A comparison of cluster and factor analytic techniques for identifying symptom-based dimensions of obsessive-compulsive disorder

Duncan H. Cameron[a,*], David L. Streiner[b,e], Laura J. Summerfeldt[d], Karen Rowa[a,b], Margaret C. McKinnon[b,c], Randi E. McCabe[a,b]

[a] Anxiety Treatment and Research Clinic, St. Joseph's Healthcare Hamilton, Canada
[b] Department of Psychiatry and Behavioural Neurosciences, McMaster University, Canada
[c] Mood Disorders Program, St. Joseph's Healthcare Hamilton, Canada
[d] Department of Psychology, Trent University, Canada
[e] Department of Psychiatry, University of Toronto, Canada

ABSTRACT

A growing body of literature suggests that obsessive-compulsive disorder (OCD) is a heterogeneous condition. The studies investigating symptom dimensions have been limited by numerous methodological differences and sample characteristics. The purpose of this study was to compare the two most commonly applied statistical techniques used in addressing this question in the same large cohort of individuals with OCD. Both cluster analysis and factor analysis were used to examine OCD symptom data as measured by the Yale-Brown Obsessive Compulsive Scale (Y-BOCS) Symptom Checklist for 355 individuals with a primary diagnosis of OCD. The factor analysis revealed a three-factor model best described as symmetry obsessions/ordering compulsions, contamination obsessions/cleaning compulsions and aggressive obsessions/checking compulsions. In contrast, the cluster analysis yielded a stable four-cluster solution best described as symmetry obsessions/ordering compulsions, contamination obsessions/cleaning compulsions, aggressive-somatic-religious obsessions/checking compulsions and a mixed symptom profile. Although there was overlap in the models resulting from these two statistical approaches, cluster analysis better captured the dimensional nature of OCD by demonstrating the prevalence of symptom categories in each subgroup. Though both analyses are capable of providing similar outputs, the validity of these results is limited given the input of a priori symptom categories from the Y-BOCS.

1. Introduction

A significant body of literature has developed over the past two decades investigating whether obsessive-compulsive disorder (OCD) can—or should—be classified into subgroups based on symptom presentation (McKay et al., 2004; Radomsky and Taylor, 2005; Starcevic and Brakoulias, 2008). The fact that several individuals could be provided the same diagnosis while presenting with vastly different symptoms has led many clinicians and researchers to accept that OCD is a heterogeneous disorder, but whether, and how, this heterogeneity can be defined in any meaningful way remains a debatable topic. Further, given that no concrete dimensions have been universally accepted, their utility for both clinical and research applications remains to be seen.

Despite these lingering issues, the increasingly large pool of research addressing the topic continues to provide some insight into possible dimensions of OCD. Existing studies that have attempted to define symptom dimensions are listed in Table 1, which shows the method of analysis and the dimensions found. As can be seen here, the large majority of studies used principal components analysis (PCA), exploratory factor analysis (EFA), or confirmatory factor analysis (CFA), while some also use cluster analysis. Latent class analysis (LCA) is becoming an increasingly popular method for defining latent subgroups within a larger population, and one study has attempted to do this in OCD. Table 1 also shows that these statistical analyses are applied to the most frequently used symptom measure for this population, the Yale-Brown Obsessive Compulsive Scale (Y-BOCS; Goodman et al., 1989), which consists of a symptom checklist (with options for present, absent, or lifetime) and a symptom severity rating scale.

Researchers have attempted several methods of coding responses on the Y-BOCS symptom checklist (Y-BOCS-SC) to yield meaningful category scores which lend themselves to classification or dimension reduction statistical analyses. The most ubiquitous method—originally

---

Table 1
Summary of past approaches to defining symptom dimensions in OCD.

Author, Year	Population (child/adult)	Symptoms	Scoring method	Analysis type (item level or category)	Analysis method	Factors/clusters
Baer (1994)	Adult	Current	0,1,2	Category	PCA Varimax	Factor 1: Hoarding obsessions and compulsions, repeating, ordering, counting (20.7%) Factor 2: Contamination, cleaning, somatic, checking (16.0%) Factor 3: Aggression, sexual, religious (11.3%)
Hantouche and Lancrenon (1996)	Adult	Current	0,1	Both	PCA Varimax	Factor 1: Hoarding obsessions and compulsions, symmetry, repeating, ordering, and checking Factor 2: Aggressive, sexual, religious, miscellaneous obsessions and compulsions Factor 3: Cleaning, contamination, and somatic obsessions
Leckman et al (1997)	Adult	Lifetime	Sum total symptoms per category	Category	PCA Varimax	Factor 1: Aggression, sexual, religious, somatic, checking (30.1%) Factor 2: Symmetry, ordering, repeating, counting (13.8%) Factor 3: Contamination, cleaning (10.2%) Factor 4: Hoarding obsessions and compulsions (8.5%)
Mataix-Cols et al. (1999)	Adult	Current	0,1,2	Category	PCA Varimax	Factor 1: Symmetry, ordering, repeating, counting (19.0%) Factor 2: Hoarding obsessions and compulsions (13.8%) Factor 3: Contamination, cleaning (12.7%) Factor 4: Aggressive, checking (10.4%) Factor 5: Sexual, religious (9.7%)
Tek and Ulug (2001)	Adult	Current	0,1	Category	PCA Varimax	Factor 1: Contamination, cleaning, repeating (17.8%) Factor 2: Symmetry, somatic, ordering (15.2%) Factor 3: Aggressive, counting (13.7%) Factor 4: Sexual, religious (9.7%) Factor 5: Hoarding compulsions and checking (9.2%)
Mataix-Cols et al. (2002)	Adult	Current	0,1,2	Category	PCA Varimax	Factor 1: Aggressive, religious, checking, repeating, and counting (23.3%) Factor 2: Contamination, cleaning (13.3%) Factor 3: Symmetry, ordering, counting (10.9%) Factor 4: Hoarding obsessions and compulsions (8.3%) Factor 5: Sexual, somatic (7.9%)
Cavallini et al. (2002)	Adult	Lifetime	0,1	Category	PCA Varimax	Factor 1: Contamination, cleaning (17.0%) Factor 2: Hoarding obsessions and compulsions (13.0%) Factor 3: Aggressive, sexual, somatic, religious, checking, repeating (11.5%) Factor 4: Symmetry, ordering, religious (9.5%) Factor 5: Repeating, counting, symmetry (8.8%)
Feinstein et al. (2003)	Adult	Current	0,1	Both	PCA Varimax	Factor 1: Symmetry, ordering, repeating, counting (14.2%) Factor 2: Contamination, cleaning, aggressive, checking (14.2%) Factor 3: Hoarding obsessions and compulsions (13.9%) Factor 4: Sexual, religious (11.8%)
Denys et al. (2004a)	Adult	Current	0,1,2	Item	PCA Varimax	Factor 1: Contamination, cleaning (16.4%) Factor 2: Aggressive, sexual, religious (9.8%) Factor 3: Somatic and checking (5.9%) Factor 4: Symmetry, repeating, ordering, counting, hoarding obsessions and compulsions (5.6%) Factor 5: High-risk and checking (4.0%)
Denys et al. (2004b)	Adult	Current	0,1,2	Item	PCA Varimax	Factor 1: Aggressive, sexual, religious (14.5%) Factor 2: Contamination, cleaning, washing (11%) Factor 3: Somatic, checking (6.3%) Factor 4: Symmetry, exactness, arranging, ordering (5.8%) Factor 5: High risk assessment and checking (4.8%)

