Severity benchmarks and contemporary clinical norms for the Obsessive-Compulsive Inventory-Revised (OCI-R)

Amitai Abramovitch [a,*], Jonathan S. Abramowitz [b], Bradley C. Riemann [c], Dean McKay [d]

[a] *Department of Psychology, Texas State University, 601 University Drive, San Marcos, TX, 78666, USA*
[b] *Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, 235 E. Cameron Avenue, Chapel Hill, NC, 27599-3270, USA*
[c] *Rogers Behavioral Health System, 34700 Valley Road Oconomowoc, WI, 53066, USA*
[d] *Department of Psychology, Fordham University, Bronx, NY, 10458, USA*

ABSTRACT

Introduction: The obsessive-compulsive inventory revised (OCI-R) is commonly used for the assessment of obsessive-compulsive symptomseverity, yet no severity benchmarks have been developed, limiting utilization in clinical practice, clinical research, and analogue studies. Additionally, since its development, no large normative work has been published. We developed empirically derived cutoff scores, examined psychometric properties, and report contemporary norms for the OCI-R.
Methods: Data from 1339 participants with obsessive-compulsive disorder (OCD) were subjected to confirmatory factor analyses, internal consistency analyses, and ROC (sensitivity/specificity) analyses. We used the Yale-Brown Obsessive-Compulsive Scale (Y-BOCS) to determine empirically derived OCI-R severity benchmarks.
Results: A six-factor solution, and good to excellent internal consistency across factors were found. OCI-R severity benchmarks were identified with satisfactory sensitivity and specificity, especially for the cutoff between mild and moderate symptoms. Cutoffs at greater levels of severity were less sensitive and specific.
Discussion: We provide contemporary clinical norms and severity benchmarks for the OCI-R total score with acceptable sensitivity and specificity, and discuss limitations regarding their use. Total scores should be interpreted with caution given the structure of the OCI-R and the heterogeneous nature of OCD.

1. Introduction

Obsessive–compulsive disorder (OCD) is a prevalent, chronic, and often disabling psychological condition. Effective treatments are available in the form of cognitive behavioral therapy (CBT) with exposure and response prevention (ERP) and pharmacotherapy using serotonin reuptake inhibitors (Romanelli, Wu, Gamba, Mojtabai, & Segal, 2014). Decisions about the type and intensity of treatment (e.g., outpatient, partial hospitalization, residential) are typically based on symptom severity (Koran, Hanna, Hollander, Nestadt, & Simpson, 2007), which is commonly defined by scores on the Yale–Brown Obsessive-Compulsive scale (Y-BOCS; Goodman, Price, Rasmussen, Mazure, Fleischmann, et al., 1989). The Y-BOCS includes a comprehensive (up to 60-items) symptom checklist and 10-item severity scale, which makes it ideal for clinical research settings and OCD specialty clinics where it enjoys abundant use. Indeed, the scale is widely considered the "gold standard"

assessment for OCD severity (Deacon & Abramowitz, 2005). Yet because the Y-BOCS requires proper training and can take over an hour to administer, it is less well suited for rapid screening or for use in typical outpatient mental health settings where the majority of individuals with OCD symptoms are evaluated and treated. Accordingly, clinicians in such settings often turn to briefer self-report measures of OCD symptoms for these purposes.

The Y-BOCS was also created as a tool for assessing OCD symptoms in medication trials (Goodman, Price, Rasmussen, Mazure, Fleischmann, et al., 1989). Thus, it is inherently designed for administration to individuals who experience clinically significant obsessions and compulsions. In fact, the measure's manual instructs interviewers to provide the definition of compulsions and obsessions before assessing symptom severity (Marques et al., 2010). Thus, the measure is not an adequate tool for assessing obsessive-compulsive phenomena in non-OCD populations. Yet a great deal of research on OCD uses large analogue

---

samples of individuals without clinically severe symptoms. In such instances, the Y-BOCS may yield a skewed score. Accordingly, like many clinicians, researchers also opt for briefer self-report assessments that can be easily used to screen large numbers of individuals.

The Obsessive-Compulsive Inventory-Revised (OCI-R; Foa et al., 2002) is an 18-item self-report measure that assesses six prototypical domains of OCD symptoms: washing, checking, ordering, neutralizing, obsessing, and hoarding.[1] Items include descriptions of symptoms (e.g., *I check things more often than necessary*) that respondents rate on a scale from 0 to 4 based on the degree of associated distress. The scale takes 7–10 min to complete and is therefore well suited for screening in general clinical and research settings. In fact, in their original investigation, Foa et al. (2002) identified optimal cutoff scores for differentiating between individuals with OCD and (a) those without a psychiatric diagnosis (a score of 21) and (b) those with other anxiety disorders (a score of 18). The OCI-R is one of the most commonly used measures for the assessment of obsessive-compulsive symptoms. Indeed, a meta-analysis conducted in 2013 identified in excess of 450 studies utilizing the OCI-R in various contexts and populations (Abramowitz et al., 2014). Research has also shown the OCI-R to be sensitive to the effects of treatment in adults (Taylor, Abramowitz, McKay, & Garner, in press). Yet despite being one of the more widely used measures of OCD symptoms worldwide, empirically established cutoff scores for delineating symptom severity levels have not been established. Thus, it is unclear how to interpret OCI-R scores in the context of clinical severity.

The absence of empirically derived OCI-R severity cutoff scores has important implications for evidence-based assessment in both clinical work and research within the field of OCD. Indeed, without empirically-derived severity thresholds, clinicians who rely on this widely used measure are making judgments and decisions based on subjective interpretations of patient responses, which potentially diminishes the quality of care and could raise the risk of adverse effects. In the context of clinical research, inclusion criteria are often set for a specific degree of symptom severity, which precludes the reliable use of the OCI-R for such purposes. Finally, because obsessions and compulsions appear on a continuum of severity in the general population, a large body of research with analogue samples (i.e., unscreened participants with varying levels of symptoms) has contributed a great deal to our understanding of the nature and treatment of OCD. Owing to its brevity and simplicity, the OCI-R is among the most commonly used measures in such studies. However, the lack of clinical severity benchmarks prohibits the accurate classification of such analogue samples, as well as benchmarking of results to clinical studies using the Y-BOCS.

Given that the OCI-R is ideal for use in clinical and research settings, contemporary norms, factor structure, reliability, and clinical severity benchmarks are needed to facilitate the comparison of treatment outcomes and establish uniform criteria for clinical cutoffs and improvement. To our knowledge, no study to date has empirically established severity benchmarks for the OCI-R, and large contemporary norms for the measure are not available. Accordingly, we undertook the present study to address these gaps in the literature.

