

Possible solutions to the shortcomings of the Yale-Brown Obsessive-Compulsive Scale

Juliana Diniz¹, Victor Fossaluzza¹, Cristina Belotto-Silva¹, Roseli Gedanke Shavitt¹, Carlos Alberto Pereira^{II}

^I Universidade de São Paulo, Faculdade de Medicina, Hospital das Clínicas, Instituto de Psiquiatria, São Paulo, Brazil

^{II} Universidade de São Paulo, Instituto de Matemática e Estatística, Departamento de Estatística, São Paulo, Brazil

OBJECTIVE: The Yale-Brown Obsessive-Compulsive Scale is the most frequently used instrument to measure obsessive-compulsive symptom severity. We describe its shortcomings and propose new methods of evaluating current severity and treatment response.

METHOD: The Yale-Brown Obsessive-Compulsive Scale total and subscale scores were pooled from one cross-sectional study database containing information on 1,000 obsessive-compulsive disorder patients from seven specialized mental health care centers. Additional longitudinal data were pooled for 155 patients who participated in a 12-week trial that evaluated the effectiveness of fluoxetine vs. cognitive-behavior therapy as first-line treatment options. All patients were followed by a clinician who provided a clinical opinion of improvement. Neither patients nor clinicians were aware of the classifications proposed in this study. New methods for using the severity scores were compared with the clinical opinion of improvement.

RESULTS: In the Yale-Brown Obsessive-compulsive scale, the summing-up of subscale scores to compose a total score does not accurately reflect clinical severity. In addition, the reduction of scores with treatment does not usually reach score zero in either subscale. To overcome such problems, we suggest (a) use of the maximum score of any of the subscales; (b) use of a minimum score of 4 in each subscale or 5 for the maximum in any subscale as the goal after treatment. This method performed better than traditional ones regarding sensitivity and specificity against the gold standard represented by the clinical opinion of improvement.

CONCLUSION: The new proposed response criteria are coherent with the clinical opinion of improvement and perform better than the traditional methodology.

KEYWORDS: Obsessive/compulsive disorder; Clinical trials; Obsessive/compulsive disorder evaluation; Instruments.

Diniz J, Fossaluzza V, Belotto-Silva C, Shavitt R, Pereira CA. Possible solutions to the shortcomings of the Yale-Brown Obsessive-Compulsive Scale. *MedicalExpress* (São Paulo, online). 2015; 2(4):M150403

Received for Publication on May 4, 2015; First review on May 12, 2015; Accepted for publication on June 19, 2015

E-mail: julianabelo.diniz@gmail.com

INTRODUCTION

The direct observation and rating of the psychopathological phenomena is technically difficult and costly. As there is a natural fluctuation of psychopathological symptoms over time, ideal ratings should be collected over extended periods of direct observations. In addition, definitions of psychopathological phenomena can be broad, and different raters might disagree about the presence or absence of a specific symptom; thus, long reliability

training, or multiple raters for the same phenomena are an essential aspect of this scenario. Therefore, most researchers choose to rely on the patients' verbal reporting of symptoms' improvement with basis on previously structured interviews to determine treatment outcomes in psychiatric trials.¹

The Yale-Brown Obsessive-Compulsive Scale (Y-BOCS)^{2,3} is one of the most widely used outcome tools in treatment studies of obsessive-compulsive disorder (OCD).⁴ Total Y-BOCS scores vary from 0-40 and intend to grade severity on the basis of time spent with symptoms, interference, associated anxiety, attempts to resist and ability to successfully control obsessions and compulsions. It contains ten question items scoring

DOI: 10.5935/MedicalExpress.2015.04.03

from 0-4; five questions are allotted to obsessions and five to compulsions. Total scores, which may theoretically range from 0 (no symptoms) to 40 (maximum severity) are composed of the sum of the marginal scores for obsessions (0-20) and compulsions (0-20).