Table 1 (*continued*)

Author, Year	Population (child/adult)	Symptoms	Scoring method	Analysis type (item level or category)	Analysis method	Factors/clusters
Kim et al. (2005)	Adult	Current	0,1,2	Category	PCA Varimax	Factor 1: Hoarding obsessions and compulsions, repeating, ordering, counting (34.0%); Factor 2: Contamination, cleaning (11.1%); Factor 3: Aggression, sexual (10.0%); Factor 4: Religious, somatic (7.7%)
Delorme et al. (2006)	Child	Current	Total symptoms per category	Category	PCA Varimax	Factor 1: Symmetry, checking, repeating, ordering (35.2%); Factor 2: Aggressive, sexual, somatic, counting (12.7%); Factor 3: Contamination, religious, cleaning (11%); Factor 4: Hoarding obsessions and compulsions (8.1%)
McKay et al. (2006)	Child	Current	Symptom number	Category	PCA Oblimin	Factor 1: Cleaning, checking, repeating, counting, ordering, superstitious behaviours, rituals involving others (17.3%); Factor 2: Aggressive, sexual, magical thinking (15.9%); Factor 3: Contamination, aggressive,sexual, magical thoughts, somatic, religious, repeating, counting, symmetry, rituals involving others (12.7%); Factor 4: Hoarding somatic, counting, ordering (11.9%)
Pinto et al. (2007)	Adult	Current	Number endorsed divided by total number in category	Category	PCA Varimax	Factor 1: Symmetry, ordering, repeating, counting (22.5%); Factor 2: Hoarding obsessions and compulsions (13.3%); Factor 3: Pathological doubt, somatic, checking (12.2%); Factor 4: Contamination and cleaning (10.3%); Factor 5: Aggressive, sexual, religious (7.4%)
Cullen et al. (2007)	Adult	Lifetime	0,1	Category	Dichotomous EFA Oblique	Factor 1: Aggressive, sexual, religious, somatic Factor 2: Cleaning, contamination Factor 3: Symmetry obsessions, repeating, ordering, counting, sensory/ motor compulsions Factor 4: Hoarding obsessions and compulsions
Hasler et al. (2007)	Adult	Lifetime	0,1	Category	PCA Promax	Factor 1: Aggressive, sexual, religious, somatic, and checking (17.7%); Factor 2: Symmetry, repeating, counting, ordering (15.4%); Factor 3: Cleaning, contamination (15.4%); Factor 4: Hoarding obsessions and compulsions (14.2%)
Stewart et al. (2007)	Child	Lifetime	Total symptoms per category	Category	PCA Promax	Factor 1: Symmetry, ordering, repeating, checking (27.0%); Factor 2: Contamination, cleaning, aggressive, somatic (14.4%); Factor 3: Hoarding obsessions and compulsions (11.1%); Factor 4: Religious, sexual (9.1%)
Matsunaga et al. (2008)	Adult	Lifetime	0,1,2	Category	PCA Varimax	Factor 1: Cleaning, contamination (21.2%); Factor 2: Hoarding obsessions and compulsions (14.3%); Factor 3: Symmetry, repeating, ordering (11.9%); Factor 4: Aggression, checking (10.3%)
Mataix-Cols et al. (2008)	Child	Lifetime	Total symptoms per category	Category	PCA Varimax	Factor 1: Hoarding obsessions and compulsions, checking (14.1%); Factor 2: Aggressive, sexual, religious (13.7%); Factor 3: Contamination, cleaning, somatic (13.6%); Factor 4: Symmetry, ordering, repeating, checking (13.1%)
Pinto et al. (2008)	Adult	Lifetime	0,1	Item	Dichotomous Factor Analysis Varimax	Factor 1: Taboo Thoughts (22.4%); Factor 2: Symmetry, ordering (11.6%); Factor 3: Hoarding (6.9%); Factor 4: Contamination, cleaning (6.6%); Factor 5: Doubt, checking (4.9%)
Katerberg et al. (2010)	Adult	Lifetime	Number endorsed divided by total number in category	Both	PCA Promax	Factor 1: Symmetry, ordering, counting, repeating; Factor 2: Aggressive, sexual, religious, checking; Factor 3: Contamination, cleaning; Factor 4: Hoarding

Table 1 (*continued*)