2. Methods

2.1. Participants

The study sample included 1339 adult participants with a DSM-IV or DSM-5 primary diagnosis of OCD, and was comprised of data drawn from outpatient, partial hospitalization, and residential treatment sites across the United States, including the Mayo Clinic outpatient OCD

program in Rochester, MN (n = 53), the Anxiety and Stress Disorders Clinic outpatient OCD program at the University of North Carolina at Chapel Hill (n = 482), and various residential, partial hospitalization, and intensive outpatient centers within the multi-site Rogers Behavioral Health System network of OCD-specific treatment programs (n = 804). Diagnosis was established using a semi-structured or structured diagnostic interview (e.g., the Structured Clinical Interview for DSM disorders [SCID]; Anxiety and Related Disorders Interview Schedule [ADIS]) depending on the screening protocol at these centers.

As is shown in Table 1, the sample included a slightly greater percentage of females than males, as well as a wide age range. The average participant was in their early 30s, had completed at least some college, and had never been married. The vast majority of the sample was White. The group's mean scores on the Y-BOCS total and subscales fell within the moderate to moderate-severe range of OCD symptom severity, yet scores ranged widely.

2.2. Measures

Obsessive-Compulsive Inventory – Revised (OCI-R; Foa et al., 2002). The OCI-R is an 18 item self-report questionnaire assessing obsessive-compulsive symptom severity. Participants indicate their level of distress associated with each of 18 common OCD and hoarding symptoms on a 5-point Likert scale ranging from 0 (not at all) to 4 (very much), resulting in a total score that ranges between 0 and 72. The OCI-R is comprised of 6 factors representing the following symptom domains: checking, ordering, neutralizing, washing, obsessing, and hoarding. Each factor is comprised of 3 items (possible range = 0–12). Overall, the measure demonstrates good internal consistency across populations and geographic locations (Cronbach's α range 0.81- 0.95; Hajcak, Huppert, Simons, & Foa, 2004; Hon, Siu, Cheng, Wong, & Foa, 2019), yet the hoarding and neutralizing subscales demonstrate somewhat weaker internal consistency.

Table 1
Demographic and clinical characteristics.

Variable	Mean/% (n)	SD	Range
Gender			
Female	52.8% (707)		
Male	46.4% (621)		
Age (years)	31.8	12.4	18–79
Education			
Did not complete high school	8.5% (66)		
High school diploma	24.3% (189)		
Some vocational school or college	27.4% (213)		
Vocational or college degree	18.1% (140)		
Graduate degree	21.7% (169)		
Marital Status			
Married	56.3% (519)		
Not married	33.0% (304)		
Separated/Divorced	6.5% (60)		
Widowed	4.2% (39)		
Ethnicity			
Asian	3.1 (32)		
Black/African American	1.8 (19)		
Hispanic or Latino	1.7 (18)		
White	92.0% (948)		
Other/not identified	1.2% (13)		
OCD Severity*			
Y-BOCS Total Score	25.40	6.51	0–40
Y-BOCS Obsessions	12.98	3.39	0–20
Y-BOCS Compulsions	12.42	3.80	0–20

Total sample: N = 1339; Demographic data were missing for some participants: Gender n = 1328 (missing 11, 0.8%); Education n = 777 (missing, n = 562, 42%); Marital status n =922 (missing, n = 419, 31%; Ethnicity n = 1030 (missing 309, 23.1%); OCD severity (Y-BOCS) n = 606 (missing, n = 733, 45.3%). SD = standard deviation; OCD= Obsessive-Compulsive disorder; Y-BOCS = Yale-Brown Obsessive-Compulsive Scale.*All participants were given a primary diagnosis of OCD, and those with a total score of zero on the OCI-R were not included in any analyses. Twof participants had a Y-BOCS total score of zero.

---

[1] In recent years it has become evident that hoarding is not a symptom of OCD (e.g., Abramowitz, Wheaton, & Storch, 2008), as reflected in the conception of Hoarding Disorder as a separate diagnosis in DSM-5 (American Psychiatric Association, 2013).

Yale-Brown Obsessive-Compulsive Scale (Y-BOCS; Goodman, Price, Rasmussen, Mazure, Delgado, et al., 1989; Goodman, Price, Rasmussen, Mazure, Fleischmann, et al., 1989). The Y-BOCS is considered the gold standard measure of OCD symptom severity (Rapp, Bergman, Piacentini, & McGuire, 2016). It includes 10 items that assess the following five parameters of obsessions (items 1–5) and compulsions (items 6–10): time, interference, distress, resistance, and control. Items are rated from 0 (none) to 4 (severe), yielding a total score (range = 0–40) and subscale scores for obsessions and compulsions (range = 0–20 for each). Originally developed as a clinician administered semi-structured interview, the Y-BOCS demonstrates good to excellent internal consistency, test–retest reliability, and interrater reliability (Goodman, Price, Rasmussen, Mazure, Fleischmann, et al., 1989; Rapp et al., 2016; Storch et al., 2005). A self-report version (Y-BOCS-SR) was later developed (Baer, Brown-Beasley, Sorce, & Henriques, 1993), also demonstrating good internal consistency and strong convergent validity with the interview version (Federici et al., 2010; Steketee, Frost, & Bogart, 1996). Data were available from 606 participants who completed the YBOCS: 375 (62%) completed the self-report version and 231 (38%) were administered the interview version. Previous research indicates the two versions may be used interchangeably (Steketee et al., 1996), with the most recent investigation suggesting that the clinician administered version produces a slightly higher total score (M_{Diff} = 1.4, Cohen's d = 0.21; Federici et al., 2010). We, however, found no significant differences in mean total scores across the two administration modalities (interview = 24.90 [5.61]; self-report = 25.44 [6.73]), t (604) = 0.80, p = .42.

2.3. Procedure

All participants presented for services at one of the study sites and were assessed using one of the following instruments administered by a trained assessor: The Mini International Neuropsychiatric Interview (MINI; Sheehan et al., 1998), The Structured Clinical Interview for DSM-IV (SCID; First, Spitzer, Gibbon, & Williams, 2002), and the Anxiety Disorder Interview Schedule for DSM-IV (ADIS; DiNardo, Brown, & Barlow, 1994). Subsequently, the interviewer presented this assessment data to an expert clinician (i.e., the site director or senior clinician) who then met with each patient to review the data. Formal interrater reliability was not assessed at all sites, but only patients for which the interviewer and the expert clinician reached full diagnostic consensus that OCD was the principal diagnosis were included in the study. All participants were administered the Y-BOCS interview or completed the Y-BOCS-SR as part of their assessment and provided consent to allow their responses to be used for both clinical and research purposes.