Although the usefulness of the Y-BOCS is widely accepted, there are divergences amongst researchers that may hamper the interpretations of the results obtained with this instrument. For example, authors vary in their way of establishing clinical response criteria and cut-off points for remission according to this instrument.^{4,5} The percentage of reduction of the initial Y-BOCS score is frequently used as a measure of response. The standard formula used to calculate reduction is as follows: $(\text{initial Y-BOCS} - \text{final Y-BOCS}) / \text{initial Y-BOCS}$. The cut-offs of meaningful response also differ across studies. The criteria most frequently used are the 25% and 35% reduction from initial scores.^{4,6} Tolin et al.⁷ employed signal detection analysis methodology with the results from the clinical global impression scale as reference and concluded that a Y-BOCS reduction criterion of 30% was optimal for determining improvement, whereas a 40% to 50% reduction criterion was appropriate for predicting a condition of mild illness as the outcome. Similarly, Farris et al.⁸ showed that the widely used 35% reduction criterion of response based on Y-BOCS scores does not represent improvement as reliably as other measures such as clinical global impression, quality of life and social adaptation. However, instead of reestablishing the percent reduction threshold, the authors proposed the use of Y-BOCS cut-off of 14 points to define remission and the use of additional instruments to compose a criterion of wellness.

In addition to the issue of determining the threshold of Y-BOCS percent reduction, it should be noted that it is unusual for patients to reach full-remission, namely to arrive at a final Y-BOCS score of 0 on both obsessions and compulsions.⁹ Therefore, the gap for improvement for a patient whose initial Y-BOCS score was, for example, 30 points, is not the total 30 points but rather the difference between 30 and the aimed cut off to be considered as OCD remission. In this article, we propose new possible ways of calculating an adequate reduction of initial Y-BOCS scores on the basis of realistic expectations of improvement.

To illustrate our proposal, we show the distribution of Y-BOCS results according to the assessment of 1,000 patients with measured Y-BOCS severity scale (clinician administered). In the sequence, we show the results of a treatment study and propose the inclusion of a minimum final Y-BOCS score for the calculation of the percent reduction as a parameter of symptom improvement. We also compare our proposed new response criterion and standardized methods with the clinical opinion of improvement in a clinical trial designed for other purposes.

■ METHODS

The complete methodology involved in data collection has been described elsewhere.¹⁰ Briefly, a national consortium was built for cross-sectional data collection from patients whose primary diagnosis was OCD. The data from the Y-BOCS severity measurements were gathered for 1000 patients at the moment they were admitted for treatment in one of the seven participating centers of specialized mental health care. Y-BOCS severity scores were obtained by trained raters with experience in OCD diagnosis.

The treatment study which provided the pre- and post-treatment Y-BOCS scores has also been described elsewhere.¹¹ In that trial, 155 OCD patients were treated either with a serotonin reuptake inhibitor (fluoxetine was given preference over other antidepressants) or group cognitive behavior therapy and followed for 12-weeks. Raters blinded to the evaluation procedures obtained Y-BOCS scores at week 0 (pre-treatment) and at week 12 (end of trial).

The clinical trial used to compare the conventional clinical impression of improvement with the classification of response proposed in this study was described elsewhere.¹² Fifty-four OCD patients were enrolled to receive add-on treatment as they were considered non-responders to 12 weeks of fluoxetine monotherapy at maximum dosage. All patients were followed in every consultation by a clinician who, in addition to the usual outcome measures, reported his/her clinical opinion on whether the patient was a responder or non-responder to treatment (only endpoint measures were entered in the following analysis). At the time clinicians provided their clinical opinion of improvement they were not aware of the classifications proposed in this study.

■ RESULTS

The distribution of Y-BOCS scores among the 1000 OCD patients is shown in Figure 1.

The total Y-BOCS scores reported for the 1,000 patients are shown in the x-axis, while the maximum Y-BOCS score obtained for obsessions or compulsions is shown in the y-axis. For each possible total Y-BOCS score, various maximum marginal scores are possible. Seven hundred and fifty-six patients (76%) reported Y-BOCS marginal maximum scores that were higher than the half of their total Y-BOCS scores. Therefore, when using the total score to represent symptom severity we are not able to predict the marginal scores and the severity of obsessions or compulsions independently. Obsessions and compulsions marginal scores were significantly correlated (Pearson correlation = 0.73).

The distribution of pre- and post-treatment scores is shown in Figure 2.