Author, Year	Population (child/adult)	Symptoms	Scoring method	Analysis type (item level or category)	Analysis method	Factors/clusters
Asadi et al. (2016)	Adult	Current	0,1 and severity scores	Item	Maximum Likelihood Varimax	Factor 1: Aggression, checking (19.5%); Factor 2: Contamination, cleaning (5.8%); Factor 3: Symmetry, ordering, counting, repeating, hoarding (3.5%); Factor 4: Sexual (2.7%); Factor 5: Somatic (2.7%)
Confirmatory Factor Analysis						
Summerfeldt et al. (1999)	Adult	Current	0,1	Category	CFA Maximum Likelihood	Factor 1: Aggressive, sexual, religious, somatic, checking; Factor 2: Symmetry, repeating, counting, ordering; Factor 3: Contamination, cleaning, washing; Factor 4: Hoarding, saving, collecting
Summerfeldt et al. (2004)	Adult	Current	Number endorsed divided by number in category; weighted scores summed into CFA categories	Both	Logistic Regression/CFA	Factor 1: Obsessions and checking; Factor 2: Symmetry; Factor 3: Contamination and cleaning; Factor 4: Hoarding
Stewart et al. (2008)	Adult/child	Current	Total symptoms per category	Category	CFA Maximum Likelihood	Factor 1: Aggressive, sexual, religious, somatic, checking; Factor 2: Symmetry, ordering, counting, repeating; Factor 3: Contamination, cleaning; Factor 4: Hoarding
Katerberg et al. (2010)	Adult	Lifetime	Number endorsed divided by total number in category	Item	CFA Weighted Least Squares	Factor 1: Taboo (sexual, aggressive, religious); Factor 2: Contamination, cleaning; Factor 3: doubts (obsessions related to fear, compulsions related to these fears); Factor 4: rituals/superstition (superstitions obsessions, eating and mental rituals); Factor 5: hoarding/symmetry (hoarding, symmetry, ordering, arranging, fear of losing things)
Bernstein et al. (2013)	Child	Current	Total symptoms per category	Item	CFA Maximum Likelihood	Factor 1: Contamination, somatic, cleaning; Factor 2: Magical, checking, repeating, counting, ordering; Factor 3: Hoarding; Factor 4: Aggressive, sexual, religious
Cluster Analysis						
Calamari et al. (1999)	Adult	Current	0,1,2	Items	Ward's Method	Cluster 1: Harming; Cluster 2: Hoarding; Cluster 3: Contamination; Cluster 4: Certainty; Cluster 5: Obsessionals
Calamari et al. (2004)	Adult	Current	0,1,2	Items	Ward's Method and K-means	Cluster 1: Contamination; Cluster 2: Harming; Cluster 3: Obsessionals; Cluster 4: Certainty; Cluster 5: Contamination/Harming; Cluster 6: Symmetry; Cluster 7: Low symptoms
Lochner et al. (2008)	Adult	Current	0,1	Items	CA Ward's Method	Cluster 1: Contamination, washing; Cluster 2: Hoarding, collecting; Cluster 3: symmetry, order, arranging, repetitive rituals, counting, checking; Cluster 4: Sexual; Cluster 5: Somatic, religious, diverse

Table 1 (*continued*)

Author, Year	Population (child/adult)	Symptoms	Scoring method	Analysis type (item level or category)	Analysis method	Factors/clusters
Hasanpour et al. (2017)	Adult	Current	Symptom Checklist and Severity	Both	Comparison of 5 CA techniques	Cluster 6: Aggressive, harm-related Cluster 1: Higher symptom severity Cluster 2: Lower symptom severity Significant clusters did not differ as a result of symptom presentation, but rather by symptom severity only
Latent Class Analysis						
Delucchi et al. (2011)	Adult	Lifetime	0,1 and Severity	Items	LCA	Classes did not differ in terms of symptom presentation but by level of symptom endorsement.

Legend: CA = cluster analysis; CFA = confirmatory factor analysis; EFA = exploratory factor analysis; LCA = latent class analysis; PCA = principal components analysis.

proposed in Baer's (1994) seminal work in this area, modified slightly—has been to provide a score of 0 if any symptom within a category is rated as absent, 1 if a symptom is present, and 2 if that symptom is rated as present *and* also listed as one of the respondent's most troubling obsessions or compulsions in the Y-BOCS severity rating scale. These scores are provided for each of the 13 rationally derived categories of the Y-BOCS, which include aggressive, contamination, sexual, hoarding, religious, symmetry, and somatic obsessions, as well as cleaning, checking, repeating, counting, ordering, and hoarding compulsions. There are also two miscellaneous categories to capture any additional symptoms, although these are almost always left out of analyses given their heterogeneous nature (see Summerfeldt et al., 2004 for item-level confirmatory factor analysis including miscellaneous symptoms).

Two other scoring methods are commonly seen. One is to simply rate the symptom as present versus absent, as 1 or 0, respectively, though this method is criticized for not capturing the important symptoms which may be most dominant in a patient's presentation or for artificially inflating the importance of less dominant symptoms (e.g., Pinto et al., 2008). Another strategy is to generate a weighted score for each symptom category by dividing the number of total symptoms endorsed as present by the total number of symptoms in that category. This method is likely the most appropriate for confirmatory factor analysis techniques which might require summing these weighted scores into hypothesized factors (as in Summerfeldt et al., 2004). However, for exploratory factor analysis this method provides scores that are mostly a reflection of the number of items in each symptom category. For example, an individual might endorse 5/9 possible symptoms in the cleaning category, yielding a weighted score of 0.556, and might also list that symptom as their most troubling. But if that person also endorses the one possible item in the counting category, this would yield a score of 1.0 (the highest possible) even if this symptom is only very rarely experienced. The 0-1-2 scoring system is not without its own pitfalls. In contrast to the weighted scores, this method reflects the severity of an individual's symptoms but only for those listed as currently most troubling, and the total number of symptoms is disregarded.

The rightmost column of Table 1 displays the symptom dimensions found in each study. As can be seen, there is relatively consistent overlap in at least three groups. These most often consist of aggressive and/or uncertainty (sometimes including sexual and religious) obsessions and checking compulsions, contamination obsessions and cleaning compulsions, and symmetry obsessions with ordering/arranging compulsions. A fourth dimension, hoarding, is also commonly reported. However, given that the majority of these studies were published prior to the release of the Diagnostic and Statistical Manual, 5th Edition (DSM-5; American Psychiatric Association, 2013) it cannot be determined whether these symptoms described hoarding within the context of OCD or whether they constituted hoarding disorder criteria. Hoarding is also the only symptom on the Y-BOCS to have one item only in each of the obsessions and compulsions sections with the same label ("hoarding") for both, making it highly likely that these two items would be endorsed together, and consequently load together as a factor of their own in factor analysis. Other than these four domains, those remaining—if more than three to four are reported—tend to have great variability in their labels and in the symptoms included. A decade has passed since the publication of a meta-analysis on the existing studies of the factor structure of OCD (Bloch et al., 2008). Here, a four factor structure was suggested with symmetry (including symmetry obsessions with counting, repeating and ordering compulsions), forbidden thoughts (including aggressive, religious, sexual and somatic obsessions with checking compulsions), contamination (with contamination obsessions and cleaning compulsions), and hoarding (with hoarding obsessions and compulsions only).

As evidenced by the number of studies listed in Table 1 (as well as the existence of a meta-analysis on the topic), many researchers have chosen to attempt to define symptom dimensions of OCD using factor

analysis. Though factor analysis is by far the most commonly employed method for defining these dimensions, there is great variability in the type of factor analysis that can be carried out. PCA—which is not, technically speaking, factor analysis—is a common factor extraction method and provides the expression of each component (factor) as a linear combination of the input variables, and requires that the components be orthogonal. This technique attempts to create components that maximize inter-individual variance. This is, in a sense, contrary to the goal of attempting to define symptom dimensions of OCD. Principal axis factoring (PAF), conversely, is an exploratory factor analysis technique—and often referred to as "common factor analysis" or "exploratory factor analysis"—that attempts to define latent constructs, or underlying similarities that we cannot overtly observe, and it is assumed that the variables are linear combinations of these latent factors. In the context of OCD symptom dimensions, employing this method would imply that there is some hypothesized latent structure in obsessive-compulsive symptom presentation that can be viewed by similarities in response patterns of individuals within each of these latent constructs.