2.4. Data analysis

Data were initially screened prior to analyses to ensure that relevant statistical assumptions regarding normality were met. Next, data analysis involved four steps. First, we conducted two confirmatory factor analyses (CFA) of the OCI-R: the first to verify the previously determined six-factor solution and the second to examine whether a single higher order factor accounts for the interrelationships among the lower order factors. Second, we investigated reliability via internal consistency coefficients and subscale-total score correlations. Third, we used ROC analyses to examine the OCI-R's ability to differentiate individuals at various levels of OCD severity. Finally, using sensitivity and specificity analyses, we established cutoff scores on the OCI-R with an optimal degree of accuracy for correctly classifying individuals with different levels of OCD severity using the published Y-BOCS benchmarks.

3. Results

3.1. Confirmatory factor analyses

CFAs were conducted using AMOS version 26 (Arbuckle, 2019) with data from 1040 of the 1339 participants (77.7%) for whom individual OCI-R item scores were available (only subscale scores were available for the remaining 299 participants). As suggested by Pincus et al. (2009), we used multiple complementary fit indices to evaluate the specified factor structure. Although chi-square is often used for examining model fit, however, as sample size (and therefore, power) increases, this statistic overestimates lack of fit (Bollen, 1989). Joreskog and Sorbom (1989) and Bentler (1990) therefore advise against using chi-square to judge overall model fit. Accordingly, as suggested by Hu and Bentler (1999), we also evaluated goodness of fit using the standardized root mean-square residual (SRMR), root-mean-square error of approximation (RMSEA), the comparative fit index (CFI), and the Tucker-Lewis index (TLI). Good model fit was defined by the following criteria (Hu & Bentler, 1999): RMSEA ≤0.06; SRMR ≤ 0.05; CFI ≥ 0.95; and TLI ≥0.95. The use of multiple indices provides a conservative and reliable evaluation of model fit relative to the use of a single-fit index.

Fig. 1 presents the factor loadings and correlations among the latent factors for the CFA examining the six-factor solution. As expected, chi-
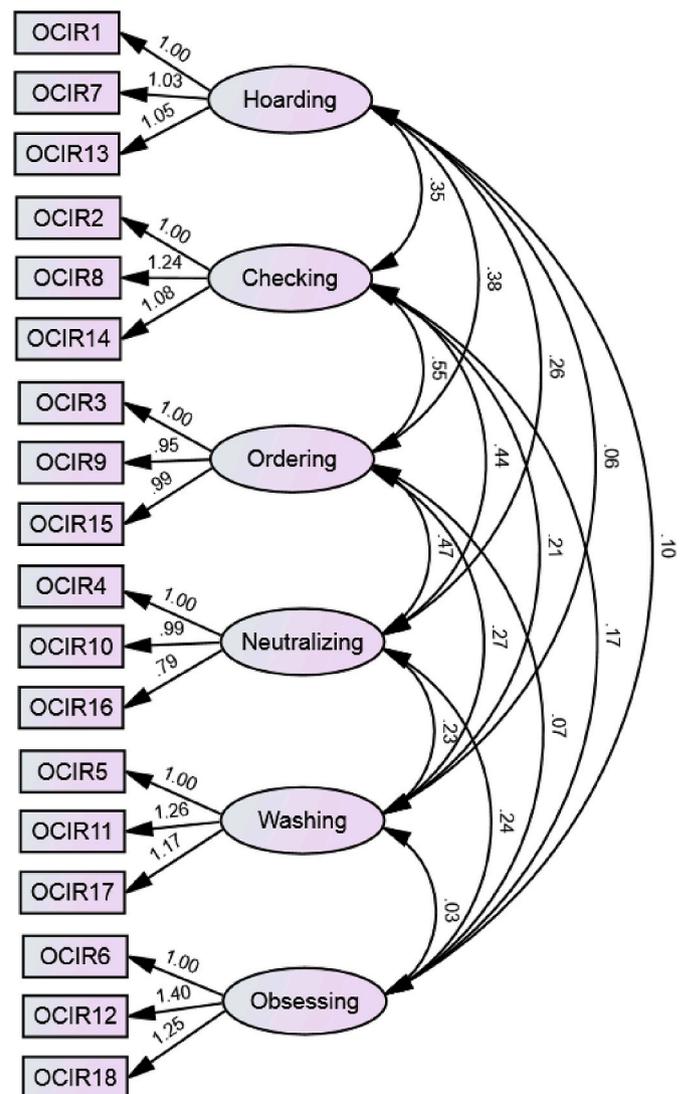

Fig. 1. Factor loadings and correlations among the latent factors, for the six-factor solution of the Obsesasive-Compulsive Inventory-Revised (OCI-R).

square was significant, CHISQ (120, $N = 1040$) = 137.174, $p < .001$; however, all of the other goodness-of-fit indices converged in supporting the fit of the data to the six-factor model: RMSEA = 0.044, SRMR = 0.042, TLI = 0.96, CFI = 0.97. Factor loading estimates revealed that the indicators were strongly related to their purported latent factors, consistent with the position that the OCI-R measures six dimensions of symptoms. Moreover, the latent factors were only weakly to moderately correlated with one another. Means, standard deviations, medians, and interquartile ranges of the OCI-R subscales for the entire OCD sample are presented in Table 2. These scores are similar to those reported in other studies of the OCI-R in individuals with OCD (Abramowitz & Deacon, 2006; Foa et al., 2002; Huppert et al., 2007).

Next, we tested a higher order CFA model to determine whether a single higher order factor accounted for the interrelationships between the lower order factors. The factor loadings are shown in Fig. 2. Again, aside from the expected chi-square result, all of the goodness-of-fit indices suggested that the higher order model fit the data well, CHISQ (129, $N = 1040$) = 469.84 $p < .001$, RMSEA = 0.044, SRMR = 0.044, TLI = 0.958, CFI = 0.969. Inspection of standardized residuals indicated no localized points of ill fit in the solution (e.g., largest standardized residual = 0.12). The first-order factors loaded moderately to very strongly on the higher order factor (range of loadings = 0.45–1.62). The results indicated that the higher order factor accounted for a significant proportion of the variance in the first-order factors (R^2s: Hoarding = 0.48, Checking = 0.73, Ordering = 0.64, Neutralizing = 0.54, Washing = 0.25, Obsessing = 0.23). Given that the higher order solution did not result in a significant decrease in model fit, we concluded that the model provided a good account for the correlations among the first order factors (Brown, 2006).

3.2. Reliability

We used Cronbach's coefficient alpha to assess internal consistency. As is shown in Table 3, values for the OCI-R total score and each of the six subscales were in the good to excellent range (Nunnally & Bernstein, 1994). Correlations between each of the six subscales and the total OCI-R score (also presented in Table 3) indicate generally strong associations with the total score.