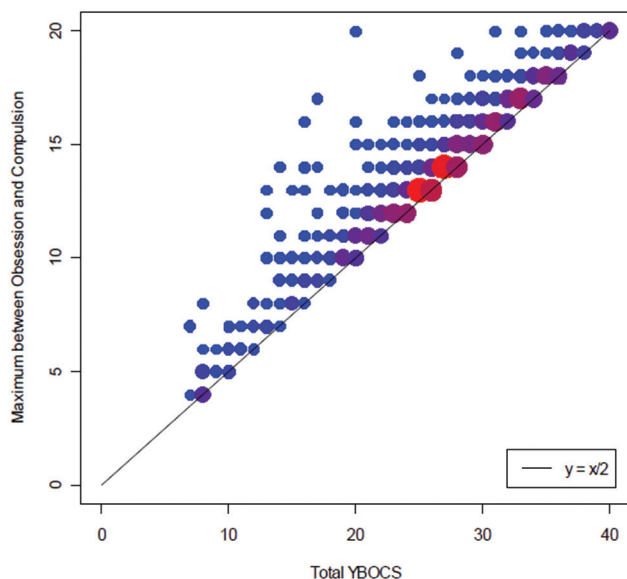


Figure 1 - Distribution of total Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) scores versus maximum Y-BOCS scores between obsessions and compulsions as reported by 1,000 OCD patients admitted to specialized outpatient clinics. The size and color of the points in the graph represent the frequency of occurrence of each value. Higher frequencies are represented by bigger red circles. The diagonal line represents the values for which the marginal scores for obsessions are equal to the marginal scores for compulsions.

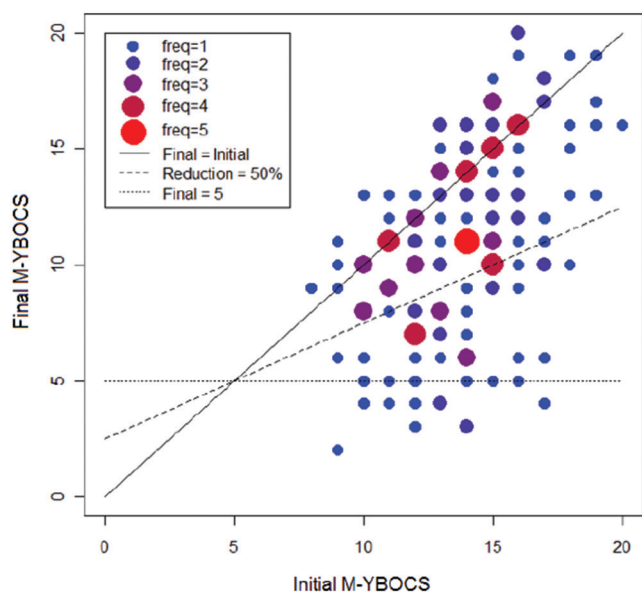


Figure 2 - Distribution of pre-treatment (initial) and post-treatment (final) maximum Yale-Brown Obsessive-Compulsive Scale (M-Y-BOCS) scores between obsessions and compulsions for 155 patients who participated in a trial which evaluated effectiveness of first line treatments for obsessive compulsive disorder. The size and color of the points in the graph represent the frequency of occurrence of each value. Higher frequencies are represented by bigger, red circles. The diagonal line is composed by points for which initial and final M-Y-BOCS scores are the same. The 50% reduction line delimitates the area where patients who improved more than 50% of initial M-Y-BOCS scores are located. The horizontal line marks the final M-Y-BOCS score of 5.

The pre-treatment (initial) maximum Y-BOCS scores are shown in the x-axis while the post-treatment (final) maximum Y-BOCS scores are shown in the y-axis. Sixteen patients (10%, N = 155) present a final marginal maximum Y-BOCS score below or equal to 5 points. Patients who improved are represented below the central continuous diagonal line (when $x = y$). Given a 50% reduction from initial maximum Y-BOCS scores as a cut-off to determine response, the fifty-two patients (34%) represented below the 50% reduction line would be classified as responders.

Results from the comparison between our proposed clinical criteria with the conventional methodologies (25% and 35% reduction from initial Y-BOCS scores) are shown in table 1.