In addition to factor extraction method, there are several options for rotation methods, which can generally be classified as being either oblique or orthogonal. Varimax is a common orthogonal rotation where the factors are kept at right angles (i.e., uncorrelated) to one another. On the other hand, promax is an oblique rotation that begins as an orthogonal varimax solution but relaxes the orthogonality such that the factors are not required to be uncorrelated. This method has the ability of providing a middle ground between methods because if there truly is no correlation between the factors, the rotation will still be relatively orthogonal.

The overview of the various methodological approaches above serves to highlight the incongruity between the type of factor analytic approach frequently used and the question being answered by this approach. Table 1 shows that almost all existing studies have used PCA with varimax rotation. It stands to reason, however, that obsessive-compulsive symptom dimensions—at least when expressed as items on the Y-BOCS—will exhibit a latent structure. Additionally, obsessive-compulsive symptoms might be expected to have some degree of correlation amongst their latent factors. Therefore, although many datasets will yield similar results following any combination of factor analytic techniques, principal axis factoring with promax rotation is likely the best reflection of reality for these data.

In addition to choice of extraction and rotation method, there are certain diagnostic audits which should be carried out prior to performing factor analysis of a dataset. Failure to follow these steps may have significant implications for the viability of the analysis and the validity of findings. Factor analysis begins with the production of a correlation matrix and, particularly in IBM® SPSS, the values represented are Pearson's r correlations. However, Pearson's r is not an appropriate correlation method for the use of dichotomous data (Babakus, 1985), such as those in the Y-BOCS symptom checklist and frequently used statistical software (e.g., SPSS) is not capable of producing tetrachoric or polychoric correlations (although there is an available plugin to use the open source program R with SPSS; see Basto and Pereira, 2012). Though not always the case, there can be a vast difference between Pearson's r and tetrachoric or polychoric correlations (Olsson, 1979).

After performing the appropriate correlations, there are several important values that should be inspected prior to moving forward with factor extraction: the Kaiser-Meyer-Olkin measures of sampling adequacy (MSA), both for the whole model and for each variable, and the communality values for each variable. The MSA is a measure of the proportion of variance among input variables that might be common variance, so a value close to zero means that there are large partial correlations compared to the sum of correlations which can be a problem for factor analysis. Communalities, however, are squared multiple correlations and provide a measure of how an item correlates with all

other items. A stringent cut-off for individual variable MSA values is ≥0.70 while ≥0.50 is appropriate for communality values (Norman and Streiner, 2008). Kaiser (1970) suggests that any value >0.60 for the overall model is "miserable" or "unacceptable" and any dataset providing this value, after removing individual variables with values below the cut-offs mentioned above, is likely not suitable for factor analysis. Only two studies of all those included in Table 1 report an MSA (Denys et al., 2004a; KMO = 0.74; and Asadi et al., 2016; KMO = 0.808), and most do not report communalities. Finally, although statisticians debate the exact number, factor analysis is generally most valid with a participant-to-variable ratio of at least five-to-one, if not ten-to-one, so studies with smaller samples should be interpreted with caution. In some cases, a large enough sample can help to address lower communalities although values ideally should still not fall below 0.50 (MacCallum et al., 1999). While these issues may not be problematic in all cases, they raise some concern about the interpretability of previous results from factor analytic investigations of the Y-BOCS.

Some of the methodological limitations of factor analysis when attempting to examine latent constructs of the Y-BOCS have led researchers to consider other approaches. Calamari et al. (1999) were the first to suggest that cluster analysis might be a more appropriate method because, although factor analysis is ideal for determining underlying structure in a dataset of *variables*, cluster analysis can be used to group *cases* to find smaller groups that are representative of data as a whole. Furthermore, factor analysis has no way of ensuring that one individual's responses are not partitioned across several factors, whereas cluster analysis can provide homogeneous groups. Cluster analysis, too, has many possible clustering techniques. The most prevalent of these are hierarchical cluster analysis, which organizes observations in a hierarchical manner based on cluster similarity (or dissimilarity); and *k*-means clustering which requires the input of a pre-specified number of clusters and attempts to fit observations to those clusters. Hasanpour et al. (2017) conducted a study comparing a multitude of clustering strategies with Y-BOCS data, and found that no one strategy stood out as significantly better than another for defining symptom clusters.

Although there is some apparent overlap in the results presented in Table 1, at least insofar as the top three typical symptom dimensions are concerned (i.e., washing, checking and symmetry), the differences in methodologies between studies raise doubts about the validity of these findings. Foremost, while the methodological differences noted between studies are slight, there are corresponding minor but important differences between the results of each study. An example of this can be seen by comparing the symptom categories that fall under the aggressive/checking dimension. In some cases this dimension includes only those two symptom types, while other cases might see the inclusion of sexual, religious, repeating, counting or somatic symptoms in any combination. These differences, though small, are not trivial and may be partially, if not completely explained by differences in coding or analysis methodology. It indeed remains a possibility that the general overlap observed in the results across these studies is largely due to chance. It is, therefore, crucial that researchers wishing to pursue this endeavour adopt a consistent and appropriate methodology that allows replication in order to produce valid findings.

Given the consistent interest in the structure of obsessive-compulsive symptoms for the past two decades and the aforementioned potential methodological challenges, the objectives of the present study were twofold. With data from a large clinical sample of individuals with OCD, this study aimed to perform 1) a factor analysis using best-practice strategies, that is, principal axis factoring with promax rotation based on tetrachoric correlations, and compare results against previous findings, as well as 2) a cluster analysis, and investigate differences in the outcomes of each method, both with the aim of illustrating the differences inherent in the methods and thus to help to elucidate the most accurate choices for characterizing OCD symptom dimensions to

Table 2
Sample demographics.

Variable	Frequency
Age, mean (SD)	33.51 (12.02)
Sex (% female)	55.7%
Age of OCD onset, mean (SD)	17.83 (9.74)
Comorbidity (≥ 1 secondary Axis I diagnosis)	77.9%
Y-BOCS Severity Score	24.17 (5.20)
Mean number of current symptoms endorsed on the Y-BOCS checklist	12.50 (6.67)

Note: Sample size $N = 355$.

improve consistency in future research.