3.3. Differentiating symptom severity levels

The first step in evaluating the OCI-R's ability to differentiate individuals at various levels of OCD severity was to choose a pre-determined benchmark for identifying the severity level of each participant. We chose the Y-BOCS total score to serve in this capacity because it is commonly considered the "gold standard" OCD severity measure. There currently exist two sets of Y-BOCS severity cutoff scores. The first is an intuitively derived set that has traditionally been used for decades to demarcate levels of OCD severity (0–7 = subclinical; 8–15 = mild; 16–23 = moderate; 24–31 = severe; 32–40 = extreme). This system, however, has never been subjected to empirical examination (and despite personal communications with numerous experts, we could not

Table 2
Means, Standard deviations, medians, and interquartile ranges for the OCI-R subscales for OCD (n = 1339).

Measure	Mean	SD	Median	IQR	Range
Total Score	25.11	12.72	23	16	1–72
Washing	4.62	4.30	4	9	0–12
Checking	4.09	3.56	3	5	0–12
Ordering	4.00	3.68	3	6	0–12
Obsessing	6.89	3.76	7	6	0–12
Hoarding	2.72	3.26	2	4	0–12
Neutralizing	2.77	3.49	1	4	0–12

OCD: Obsessive-Compulsive disorder; OCI-R: Obsessive-Compulsive Inventory Revised; SD: standard deviation; IQR: Interquartile range.
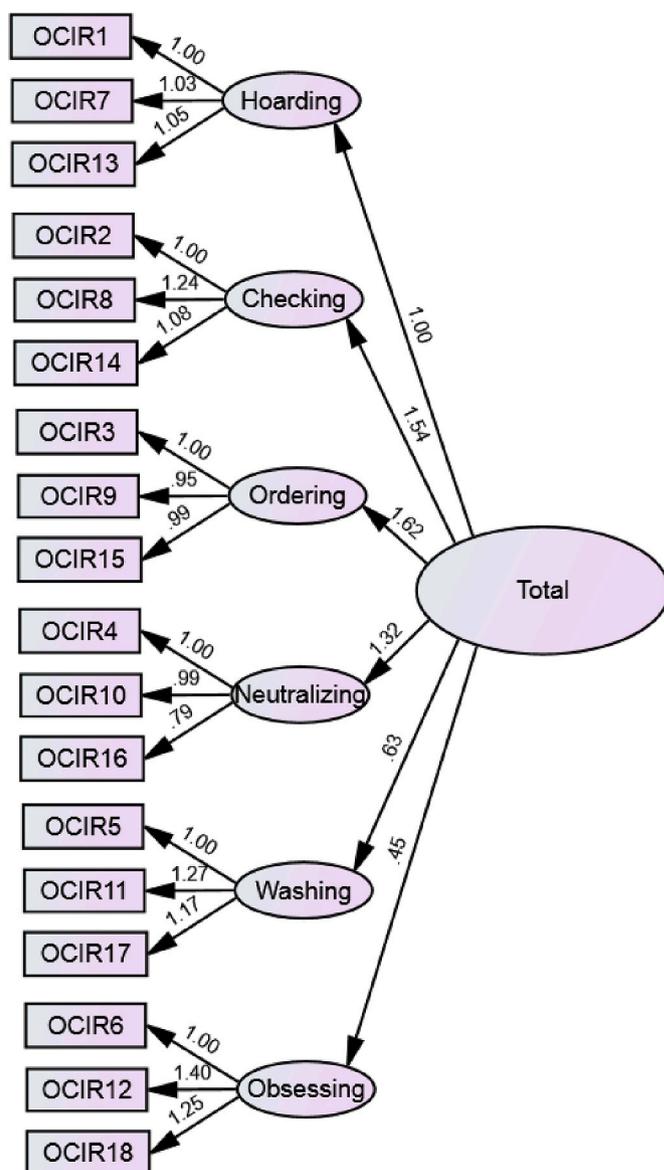

Fig. 2. Factor loadings for the higher order confirmatory factor analysis model of the Obsesasive-Compulsive Inventory-Revised (OCI-R).

Table 3
Reliability of the OCI-R total and subscales.

Measure	Cronbach's α	Correlation with total score
Total score	.83	–
Washing	.90	.51*
Checking	.85	.69*
Ordering	.91	.68*
Obsessing	.84	.45*
Hoarding	.87	.54*
Neutralizing	.83	.63*

*$p < .001$; OCI-R: Obsessive-Compulsive Inventory Revised.

locate its precise origin). Recently, Storch et al. (2015) derived a system of Y-BOCS cutoffs keyed to scores on the Clinical Global Impressions Scale (NIMH, 1976), resulting in the following severity benchmarks: 0–13 = mild; 14–25 = moderate; 26–34 = moderate to severe; and 35–40 = severe. Since these cutoffs were determined empirically, we elected to use them as our benchmark in the present study. Table 4 presents the sample sizes, and OCI-R total and subscale means, standard deviations, and ranges for individuals falling within each Y-BOCS

Table 4
OCI-R total and subscale mean scores, standard deviations, and ranges by Y-BOCS severity group.

Severity level	N	OCI-R						
		Total score	Washing	Checking	Ordering	Obsessing	Hoarding	Neutralizing
Mild	27	14.59 (7.73) 1-30	2.48 (3.63)[a] 0-12	1.74 (1.93) 0-12	1.52 (1.67) 0-6	5.48[d] (3.74) 0-12	2.63[e] (2.69) 0-12	0.74 (1.23) 0-12
Moderate	260	22.52 (11.03) 3-68	3.74 (4.00)[a] 0-12	3.76[b] (3.16) 0-12	3.40 (3.29) 0-12	6.75[d] (3.85) 0-12	2.49[e] (3.03) 0-12	2.38 (3.01) 0-12
Moderate-Severe	278	28.17 (12.58) 3-70	5.01 (4.45) 0-12	4.78 (3.79) 0-12	4.61[c] (3.87) 0-12	7.63 (3.58) 0-12	2.93[e] (3.38) 0-12	3.22[f] (3.66) 0-12
Severe	37	31.60 (11.63) 13-64	6.03 (4.66) 0-12	4.57[b] (3.75) 0-12	5.24[c] (3.75) 0-12	9.00 (3.22) 2-12	3.60[e] (3.56) 0-12	3.14[f] (3.36) 0-12

Means on each subscale with the same superscript were not significantly different from one another ($p > .02$); OCI-R: Obsessive-Compulsive Inventory Revised; Y-BOCS: Yale-Brown Obsessive-Compulsive Scale.

severity group. As can be seen, there were no between group differences on the hoarding subscale; the moderate to severe group failed to differ from the severe group on the ordering and neutralizing subscales; the mild and moderate groups failed to differ with respect to the washing subscale; and the moderate and severe groups failed to differ on the checking subscale.