DISCUSSION

From a clinician's standpoint, a patient who, for instance, scores 0 on obsessions and 20 on compulsions may be more severe than a patient who scores 10 both on obsessions and compulsions although both have identical total Y-BOCS scores. Similarly, a patient who scores, say, 20 on obsessions and 0 on compulsions is not half as severe as a patient who scores 20 on obsessions and 20 on compulsions. The distribution of Y-BOCS scores showed in Figure 1 confirms that, for each total Y-BOCS score, different compositions of marginal scores for obsessions and compulsions are possible according to the patients' report.

The high correlation found between obsessions and compulsions suggests that by summing up the marginal scores of a patient, one may inadequately be doubling the information and increasing it artificially; the actual measurement of possible severity (Y-BOCS instead of assuming an integer between 0 and 20, may in fact assume an integer between 0 and 40). This doubling does not correspond to any phenomena observed in the real clinical condition: a patient with a marginal score of 20 may be as severe as one with a total score of 40 – or even worse than one with total of 30. As far as we are aware, this specific issue regarding the summing up of the Y-BOCS marginal scores has never been previously discussed. In previous studies, the threshold of Y-BOCS percent reduction from baseline had already been questioned.^{7,8} However, by simply reestablishing the threshold we do not solve the problem of misrepresentation of a doubled score (i.e., the summing up of subscales that are highly correlated) and do not correct for a more realistic expectation of improvement.

Clinicians are used to giving grades to symptom severity by creating an order based on arbitrary classifications (such as mild, moderate, severe and extremely severe). Nonetheless, the classifications

Table 1 - Cross-tabs and sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) calculations for each criterion of response

Max-Y-BOCS 50%				
			Sensitivity	56%
Clinical opinion	Non-Respondent	Respondent	Specificity	83%
Non-Respondent	30	6	PPV	62%
Respondent	8	10	NPV	79%
			p-value*	0.003
Y-BOCS 25%				
			Sensitivity	61%
	Non-Respondent	Respondent	Specificity	72%
Non-Respondent	26	10	PPV	52%
Respondent	7	11	NPV	79%
			p-value*	0.018
Y-BOCS 35%				
			Sensitivity	39%
	Non-Respondent	Respondent	Specificity	83%
Non-Respondent	30	6	PPV	54%
Respondent	11	7	NPV	73%
			p-value*	0.072

Y-BOCS = Total Score (obsessions plus compulsions) from the Yale-Brown Obsessive-Compulsive Scale; Max-Y-BOCS: Maximum Score (either obsessions or compulsions) from the Yale-Brown Obsessive-Compulsive Scale; Y-BOCS 35%; $((\text{initialY-BOCS-finalY-BOCS})/\text{initialY-BOCS}) * 100 > 35\% = \text{respondent}$; $((\text{initialY-BOCS-finalY-BOCS})/\text{initialY-BOCS}) * 100 > 25\% = \text{respondent}$; M-Y-BOCS 50%; $((\text{initialM-Y-BOCS-finalM-Y-BOCS})/\text{initialM-Y-BOCS-5}) * 100 > 50\% = \text{respondent}$.

we create have not the same properties of continuous scales^{13,14} such as, for example, the Likert scale.¹⁵ It means, for instance, that the distance between mild and moderate is not mathematically the same as the distance between moderate and severe. Therefore, mathematical operations should be applied with caution to these grades.

Even though it does not completely solve the problem, we propose an alternative to the sum of marginal scores to compose the total Y-BOCS score: this alternative would be to use the maximum score obtained for obsessions or compulsions. As, for instance, faced with two patients with a total Y-BOCS score of 20, we should realize that patient A who scores 20 on obsessions is more severe and has a larger range to improve than patient B who scores 10 on obsessions and 10 on compulsions. In our suggested procedure, patient A would be rated 20, whereas Patient B would be rated 10.

