers GP, Hoogduin CA, et al. Addition of cognitive-behaviour therapy for obsessive-compulsive disorder patients non-responding to fluoxetine. Acta Psychiatr Scand 2002;106:314–319
20. Weiss EL, Potenza MN, McDougle CJ, et al. Olanzapine addition in obsessive-compulsive disorder refractory to selective serotonin reuptake inhibitors: an open-label case series. J Clin Psychiatry 1999;60: 524–527
21. DeVeaugh-Geiss J, Katz R, Landau P, et al. Clinical predictors of treatment response in obsessive compulsive disorder: exploratory analyses from multicenter trials of clomipramine. Psychopharmacol Bull 1990; 26:54–59
22. McDougle CJ, Goodman WK, Leckman JF, et al. Haloperidol addition in fluvoxamine-refractory obsessive-compulsive disorder: a double-blind, placebo-controlled study in patients with and without tics. Arch Gen Psychiatry 1994;51:302–308
23. Swets JA, Pickett RM. Evaluation of Diagnostic Systems: Methods From Signal Detection Theory. New York, NY: Academic Press; 1982
24. Guy W. Assessment Manual for Psychopharmacology. Washington, DC: US Government Printing Office; 1976
25. Eisen JL, Goodman WK, Keller MB, et al. Patterns of remission and relapse in obsessive-compulsive disorder: a 2-year prospective study. J Clin Psychiatry 1999;60:346–351
26. Hollander E, Zohar J. Beyond refractory obsessions and anxiety states: toward remission. J Clin Psychiatry 2004;65(suppl 14):3–5
27. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision. Washington, DC: American Psychiatric Association; 2004
28. Brown TA, DiNardo PA, Barlow DH. Anxiety Disorders Interview Schedule for DSM-IV. San Antonio, Tex: The Psychological Corporation; 1994
29. Sheehan DV, Lecrubier Y, Sheehan KH, et al. The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. J Clin Psychiatry 1998;59(suppl 20):22–33
30. Steer RA, Ball R, Ranieri WF, et al. Dimensions of the Beck Depression Inventory-II in clinically depressed outpatients. J Clin Psychol 1999;55:117–128
31. Steer RA, Clark DA, Beck AT, et al. Common and specific dimensions of self-reported anxiety and depression: the BDI-II versus the BDI-IA. Behav Res Ther 1999;37:183–190
32. Beck AT, Steer RA, Ball R, et al. Comparison of Beck Depression Inventories-IA and -II in psychiatric outpatients. J Pers Assess 1996;67: 588–597
33. Spielberger CD, Gorsuch RL, Lushene RE, et al. Manual for the State-Trait Anxiety Inventory. Palo Alto, Calif: Consulting Psychologists Press; 1983
34. Leon AC, Olfson M, Portera L, et al. Assessing psychiatric impairment in primary care with the Sheehan Disability Scale. Int J Psychiatry Med 1997;27:93–105
35. McLean PD, Whittal ML, Thordarson DS, et al. Cognitive versus behavior therapy in the group treatment of obsessive-compulsive disorder. J Consult Clin Psychol 2001;69:205–214
36. Kozak MJ, Foa EB. Mastery of obsessive-compulsive disorder: a cognitive-behavioral approach. San Antonio, Tex: The Psychological Corporation; 1997
37. Foa EB, Wilson R. Stop Obsessing!: How to Overcome Your Obsessions and Compulsions. New York, NY: Bantam Books; 2001
38. Kraemer HC. Evaluating Medical Tests: Objective and Quantitative Guidelines. Newbury Park, Calif: Sage Publications; 1992
39. Jacobson NS, Follette WC, Revenstorf D. Toward a standard definition of clinically significant change. Behav Ther 1984;17:308–311
40. Jacobson NS, Truax P. Clinical significance: a statistical approach to

© COPYRIGHT 2005 PHYSICIANS POSTGRADUATE PRESS, INC.

defining meaningful change in psychotherapy research. J Consult Clin Psychol 1991;59:12–19
41. Steketee G, Frost R, Bogart K. The Yale-Brown Obsessive Compulsive Scale: interview versus self-report. Behav Res Ther 1996;34:675–684
42. O'Connor K, Todorov C, Robillard S, et al. Cognitive-behaviour therapy and medication in the treatment of obsessive-compulsive disorder: a controlled study. Can J Psychiatry 1999;44:64–71
43. van Balkom AJ, de Haan E, van Oppen P, et al. Cognitive and behavioral therapies alone versus in combination with fluvoxamine in the treatment of obsessive compulsive disorder. J Nerv Ment Dis 1998;186:492–499
44. Foa EB, Liebowitz MR, Kozak MJ, et al. Randomized, placebo-controlled trial of exposure and ritual prevention, clomipramine, and their combination in the treatment of obsessive-compulsive disorder. Am J Psychiatry 2005;162:151–161
45. Maltby N, Mayers MF, Allen AJ, et al. Anxiety sensitivity: stability in prospective research. J Anxiety Disord 2005;19:708–716
46. Arrindell WA. Changes in waiting-list patients over time: data on some commonly-used measures: beware! Behav Res Ther 2001;39:1227–1247
47. Storch EA, Murphy TK, Geffken GR, et al. Psychometric evaluation of the Children's Yale-Brown Obsessive Compulsive Scale. Psychiatry Res 2004;129:91–98
48. Ogles BM, Lunnen KM, Bonesteel K. Clinical significance: history, application, and current practice. Clin Psychol Rev 2001;21:421–446
49. Kazdin AE. Almost clinically significant ($p < .10$): current measures may only approach clinical significance. Clin Psychol Sci Pract 2001; 8:455–462