Next, we conducted a series of receiver operating characteristic (ROC) analyses, which use the association between sensitivity and specificity to estimate the area under the curve (AUC) and indicate how well scores on the OCI-R and the six subscales distinguish between the previously determined states (e.g., moderate vs. severe OCD symptoms) as determined by the Y-BOCS. An AUC value of 1.0 indicates perfect prediction, whereas a value of 0.50 indicates the level of chance.

Table 5 shows the AUC estimates, which ranged from fairly close to chance prediction to well above chance. The OCI-R total score demonstrated the highest AUC in differentiating between the mild and moderate groups and between the moderate and moderate-severe groups; yet the obsessions subscale was best at distinguishing between the moderate-severe and severe groups. The AUCs also indicate that the total score discriminates fairly well at the milder end of the severity continuum, but not as well at the more severe end.

3.4. Establishing optimal cutoff scores for OCD severity levels

Because the OCI-R total score best distinguished between severity groups in two of the three analyses reported above, and for ease of use and interpretation, we determined the total scores with the optimal degree of accuracy for correctly classifying individuals with different levels of Y-BOCS-determined OCD severity. Classification accuracy was evaluated by calculating the sensitivity and specificity of various OCI-R total scores. To illustrate, in differentiating mildly from moderately severe patients, sensitivity refers to the percentage of patients correctly classified by a given OCI-R score as having moderate OCD as determined by the Y-BOCS (i.e., true positives), while specificity refers to the percentage correctly classified as having mild OCD (i.e., true negatives). Table 6 shows the OCI-R cutoff scores that provided the best balance between sensitivity and specificity in classifying patients of different severity levels along with each score's sensitivity and specificity. For

Table 6
Optimal OCI-R total cutoff scores between Y-BOCS severity groups.

Severity level	Y-BOCS cutoff score	Optimal OCI-R total cutoff score	Sensitivity	Specificity
Mild vs. moderate	13	15	.70	.59
Moderate vs. moderate-severe	25	23	.60	.58
Moderate-severe vs. severe	34	28	.53	.51

OCI-R= Obsessive-Compulsive Inventory Revised; Y-BOCS= Yale-Brown Obsessive-Compulsive Scale.

example, an OCI-R total score of 15 correctly classified 70% of moderate (sensitivity) and 59% of mild (specificity) OCD patients as determined by the Storch et al. (2015) Y-BOCS cutoff score of 13 for this dichotomy.

As can be seen in Table 6, the optimal OCI-R cutoff scores were more sensitive and specific at lower levels of severity than at higher levels. Since the moderate-severe category in the Storch et al. (2015) classification scheme had a range of only 4 points on the OCI-R (i.e., from 24 to 28), we tested whether combining participants classified as either moderate or moderate-severe based on Storch et al. into a single "moderate" group would (a) result in the OCI-R total score demonstrating the greatest AUC relative to the subscales, (b) provide a broader (and more clinically useful) range of scores for this classification, and (c) provide a more sensitive and specific optimal cutoff between moderate and severe OCD. ROC analysis indicated an AUC of .65 (95% CI = 0.57 - 0.73) for the OCI-R total score, which was superior to the subscale AUCs (which ranged from 0.52 to 0.64) and revealed that a total score of 27 optimally discriminated between participants with moderate and severe OCD with 0.62 sensitivity and 0.59 specificity. This represented an improved AUC estimate, and also slightly greater sensitivity and specificity when distinguishing severe individuals from those with moderate symptoms. Fig. 3 presents a graphic representation of the final severity benchmarks.
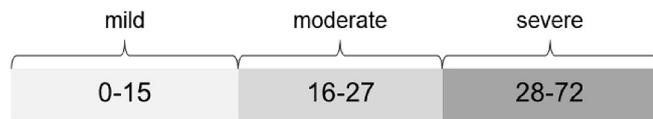
Table 5
Area under the curve (95% confidence interval) estimates for distinguishing between Y-BOCS severity groups.

Comparison	OCI-R						
	Total score	Washing	Checking	Ordering	Obsessing	Hoarding	Neutralizing
Mild vs Moderate	.73 (.62–.81)	.60 (.49–.71)	.70 (.60–.79)	.67 (.58–.76)	.60 (.49–.71)	.45 (.34–.56)	.66 (.57–.76)
Moderate vs. Moderate-Severe	.64 (.59–.68)	.58 (.53–.63)	.57 (.52–.62)	.59 (.54–.63)	.56 (.52–.61)	.53 (.48–.57)	.55 (.50–.60)
Moderate-Severe vs. Severe	.58 (.49–.67)	.56 (.47–.66)	.49 (.39–.59)	.56 (.47–.65)	.61 (.52–.71)	.55 (.45–.65)	.50 (.40–.60)

OCI-R= Obsessive-Compulsive Inventory Revised; Y-BOCS = Yale-Brown Obsessive-Compulsive Scale.

mild	moderate	severe
0-15	16-27	28-72

Fig. 3. Empirically derived severity benchmarks for the Obsessive-Compulsive Inventory-Revised (OCI-R).

4. Discussion

The aims of the present study were to examine the factor structure and reliability of the OCI-R in a large clinical sample of individuals with OCD, provide normative data, evaluate the instrument's ability to distinguish between different OCD severity levels, and determine cutoff scores for identifying individuals at different severity levels. To maximize the generalizability of our findings to those seeking treatment for OCD across the range of clinical settings, we included a large sample of individuals with the full range of symptoms presenting at various outpatient, partial hospitalization, and residential treatment facilities who met DSM criteria for OCD.

Consistent with previous research conducted in countries around the world (e.g., Foa et al., 2002; Fullana et al., 2005; Gonner, Leonhart, & Ecker, 2008; Huppert et al., 2007; Sica et al., 2009), our data supported the six-factor structure of the OCI-R, along with a single higher order factor. Regarding evidence for the reliability of OCI-R scores, we found internal consistency coefficients for the total and subscale scores that fell within the good to excellent range (≥ 0.83). In comparison with other widely used measures of OCD symptoms, these coefficients were somewhat stronger than those reported for the Y-BOCS (Woody, Steketee, & Chambless, 1995) and the most recent version of the Padua Inventory (Burns, Keortge, Formea, & Sternberger, 1996), yet weaker than those found for the Dimensional Obsessive-Compulsive Scale (Abramowitz et al., 2010). Each of the OCI-R subscales were moderately to strongly correlated with the total score; the checking and ordering subscales showing especially strong correlations.