2. Methods

2.1. Participants

Data were collected from $N = 355$ participants aged 18–65 with a primary DSM-IV diagnosis of OCD who were referred for assessment and treatment to a large outpatient anxiety disorders clinic located in an academic community hospital. All participants received a diagnostic assessment using the Structured Clinical Interview for Axis I Disorders for DSM-IV (SCID-IV; First et al., 1995) or by psychiatric consult by a physician. Those with a confirmed principal diagnosis of OCD then received the clinician-administered Y-BOCS and also completed a package of self-report questionnaires. The clinic received institutional ethics approval for an ongoing database for individuals assessed at the clinic. The data reported on in this study were collected between 2003 and 2010. Demographics for the sample are presented in Table 2. The mean age at the time of assessment was 33.51 (SD = 12.02) and the mean Y-BOCS severity score was 24.12 (SD = 5.26).

2.2. Measures

Yale-Brown Obsessive Compulsive Scale (Y-BOCS) (Goodman et al., 1989a, b). The Y-BOCS is a standardized scale with two parts. The first part is a 74-item checklist of symptoms covering seven rationally derived categories of obsessions and six categories of compulsions, with each also having a respective additional miscellaneous category. The 74-item version also provides one open-ended question ("other" symptom) per category for both obsessions and compulsions. Most previous studies excluded the miscellaneous categories and the 10 open-ended questions to increase validity of results and reduce heterogeneity due to the large variability in potential answers. We also chose to exclude hoarding items from this analysis as these data were collected prior to the release of the DSM-5. The Y-BOCS part two is a standardized severity scale with 10 items pertaining to obsessions and compulsions on a 5-point adjectival scale ranging from 0 (no symptoms) to 4 (severe symptoms), and was only used for baseline characteristics in this sample. The clinician-administered version was used for this study in an attempt to increase validity of the data as the two versions may have only moderate convergence in some samples (Federici et al., 2010).

2.3. Procedure

Y-BOCS-SC responses were coded according to Mataix-Cols et al. (1999) adapted method from Baer (1994). As described above, a score of 0, 1 or 2 is assigned to each symptom category. A score of 2 was assigned to a category containing at least one item listed as present *and* as one of an individual's most upsetting obsessions or compulsions, 1 if the symptoms in that category were present only, and 0 if the category contained only absent symptoms. Lifetime symptoms were excluded from the present study to eliminate any potential recall bias.

2.4. Data analyses

2.4.1. Factor analysis

Factor analysis was completed with the computer program RStudio, version 3.3.2. First, correlations were computed using the "polycor" package. The factor analysis was accompanied by a modified version of Glorfeld's parallel analysis and Velicer's minimum average partials (MAP) test to determine the number of factors to extract (Glorfeld, 1995; O'Connor, 2000; Velicer 1976). Principal axis factoring followed by promax rotation was then performed on the tetrachoric correlation matrix; both analyses were performed using the "paramap" package.

2.4.2. Cluster analysis

Given the finding from Hasanpour et al. (2017) that clustering methods are essentially comparable when attempting to cluster Y-BOCS data we chose to apply the method used in Calamari et al. (1999; 2004) of hierarchical cluster analysis using Ward's agglomerative procedure with squared Euclidean distance as the similarity measure. Calamari et al. (2004) followed their hierarchical cluster analysis with *k*-means clustering which, they noted, is sometimes used to address the limitations of hierarchical clustering alone (Borgen and Barnett, 1987; Milligan and Sokal, 1980) but reported that this did not result in any significant improvement in interpretability of the clusters. We, therefore, decided to perform hierarchical clustering only. Cluster analysis was completed using the "fpc" package for RStudio version 3.3.2.

3. Results

Factor analysis of the full eleven Y-BOCS categories revealed many communality and MSA values below the recommended thresholds. The sexual, religious and somatic obsessions, and repeating and counting compulsions categories were removed from the analysis, and the factor analysis was then performed with the remaining six variables. A stable three-factor structure was found explaining 83.2% of the variance, with an overall KMO of 0.707 and a significant Bartlett's test of Sphericity ($\chi^2 = 743.1$, $p < .001$). Results of the parallel analysis indicated that three factors should be retained, with eigenvalues of 1.29, 1.21, and 1.15 for factors one through three, respectively. Similarly the MAP test indicated three factors for sample size of $N = 355$ and $k = 11$ variables, where eigenvalues must be 2.23, 1.27 and 1.13 for factors one through three, respectively, using the 95th percentile and 1000 replications. The factor loadings, with communalities and MSAs for each variable are found in Table 3. The first factor, symmetry obsessions and ordering compulsions, yielded an eigenvalue of 2.56 and accounted for 42.7% of the variance. Contamination obsessions and cleaning compulsions (eigenvalue = 1.23) accounted for 20.6% of the variance, while aggressive obsessions and checking compulsions (eigenvalue = 1.20) accounted for a further 19.9%.

Cluster solutions for $n = 2$ to $n = 11$ clusters were evaluated and solutions beyond four did not improve the average silhouette or cluster interpretability. An average silhouette value of 0.30 ("fair") for the four cluster model was the highest achieved for any solution. Fig. 1 shows

Table 3
Factor loadings for principal axis factoring with promax rotation.

Variable	Factor 1	Factor 2	Factor 3	Communality	MSA
Aggressive Obsessions	–	–	.734	.547	.664
Contamination Obsessions	–	.814	–	.672	.697
Symmetry Obsessions	.826	–	–	.684	.693
Cleaning Compulsions	–	.869	–	.764	.708
Checking Compulsions	–	–	.709	.522	.773
Ordering Compulsions	.896	–	–	.803	.692