Regarding treatment response, the results observed in Figure 2 show that it is uncommon for patients to reach a score of 0 after treatment. Therefore, as we noted before, when we treat a patient we are not expecting an improvement of 100%. For a cut-off point of 35% reduction of the initial Y-BOCS score (to distinguish between responders and non-responders), it is not the same as to say that the patient improved 35% and has the chance to improve another 65% to reach remission. For instance, patients with a starting score of 20 in the Y-BOCS

who improve 10 points (50%) are unlikely to improve the additional 50%, because score zero is rarely reached. In other words, the expectation of 100% improvement given the possibility of a final score of 0 is unrealistic. Therefore, the 50% improvement is quite near the best a patient with an initial score of 20 can reach, given a remission cut-off of a score of 8. This means that if a patient with an initial score of 20 reaches 50% improvement, he can be considered in remission although he improved only half of what it is theoretically possible.

If we consider a lower limit for the Y-BOCS that is higher than 0, we may then have a better picture of each patient's situation. If a 35% reduction is reported by the patient, it means that he still have the remaining 65% to improve. But is a lower limit of 8 points is the goal for the total Y-BOCS score, or 4 for marginal scores, or 5 for maximum between marginal scores, a possible cut-off for response would be 50% reduction of the initial Y-BOCS scores.

Using the clinical opinion of improvement as reference, the criterion of 50% reduction (given a minimum cut-off of 5 points) of the baseline Y-BOCS showed high specificity and reasonable sensitivity. Due to the small sample used to test this criterion, additional trials are still needed to evaluate its performance and determine if it is a better classification of improvement than the methods that are most frequently used.

■ TRIAL REGISTRATIONS

Both trials cited in this manuscript have been registered in the clinicaltrials.gov database (NCT00680602 and NCT00466609).

■ AUTHOR CONTRIBUTIONS

Diniz JB provided intellectual contribution to the original idea that led to this manuscript and wrote most of the text in its present form. Fossaluzza V designed and performed most of the statistical analyses and built the graphs presented as figures. Belotto-Silva C was the principal investigator of the clinical trial from which we gathered data and provided guidance on how to interpret clinical opinion of improvement. Shavitt RG was the mentor behind the clinical trial and the cross-sectional study from which we gathered the data. She also contributed to the writing of the manuscript in its present form. Pereira CAB was the mentor of this work and provided the intellectual basis that led to the critique and suggestions for the use of the YBOCS.

■ CONFLICT OF INTEREST

Authors declare no conflict of interest regarding this publication.

POSSÍVEIS SOLUÇÕES PARA AS DEFICIÊNCIAS DA ESCALA DE YALE-BROWN PARA AVALIAÇÃO DO TRANSTORNO OBSESSIVO-COMPULSIVO

OBJETIVO: A escala de Yale-Brown para avaliação do transtorno obsessivo-compulsivo é o instrumento mais utilizado para medir a gravidade desse transtorno. Descrevemos as deficiências dessa escala e propomos novos métodos de cálculo dos escores para avaliação de gravidade e resposta ao tratamento.

MÉTODO: Os escores totais e subtotais da escala de Yale-Brown foram recuperados de um banco de dados de um estudo transversal com informações sobre 1.000 pacientes com transtorno obsessivo-compulsivos atendidos em sete centros especializados em saúde mental. Foram acrescentados os dados longitudinais de 155 pacientes participantes de um ensaio clínico de 12 semanas que avaliou a eficácia da fluoxetina ou da terapia cognitivo-comportamental como opções de tratamento de primeira linha. Todos os pacientes foram acompanhados por um médico que forneceu um parecer clínico de melhora. Nem os pacientes nem os médicos estavam conscientes das classificações propostas neste estudo. Novos métodos para avaliar os escores de gravidade foram comparados com o parecer clínico de melhora.

RESULTADOS: Na escala obsessivo-compulsiva Yale-Brown, a soma de sub-escalas para compor a pontuação total não reflete com precisão a gravidade clínica. Além disso, a redução da pontuação com o tratamento, normalmente, não atinge o valor zero em qualquer das sub-escalas. Para superar esses problemas, sugerimos (a) o uso da pontuação máxima de qualquer das sub-escalas antes do tratamento; (b) o uso de um score mínimo de 4 em cada sub-escala ou um escore mínimo de 5 como o máximo de qualquer das sub-escalas como a meta para o pós-tratamento. Os novos métodos propostos tiveram melhor desempenho do que os tradicionais quanto a sensibilidade e especificidade contra o padrão ouro representado pelo parecer clínico de melhora.