Our ROC analyses indicated that the OCI-R total score was better than the subscales at distinguishing between severity-based groups at the mild, moderate, and moderate-severe levels; but that the obsessions subscale optimally distinguishes between the moderate-severe and severe groups. These findings might be expected given that OCD symptoms are highly heterogeneous and idiosyncratic, whereas the OCI-R's subscales assess specific and quintessential types of obsessions and compulsions. While the total score is likely to capture this heterogeneity better than the individual subscales which assess specific symptom themes (e.g., washing, checking, ordering), the obsessions subscale is a notable exception because it assess obsessions without regard to their theme (e.g., "I am upset by unpleasant thoughts that come into my mind against my will"). Since obsessions are present in one form or another in virtually everyone with OCD, it is not surprising that this subscale would capture severity better than the other more circumscribed subscales; and even better than the total score at higher levels of severity, as we discuss in greater detail next.

Our data indicate that the OCI-R total score is more effective in distinguishing between different levels of OCD severity at the milder end of the severity continuum (e.g., in discriminating between individuals with mild and moderate symptoms) than at the more severe end (e.g., between those with severe vs. extreme symptoms). In fact, our AUC estimates were only modestly above chance prediction in differentiating between patients with moderate and severe Y-BOCS scores. This finding is likely an artifact of how the OCI-R assesses OCD symptoms differently from the Y-BOCS. Scores on the Y-BOCS are likely more sensitive to the heterogeneity and idiosyncrasies of OCD since patients first identify their primary obsessions and compulsions (using the symptom checklist) before rating only these symptoms on the five severity parameters. The OCI-R, however, assesses only a single parameter (distress) of select OCD symptoms using only three items per symptom, and thus (with the

possible exception of the obsessions subscale, as discussed above) confounds severity with the type and range of symptoms present. These differences in approach are less likely to result in discrepancies at the lower end of the severity continuum because individuals with less severe symptoms are likely to have lower scores on the Y-BOCS time, interference, resistance and control items. As severity increases, the Y-BOCS will be more likely to systematically place certain individuals—namely, those with symptoms not assessed by the OCI-R and those with a single, yet very severe type of obsession—in a more severe category than will the OCI-R. Indeed, for such individuals, the ceiling on the OCI-R is lower than it is on the Y-BOCS.

The following clinical example illustrates this problem: Consider an individual with primary contamination obsessions and extensive daily washing rituals, whose symptoms would be considered severe in that they are highly time consuming, greatly impairing, very distressing, and difficult to resist or control. She would likely receive a total Y-BOCS score reflecting a severe degree of symptoms (perhaps above 30), but has no checking, ordering, hoarding, or neutralizing to endorse on the OCI-R. In this case she would have very high scores on the OCI-R washing (and perhaps obsessing) items, but low scores on the other factors, resulting in an OCI-R total score that would incorrectly reflect mild (or perhaps moderate) severity.

A main aim of this study was to identify empirically-based clinically useful OCI-R cutoff scores that clinicians and researchers can use for interpreting OCI-R scores among treatment-seeking adults with OCD, as well as individuals included in analogue studies and in the context of other clinical (e.g., anorexia nervosa), and non-clinical samples. Our sensitivity and specificity analyses using Storch et al.'s (2015) empirically established Y-BOCS cutoffs (after combining their moderate and moderate-severe group) as the standard suggested that OCI-R total scores of up to 15 relate to mild OCD symptoms, 16–27 relate to moderate symptoms, and scores of 28 or greater relate to severe symptoms.

As with the OCI-R's ability to discriminate between different severity levels on the whole, the cutoff scores we identified varied in terms of their sensitivity and specificity. In particular, the cutoff between mild and moderate symptoms (OCI-R total = 15) was reasonably sensitive and specific, correctly identifying about three-quarters of participants in each category. Yet, the cutoff between moderate and severe (OCI-R total = 27) was somewhat less effective, correctly identifying only 58% of individuals in these two categories. Accordingly, although we have used empirical methods to identify optimal cutoff scores in our large and clinically diverse sample, caution is merited when interpreting OCI-R total scores in the context of symptom severity. When it is important to determine OCD severity before recommending treatment options in clinical settings, we recommend that additional data (e.g., clinical interview, Y-BOCS) be considered along with the OCI-R total score.

Indeed, due to the heterogeneous nature of OCD, scales like the OCI-R which potentially confound their assessment of severity with the types, and number of different types, of symptoms have difficulty providing valid estimates of severity (and therefore also effective cutoff scores) for individuals with more serious and complicated symptom presentations. This recognition has implications for the measurement of OCD. First, the vast heterogeneity and idiosyncratic nature of obsessions and compulsions forbids any given self-report measure from including an exhaustive assessment of these symptoms. Yet when scale authors select which symptoms to include, respondents whose obsessions and compulsions happen to match those assessed by items on the measure will appear more severe than those whose symptoms do not match. All three items on the OCI-R's neutralizing scale, for example, pertain to numbers and counting (which are not assessed by most other self-report measures); thus individuals with more counting rituals (as opposed to praying rituals, which are not at all assessed on the OCI-R) will have higher scores than those with fewer counting rituals. Thus, for patients with more severe OCD symptoms, we recommend the use of self-report instruments such as Dimensional Obsessive Compulsive Scale (DOCS; Abramowitz et al., 2010), whose assessment of severity is not influenced

by the type or number of obsessions or compulsions present.

Second, research shows that OCD symptom severity is multidimensional, consisting of parameters such as distress, functional interference, and the frequency or duration of obsessions and compulsions (Deacon & Abramowitz, 2005; Kim, Dysken, Pheley, & Hoover, 1994; McKay, Neziroglu, Stevens, & Yaryura-Tobias, 1998). The OCI-R's exclusive reliance on ratings of distress likely accounts for its problems with sensitivity at greater severity levels. Accordingly, we encourage the use of measures that tap into multiple domains of OCD severity, especially when assessing more severely affected patients. A third limitation of the OCI-R is that it assesses obsessions separately from compulsions, and thus treats these symptoms as disconnected clinical phenomena. Yet structural analyses indicate that OCD psychopathology does not distill into obsessions and compulsions, but rather into dimensions characterized by both obsessions *and* compulsions (Anholt et al., 2010; Deacon & Abramowitz, 2005; McKay et al., 1998). We therefore suggest the use of self-report measures that are grounded in this conceptual framework and capture the link between obsessions and compulsions, such as the DOCS. Finally, the OCI-R includes a 3-item hoarding subscale, which is now known to be distinct from OCD (e.g., Rachman, Elliott, Shafran, & Radomsky, 2009; Wootton et al., 2015). As a result, we suggest either the removal of this subscale, or the use of other self-report measures whose assessment of OCD severity does not consider hoarding symptoms.