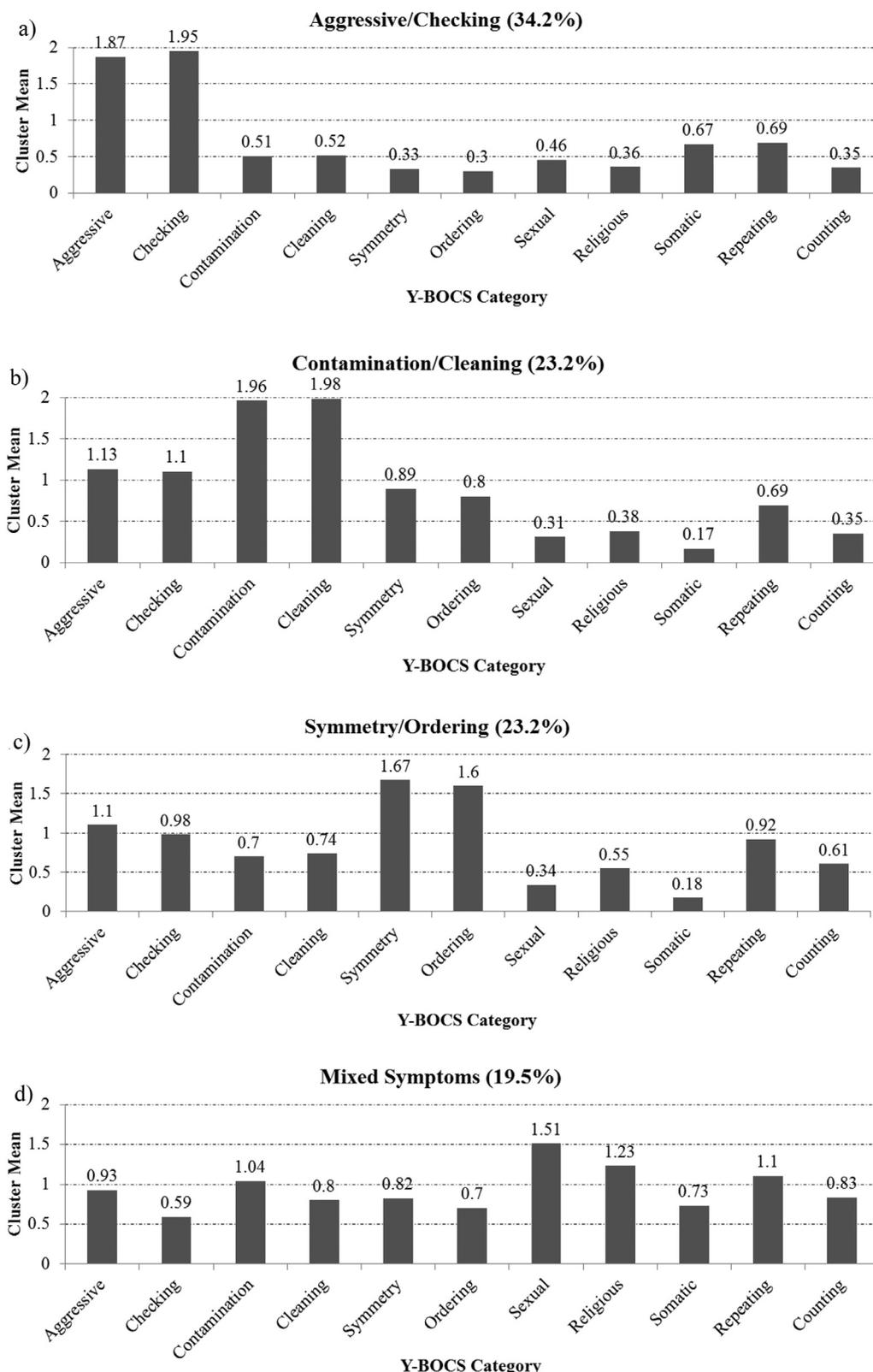Legend: MSA = Measures of sampling adequacy for individual variables.

Fig. 1. Cluster membership.

the mean values for each Y-BOCS symptom category in each of the four clusters. Cluster one, aggressive obsessions and checking compulsions, accounted for 34.2% of the sample. Clusters two and three, contamination/cleaning, and symmetry/ordering, accounted for 23.2% of the sample each. A fourth cluster representing mixed symptoms across almost all categories accounted for the final 19.5% of individuals.

4. Discussion

The purpose of this study was to employ the current best practices in factor analysis with ordinal data on responses to the Y-BOCS-SC, and compare the results of this analysis with the output from cluster analysis. Factor analysis using principal axis factoring and promax rotation

Table 4
Representation of Y-BOCS symptoms as primary or secondary.

Symptom domain	Frequency (%)		
	0	1	2
Aggressive	14.6	45.4	40.0
Checking	11.0	47.3	41.7
Contamination	23.7	42.8	33.5
Cleaning	21.7	43.9	34.4
Symmetry	34.6	48.2	17.2
Ordering	37.5	46.8	15.8
Sexual	69.6	28.2	2.3
Religious	64.2	31.8	3.9
Somatic	76.6	18.9	4.5
Repeating	38.3	54.1	7.6
Counting	65.6	33.2	1.1

Note: 0 = symptom absent; 1 = symptom present; 2 = symptom present and listed as one of currently most troubling.

resulted in three factors including symmetry/ordering, contamination/cleaning and aggressive/checking symptoms. These three factors are consistent with the majority of previous factor analytic studies attempting to define symptom dimensions of OCD using the Y-BOCS symptom checklist. Where this study differs from previous approaches is in the inclusion of the remaining Y-BOCS symptom categories, which all exhibited very low communality values. Because communalities represent the squared multiple correlations, the low values for these variables reflects the fact that these symptom categories are not strongly related to each other and are poorly represented in the final factor solution. As can be seen in Table 3, the communality and MSA values for the retained variables are barely within the recommended range for factor analysis. This finding is arguably more important than the resulting factor solution. What we have learned from a rigorous investigation of factor analysis of the Y-BOCS symptom checklist is that the data, at least, when coded in such a way that multi-item data are reduced to a single, ordinal metric, likely do not lend themselves well to this type of analysis, casting doubt on the interpretability of previous findings.

Table 4 shows the frequency of symptoms endorsed in our sample, which provides some insight into the nature of these results. It can easily be seen that there were considerably more individuals ranking symptoms of aggressive/checking, contamination/cleaning, and symmetry/ordering categories as present and/or most troubling compared to all other symptom categories, which reflects the nature of the presentation of OCD. These limitations, however, do not necessarily stop computerized statistical software from continuing with the analysis and defining what seem to be meaningful factors.

To demonstrate this, results of a factor analysis performed without excluding variables with low communalities or MSA values, and without Velicer's MAP test and parallel analysis indicating the recommended number of factors is presented in Table 5. The resulting

four factor model presented here shows all symptom checklist categories, except counting compulsions, loading significantly onto at least one factor with a value of > 0.30. The factors representing dimensions with symmetry/ordering/repeating, contamination/cleaning, aggressive/checking, and sexual/religious/somatic/repeating symptoms appears similar to those presented in previous studies, when excluding hoarding symptoms (e.g., Cavallini et al., 2002; Kim et al., 2005; Mataix-Cols et al., 1999; Pinto et al., 2007).