CONCLUSÃO: Os novos critérios propostos são coerentes com o parecer clínico de melhora e desempenham melhor do que a metodologia tradicional.

PALAVRAS-CHAVE: transtorno obsessivo/compulsivo; ensaios clínicos; avaliação; instrumentos

■ REFERENCES

1. Kraemer H, Telch C. Selection and utilization of outcome measures in psychiatric clinical trials. Report on the 1988 MacArthur Foundation Network I Methodology Institute. *Neuropsychopharmacology*. 1992;7(2):85-94.
2. Goodman W, Price L, Rasmussen S, Mazure C, Fleischmann R, Hill C, et al. The Yale-Brown Obsessive Compulsive Scale. I. Development, use, and reliability. *Arch Gen Psychiatry*. 1989;46(11):1006-11.
3. Goodman W, Price L, Rasmussen S, Mazure C, Delgado P, Heninger G, et al. The Yale-Brown Obsessive Compulsive Scale. II. Validity. *Arch Gen Psychiatry*. 1989;46(11):1012-6.
4. Lewin A, De Nadai A, Park J, Goodman W, Murphy T, Storch E. Refining clinical judgment of treatment outcome in obsessive-compulsive disorder. *Psychiatry Res*. 2010.
5. Storch E, Lewin A, De Nadai A, Murphy T. Defining treatment response and remission in obsessive-compulsive disorder: a signal detection analysis of the Children's Yale-Brown Obsessive Compulsive Scale. *J Am Acad Child Adolesc Psychiatry*. 2010;49(7):708-17.
6. Ferrão Y, Diniz J, Lopes A, Shavitt R, Greenberg B, Miguel E. [Resistance and refractoriness in obsessive-compulsive disorder]. *Rev Bras Psiquiatr*. 2007;29 Suppl 2:S66-76.
7. Tolin DF, Abramowitz JS, Diefenbach GJ. Defining response in clinical trials for obsessive-compulsive disorder: a signal detection analysis of the Yale-Brown obsessive compulsive scale. *J Clin Psychiatry*. 2005;66(12):1549-57.
8. Farris SG, McLean CP, Van Meter PE, Simpson HB, Foa EB. Treatment response, symptom remission, and wellness in obsessive-compulsive disorder. *J Clin Psychiatry*. 2013;74(7):685-90.
9. Pallanti S, Quercioli L. Treatment-refractory obsessive-compulsive disorder: methodological issues, operational definitions and therapeutic lines. *Prog Neuropsychopharmacol Biol Psychiatry*. 2006;30(3):400-12.
10. Miguel E, Ferrão Y, Rosário M, Mathis M, Torres A, Fontenelle L, et al. The Brazilian Research Consortium on Obsessive-Compulsive Spectrum Disorders: recruitment, assessment instruments, methods for the development of multicenter collaborative studies and preliminary results. *Rev Bras Psiquiatr*. 2008;30(3):185-96.
11. Silva CBD. *Estudo comparativo de efetividade da terapia cognitivo-comportamental em grupo e dos inibidores seletivos de recaptacao da serotonina em pacientes com transtorno obsessivo-compulsivo: Um ensaio clínico pragmático*. São Paulo: Universidade de São Paulo; 2009.

12. Diniz JB, Shavitt RG, Fossaluza V, Koran L, Pereira CA, Miguel EC. A double-blind, randomized, controlled trial of fluoxetine plus quetiapine or clomipramine versus fluoxetine plus placebo for obsessive-compulsive disorder. *J Clin Psychopharmacol.* 2011;31(6):763-8.
13. Rochon J. Analyzing bivariate repeated measures for discrete and continuous outcome variables. *Biometrics.* 1996;52(2):740-50.
14. Starmer CF, Lee KL. A mathematical approach to medical decisions: application of Bayes' rule to a mixture of continuous and discrete clinical variables. *Comput Biomed Res.* 1976;9(6):531-41.
15. Grant S, Aitchison T, Henderson E, Christie J, Zare S, McMurray J, et al. A comparison of the reproducibility and the sensitivity to change of visual analogue scales, Borg scales, and Likert scales in normal subjects during submaximal exercise. *Chest.* 1999;116(5):1208-17.