A key strength of the present study is the use of a large sample of treatment-seeking individuals with a DSM diagnosis of OCD. Yet this investigation also has several limitations including the lack of ethnic diversity, which unfortunately is a consistent difficulty in research on OCD (Williams, Powers, Yun, & Foa, 2010). Notably, previous reports suggest the OCD symptomatology varies across ethnic lines (Wheaton, Berman, Fabricant, & Abramowitz, 2013), thus future research with more diverse samples is badly needed. Finally, individuals with OCD often meet criteria for comorbid conditions, yet these were not formally assessed across all of the study sites and thus not reported here. However, although comorbidities may attenuate internal validity, it enhances the generalizability of our findings by allowing for the opportunity to observe the properties of the OCI-R in a clinical sample of OCD patients in naturalistic treatment-seeking settings without imposing exclusion criteria that may alter the sample's profile. Indeed, comorbidity information is not a component regularly incorporated in psychometric analyses, developments of measures, cut off scores and norms; as seen in past psychometric work in the field (Abramowitz et al., 2010; Foa et al., 2002; Goodman, Price, Rasmussen, Mazure, Delgado, et al., 1989; Goodman, Price, Rasmussen, Mazure, Fleischmann, et al., 1989; Huppert et al., 2007; Storch et al., 2015). Finally, it should be noted that the Y-BOCS benchmarks published by Storch et al. (2015) were extracted from a Brazilian sample using the Portuguese version of the Y-BOCS, while the present study utilized a US sample.

5. Conclusion

The present study provides an empirically derived set of severity cutoffs, as well as normative data for the OCI-R—one of the most widely used measures for the assessment of OCD symptom severity. However, clinicians and researchers should recognize the specific limitations of this measure, as well as the meaningful conceptual differences between the OCI-R and the Y-BOCS. The implications described previously provide directions for future research to improve the OCI-R. One possible avenue is the addition of further severity parameters, such as functional interference to the distress ratings. The brevity of the OCI-R would likely not be significantly impacted by the addition of such ratings. Another direction is to remove the hoarding subscale, which as we described previously, assesses a symptom that is no longer considered a manifestation of OCD.

References

Abramowitz, J. S., & Deacon, B. J. (2006). Psychometric properties and construct validity of the Obsessive-Compulsive Inventory–Revised: Replication and extension with a clinical sample. *Journal of Anxiety Disorders, 20*(8), 1016–1035. https://doi.org/10.1016/j.janxdis.2006.03.001.

Abramowitz, J. S., Deacon, B. J., Olatunji, B. O., Wheaton, M. G., Berman, N. C., Losardo, D., … Hale, L. R. (2010). Assessment of obsessive-compulsive symptom dimensions: Development and evaluation of the dimensional obsessive-compulsive scale. *Psychological Assessment, 22*(1), 180–198. https://doi.org/10.1037/a0018260.

Abramowitz, J. S., Fabricant, L. E., Taylor, S., Deacon, B. J., McKay, D., & Storch, E. A. (2014). The relevance of analogue studies for understanding obsessions and compulsions. *Clinical Psychology Review, 34*(3), 206–217. https://doi.org/10.1016/j.cpr.2014.01.004.

Abramowitz, J. S., Wheaton, M. G., & Storch, E. A. (2008). The status of hoarding as a symptom of obsessive-compulsive disorder. *Behaviour Research and Therapy, 46*(9), 1026–1033. https://doi.org/10.1016/j.brat.2008.05.006.

American Psychiatric Association. (2013). *Diagnostic and statistical Manual of Mental Disorders : DSM-5* (5th ed.). Arlington, Va: American Psychiatric Association.

Anholt, G., Oppen, P., Cath, D., Smit, J., Den Boer, J., Verbraak, M., et al. (2010). The Yale-Brown obsessive-compulsive scale: Factor structure of a large sample. *Frontiers in Psychiatry, 1*(18). https://doi.org/10.3389/fpsyt.2010.00018.

Arbuckle, J. L. (2019). *Amos*. Chicago: IBM SPSS Version 26.0. .

Baer, L., Brown-Beasley, M. W., Sorce, J., & Henriques, A. I. (1993). Computer-assisted telephone administration of a structured interview for obsessive-compulsive disorder. *American Journal of Psychiatry, 150*(11), 1737–1738. https://doi.org/10.1176/ajp.150.11.1737.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Burns, G. L., Keortge, S. G., Formea, G. M., & Sternberger, L. G. (1996). Revision of the Padua inventory of obsessive compulsive disorder symptoms: Distinctions between worry, obsessions, and compulsions. *Behaviour Research and Therapy, 34*(2), 163–173. https://doi.org/10.1016/0005-7967(95)00035-6.

Deacon, B. J., & Abramowitz, J. S. (2005). The Yale-Brown obsessive compulsive scale: Factor analysis, construct validity, and suggestions for refinement. *Journal of Anxiety Disorders, 19*(5), 573–585. https://doi.org/10.1016/j.janxdis.2004.04.009.

DiNardo, P., Brown, T., & Barlow, D. H. (1994). *Anxiety Disorders Interview Schedule for DSM-IV: Lifetime version (ADIS-IV-LV)*. San Antonio, TX: Psychological Ccorporation.

Federici, A., Summerfeldt, L. J., Harrington, J. L., McCabe, R. E., Purdon, C. L., Rowa, K., et al. (2010). Consistency between self-report and clinician-administered versions of the Yale-Brown obsessive-compulsive scale. *Journal of Anxiety Disorders, 24*(7), 729–733. https://doi.org/10.1016/j.janxdis.2010.05.005.

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). In *Structured Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Patient Edition. (SCID-I/P)*. New York: Biometrics Research, New York State Psychiatric Institute.

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., et al. (2002). The obsessive-compulsive inventory: Development and validation of a short version. *Psychological Assessment, 14*(4), 485–496.

Fullana, M. A., Tortella-Feliu, M., Caseras, X., Andion, O., Torrubia, R., & Mataix-Cols, D. (2005). Psychometric properties of the Spanish version of the Obsessive-Compulsive Inventory–revised in a non-clinical sample. *Journal of Anxiety Disorders, 19*(8), 893–903. https://doi.org/10.1016/j.janxdis.2004.10.004.

Gonner, S., Leonhart, R., & Ecker, W. (2008). The obsessive-compulsive inventory-revised (OCI-R): Validation of the German version in a sample of patients with OCD, anxiety disorders, and depressive disorders. *Journal of Anxiety Disorders, 22*(4), 734–749. https://doi.org/10.1016/j.janxdis.2007.07.007.

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Delgado, P., Heninger, G. R., et al. (1989). The Yale-Brown obsessive compulsive scale. II. Validity. *Archives of General Psychiatry, 46*(11), 1012–1016.

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., … Charney, D. S. (1989). The Yale-Brown obsessive compulsive scale. I. Development, use, and reliability. *Archives of General Psychiatry, 46*(11), 1006–1011.