Cluster analysis, on the other hand, may offer a slightly different insight into symptom structure, at least when using Y-BOCS-SC responses as input variables. The low representations of less frequently endorsed Y-BOCS items can still be observed by the dominance of the aggressive/checking, contamination/cleaning and symmetry/ordering categories in the largest three clusters, as well as by the relatively low mean scores for all other categories within each cluster. Not only is this affected by certain symptoms being less frequently endorsed overall, but further by the fact that even when these symptoms are endorsed, they are very rarely listed as a respondent's most troubling symptom (see Table 4) . However, the results yielded by this method differ from factor analysis in the fourth cluster, which appears to represent individuals who tend to express symptoms from a variety of different obsessive-compulsive categories with no single type of symptom being more prevalent than the next. This is a similar finding to previous cluster analytic studies, particularly those of Calamari et al. (1999; 2004) who defined "obsessional" and "certainty" subgroups which displayed similar trends to the fourth, mixed cluster described here.

This study was subject to the same limitations of all studies attempting to define symptom dimensions based on Y-BOCS symptom categories, namely (1) attempting to reduce dichotomous multi-item data to a single ordinal metric and (2) defining symptom dimensions based on a priori rather than empirically derived categories. Furthermore, although our sample size was large relative to many other investigations of this kind, it was insufficient for item-level factor analysis results to be valid. Despite a large clinical sample, more meaningful findings might result from having the less frequently endorsed symptoms (such as somatic or religious obsessions and counting or repeating compulsions) better represented in the study sample. Finally, there will always be a conceptual limitation of subtyping based on overt symptom only, and this is described in greater detail below.

The limitations of this study, however, also represent the most important implications of this study in how they should be used to inform future research. First, it is apparent that factor analysis might not be capable of defining symptom dimensions of OCD based on responses recorded on the Y-BOCS symptom checklist. It appears superficially as though cluster analysis, then, might be the more appropriate method for this research question. However, given that the goal is to define taxonomy through latent constructs of symptoms, a method which seeks to group cases rather than symptoms does not adequately address the question. If researchers choose to follow this route, an alternative to cluster analysis, Gaussian mixture modelling, holds the benefit of

Table 5
Factor loadings for principal axis factoring with varimax rotation without removing variables due to low communality or MSA, no MAP test or parallel analysis.

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Communality	MSA
Aggressive Obsessions	.193	.189	**.870**	**.456**	.758	.694
Contamination Obsessions	.234	**.885**	.198	**.309**	.795	.617
Symmetry Obsessions	**.886**	.282	.209	.259	.800	.655
Cleaning Compulsions	**.359**	**.802**	.100	**.353**	.660	.642
Checking Compulsions	**.309**	**.373**	**.580**	**.606**	.472	.794
Ordering Compulsions	**.834**	**.333**	.158	.250	.700	.664
Sexual Obsessions	.115	.026	.267	**.448**	.209	.696
Religious Obsessions	.103	.167	**.363**	**.520**	.291	.776
Somatic Obsessions	.192	.183	.218	**.444**	.199	.818
Repeating Compulsions	**.467**	.273	.253	**.508**	.343	.878
Counting Compulsions	.281	.073	.174	.244	.108	.803

Legend: MAP = Velicer's Minimum Average Partials test; MSA = Measure of sampling adequacy for individual variables.

providing a maximum likelihood model of the data. Maximum likelihood factor analysis also carries this benefit, but is only appropriate if the data do not deviate too far from normality, which, in the case of present study—and possibly in many samples employing the Y-BOCS symptom checklist—they did.

If factor analysis is chosen, researchers should take care to report communality and MSA values for individual variables and assure adequate sample size to maximize validity and interpretability of results. Though not necessitated in the present study due to the small number of factors and the strong loadings by only two variables on each factor, researchers might also choose, in future, to estimate significance of factor loadings for each variable. Stevens (2001) recommends an approach whereby a critical value for the 1% significance level of a correlation is retrieved, doubled, and is then divided by $\sqrt{(N-2)}$, yielding a cut-off value above which factor loadings may be deemed "significant." This method increases validity when one is uncertain of whether or not to include variables as a part of any factor, rather than choosing those that fall above a certain arbitrary value (0.30, for example). Furthermore, these two statistical methods are predicated upon the assumption that factor or cluster solutions of the Y-BOCS are equal to structure of OCD. It is probable that the Y-BOCS itself, despite being the most comprehensive symptom checklist available for OCD, cannot provide the data necessary for this task, and this is possibly the greatest implication of the results described in this study.

Perhaps more important than the limitations of the Y-BOCS are those imposed by the practice of subtyping based on overt symptom. Among the most prominent issues in doing this is that individuals with OCD might not present with one symptom dimension only, effectively eliminating the utility of such a classification. It has been proposed that a more etiologically valid method for defining symptom dimensions of OCD is to group based on symptom theme, or the underlying motivation behind symptoms. One possible model is seen in Summerfeldt et al. (2014), which describes a division of symptoms into either harm avoidance or incompleteness categories (or, "core dimensions"). Though research investigating this model is in its infancy, schemes such as this warrant further exploration.

One possible method for helping to elucidate symptom dimensions of OCD is to include other taxometric methods, some of which include age of onset (early vs. late, e.g. Taylor, 2011), comorbidities (such as obsessive-compulsive personality disorder, e.g. Coles et al., 2008), or neural correlates (imaging data or neuropsychological task performance). Taylor's review of early vs. late onset (2011) found that 7/13 of the Y-BOCS symptoms categories were more likely to occur in early onset OCD, and heritability studies report gene effects associated with specific symptoms (Grados et al., 2003) as well as sibling interactions (Katerberg et al., 2010; Pinto et al., 2008), suggesting link between symptom dimensions and other distinguishing features in OCD. Perhaps the most promising route is the use of neuropsychological functioning—as a cheaper alternative to structural or functional imaging—to delineate symptom dimensions. Two small meta-analyses, targeting washing and checking symptoms (Leopold and Backenstrass, 2015) and symmetry and obsessing symptoms (Bragdon et al., 2018), have summarized the majority of these results showing moderate effect size differences between symptom profiles. However, any attempt to assess the relationship between symptom dimensions and other subtyping strategies will be severely limited until more consistent methods for categorizing symptom dimensions have been adopted.