NIMH. (1976). Clinical global impressions. In W. Guy (Ed.), *ECDEU assessment manual of psychopharmacology* (pp. 218–222). Rockville, MD: United States Department of Health, Education, and Welfare.

Hajcak, G., Huppert, J. D., Simons, R. F., & Foa, E. B. (2004). Psychometric properties of the OCI-R in a college sample. *Behaviour Research and Therapy, 42*(1), 115–123.

Hon, S. K. H., Siu, B. W. M., Cheng, C. W., Wong, W. C. W., & Foa, E. B. (2019). Validation of the Chinese version of obsessive-compulsive inventory-revised. *East Asian Arch Psychiatry, 29*(4), 103–111. https://doi.org/10.12809/eaap1832.

Hu, L.t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118.

Huppert, J. D., Walther, M. R., Hajcak, G., Yadin, E., Foa, E. B., Simpson, H. B., et al. (2007). The OCI-R: Validation of the subscales in a clinical sample. *Journal of Anxiety Disorders, 21*(3), 394–406. https://doi.org/10.1016/j.janxdis.2006.05.006.

Joreskog, K. G., & Sorbom, D. (1989). *LISREL 7: A guide to the program and applications* (2nd ed.). Chicago, IL: SPSS Publications.

Kim, S. W., Dysken, M. W., Pheley, A. M., & Hoover, K. M. (1994). The Yale-Brown obsessive-compulsive scale: Measures of internal consistency. *Psychiatry Research, 51*(2), 203–211.

Koran, L. M., Hanna, G. L., Hollander, E., Nestadt, G., & Simpson, H. B. (2007). Practice guideline for the treatment of patients with obsessive-compulsive disorder. *American Journal of Psychiatry, 164*(7 Suppl), 5–53.

Marques, L., Chosak, A., Simon, N. M., Phan, D. M., Wilhelm, S., & Pollack, M. (2010). Ratings scales for anxiety disorders. In L. Baer, & M. Blais (Eds.), *Handbook of clinical rating scales and assessment in psychiatry and mental health* (pp. 37–72). New York, NY: Humana Press.

McKay, D., Neziroglu, F., Stevens, K., & Yaryura-Tobias, J. A. (1998). The Yale-Brown obsessive-compulsive scale: Confirmatory factor Analytic findings. *Journal of Psychopathology and Behavioral Assessment, 20*(3), 265–274. https://doi.org/10.1023/A:1023019419065.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York, NY: McGraw-Hill.

Pincus, A. L., Ansell, E. B., Pimentel, C. A., Cain, N. M., Wright, A. G. C., & Levy, K. N. (2009). Initial construction and validation of the pathological narcissism inventory. *Psychological Assessment, 21*(3), 365–379. https://doi.org/10.1037/a0016530.

Rachman, S., Elliott, C. M., Shafran, R., & Radomsky, A. S. (2009). Separating hoarding from OCD. *Behaviour Research and Therapy, 47*(6), 520–522. https://doi.org/10.1016/j.brat.2009.02.014.

Rapp, A. M., Bergman, R. L., Piacentini, J., & McGuire, J. F. (2016). Evidence-based assessment of obsessive-compulsive disorder. *Journal of Central Nervous System Disease, 8*, 13–29. https://doi.org/10.4137/jcnsd.S38359.

Romanelli, R. J., Wu, F. M., Gamba, R., Mojtabai, R., & Segal, J. B. (2014). Behavioral therapy and serotonin reuptake inhibitor pharmacotherapy in the treatment of obsessive–compulsive disorder: A systematic review and meta-analysis of head-to-head randomized controlled trials. *Depression and Anxiety, 31*(8), 641–652. https://doi.org/10.1002/da.22232.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., … Dunbar, G. C. (1998). The mini-international neuropsychiatric interview (M.I.N.I.): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry, 59*(Suppl 20), 22–33.

Sica, C., Ghisi, M., Altoe, G., Chiri, L. R., Franceschini, S., Coradeschi, D., et al. (2009). The Italian version of the obsessive compulsive inventory: Its psychometric properties on community and clinical samples. *Journal of Anxiety Disorders, 23*(2), 204–211. https://doi.org/10.1016/j.janxdis.2008.07.001.

Steketee, G., Frost, R., & Bogart, K. (1996). The Yale-Brown obsessive compulsive scale: Interview versus self-report. *Behaviour Research and Therapy, 34*(8), 675–684. https://doi.org/10.1016/0005-7967(96)00036-8.

Storch, E. A., De Nadai, A. S., Conceicao do Rosario, M., Shavitt, R. G., Torres, A. R., Ferrao, Y. A., … Fontenelle, L. F. (2015). Defining clinical severity in adults with obsessive-compulsive disorder. *Comprehensive Psychiatry, 63*, 30–35. https://doi.org/10.1016/j.comppsych.2015.08.007.

Storch, E. A., Shapira, N. A., Dimoulas, E., Geffken, G. R., Murphy, T. K., & Goodman, W. K. (2005). Yale-Brown obsessive compulsive scale: The dimensional structure revisited. *Depression and Anxiety, 22*(1), 28–35. https://doi.org/10.1002/da.20088.

Taylor, S., Abramowitz, J. S., McKay, D., & Garner, L. (in press). Obsessive compulsive and related disorders. In M. M. Anthony & D. H. Barlow (Eds.), Handbook of Assessment and Treatment Planning for Psychological Disorders (3 ed.). (New York, NY).

Wheaton, M. G., Berman, N. C., Fabricant, L. E., & Abramowitz, J. S. (2013). Differences in obsessive-compulsive symptoms and obsessive beliefs: A comparison between african americans, asian americans, latino americans, and European americans. *Cognitive Behaviour Therapy, 42*(1), 9–20. https://doi.org/10.1080/16506073.2012.701663.

Williams, M., Powers, M., Yun, Y. G., & Foa, E. (2010). Minority participation in randomized controlled trials for obsessive-compulsive disorder. *Journal of Anxiety Disorders, 24*(2), 171–177. https://doi.org/10.1016/j.janxdis.2009.11.004.

Woody, S. R., Steketee, G., & Chambless, D. L. (1995). Reliability and validity of the Yale-Brown obsessive-compulsive scale. *Behaviour Research and Therapy, 33*(5), 597–605. https://doi.org/10.1016/0005-7967(94)00076-v.

Wootton, B. M., Diefenbach, G. J., Bragdon, L. B., Steketee, G., Frost, R. O., & Tolin, D. F. (2015). A contemporary psychometric evaluation of the Obsessive Compulsive Inventory-Revised (OCI-R). *Psychological Assessment, 27*(3), 874–882. https://doi.org/10.1037/pas0000075.