For those wishing to assess severity based on overt symptom dimensions of OCD in clinical or research settings, the Dimensional Obsessive-Compulsive Scale (DOCS; Abramowitz et al., 2010) was developed with empirically supported, rather than rationally derived symptom categories. Another alternative is the Dimensional Yale-Brown Obsessive-Compulsive Scale (DY-BOCS; Rosario-Campos et al., 2006), which also holds the benefit of being clinician administered to increase validity of symptom ratings given the discordance between clinician-administered and self-report OCD measures (Storch et al.,

2017). Even if the statistical techniques described in this paper were without flaw, any results of an attempt to define symptom dimensions using the Y-BOCS could be rebutted given that this method relies on a priori symptom categories. This is exemplified by the item-level investigations which show that, while close to the results of the categorical approaches, several items do not match their a priori category designation. For example, previous investigations find that when analyzed at the item-level, fear of aggressive impulse items are placed in a dimension comprising "taboo thoughts" while items of excessive responsibility for harm load onto a "doubting/checking" dimension (Denys et al., 2004b; Pinto et al., 2008; Summerfeldt et al., 2004), whereas these items are lumped into a combined aggressive/sexual/religious/somatic/checking dimension when using a categorical approach, as described in Table 5. It is recommended that researchers also collect symptom level data wherever possible as this will help inform future investigations of this nature. However, future attempts to draw meaningful symptom dimensions of OCD using the Y-BOCS symptom checklist should seek to use item-level analysis only, or they risk losing potentially critical detail to rationally derived but not empirically driven categories.

Declaration of Competing Interest

The authors do not have any conflicts to disclose.

Appendix I

XXX.

References

Abramowitz, J.S., Deacon, B.J., Olatunji, B.O., Wheaton, M.G., Berman, N.C., Losardo, D., ... & Björgvinsson, T., 2010. Assessment of obsessive-compulsive symptom dimensions: development and evaluation of the Dimensional Obsessive-Compulsive Scale. Psychol. Assess. 22 (1), 180.
American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). American Psychiatric Pub.
Asadi, S., Daraeian, A., Rahmani, B., Kargari, A., Ahmadiani, A., Shams, J., 2016. Exploring yale-brown obsessive-compulsive scale symptom structure in Iranian OCD patients using item-based factor analysis. Psychiatry Res. 245, 416–422.
Babakus, E., 1985. The sensitivity of maximum likelihood factor analysis given violations of interval scale and multivariate normality. The University of Alabama.
Baer, L., 1994. Factor analysis of symptom subtypes of obsessive compulsive disorder and their relation to personality and tic disorders. J. Clin. Psychiatry.
Basto, M., Pereira, J.M., 2012. An SPSS R-menu for ordinal factor analysis. J. Stat. Softw. 46 (4), 1–29.
Bernstein, G.A., Victor, A.M., Nelson, P.M., Lee, S.S., 2013. Pediatric obsessive–compulsive disorder: symptom patterns and confirmatory factor analysis. J. Obsessive-Compulsive Relat. Disord. 2 (3), 299–305.
Bloch, M.H., Landeros-Weisenberger, A., Rosario, M.C., Pittenger, C., Leckman, J.F., 2008. Meta-analysis of the symptom structure of obsessive-compulsive disorder. Am. J. Psychiatry 165 (12), 1532–1542.
Borgen, F.H., Barnett, D.C., 1987. Applying cluster analysis in counseling psychology research. J. Couns. Psychol. 34 (4), 456.
Bragdon, L.B., Gibb, B.E., Coles, M.E., 2018. Does neuropsychological performance in OCD relate to different symptoms? A meta-analysis comparing the symmetry and obsessing dimensions. Depression Anxiety 35 (8), 761–774.
Calamari, J.E., Wiegartz, P.S., Janeck, A.S., 1999. Obsessive–compulsive disorder subgroups: a symptom-based clustering approach. Behav. Res. Ther. 37 (2), 113–125.
Calamari, J.E., Wiegartz, P.S., Riemann, B.C., Cohen, R.J., Greer, A., Jacobi, D.M., ... & Carmin, C., 2004. Obsessive–compulsive disorder subtypes: an attempted replication and extension of a symptom-based taxonomy. Behav. Res. Ther. 42 (6), 647–670.
Cavallini, M.C., Di Bella, D., Siliprandi, F., Malchiodi, F., Bellodi, L., 2002. Exploratory factor analysis of obsessive-compulsive patients and association with 5-HTTLPR polymorphism. Am. J. Med. Genet. Part A 114 (3), 347–353.
Coles, M.E., Pinto, A., Mancebo, M.C., Rasmussen, S.A., Eisen, J.L., 2008. OCD with comorbid OCPD: a subtype of OCD? J. Psychiatr. Res. 42 (4), 289–296.
Cullen, B., Brown, C.H., Riddle, M.A., Grados, M., Bienvenu, O.J., Hoehn-Saric, R., ... & Nestadt, G., 2007. Factor analysis of the Yale–Brown Obsessive Compulsive Scale in a family study of obsessive–compulsive disorder. Depression Anxiety 24 (2), 130–138.
Delorme, R., Bille, A., Betancur, C., Mathieu, F., Chabane, N., Mouren-Simeoni, M.C., Leboyer, M., 2006. Exploratory analysis of obsessive compulsive symptom dimensions in children and adolescents: a prospective follow-up study. BMC Psychiatry 6 (1), 1.
Delucchi, K.L., Katerberg, H., Stewart, S.E., Denys, D.A., Lochner, C., Stack, D.E., ... & Cath, D.C., 2011. Latent class analysis of the Yale-Brown Obsessive-Compulsive Scale

symptoms in obsessive-compulsive disorder. Comprehensive Psychiatry 52 (3), 334–341.

Denys, D., de Geus, F., van Megen, H.J., Westenberg, H.G., 2004a. Use of factor analysis to detect potential phenotypes in obsessive-compulsive disorder. Psychiatry Res. 128 (3), 273–280.

Denys, D., de Geus, F., van Megen, H.J., Westenberg, H.G., 2004b. Symptom dimensions in obsessive-compulsive disorder: factor analysis on a clinician-rated scale and a self-report measure. Psychopathology 37 (4), 181–189.

Federici, A., Summerfeldt, L.J., Harrington, J.L., McCabe, R.E., Purdon, C.L., Rowa, K., Antony, M.M., 2010. Consistency between self-report and clinician-administered versions of the Yale-Brown Obsessive–Compulsive Scale. J. Anxiety Disord. 24 (7), 729–733.

Feinstein, S.B., Fallon, B.A., Petkova, E., Liebowitz, M.R., 2003. Item-by-item factor analysis of the Yale-Brown obsessive compulsive scale symptom checklist. J. Neuropsychiatry Clin. Neurosci. 15 (2), 187–193.

First, M.B., Spitzer, R.L., Gibbon, M.W.J.B., Williams, J.B., 1995. Structured Clinical Interview For DSM-IV Axis I disorders. State Psychiatric Institute, New York: New York.

Glorfeld, L.W., 1995. An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain. Educ. Psychol. Measur. 55 (3), 377–393.

Goodman, W.K., Price, L.H., Rasmussen, S.A., Mazure, C., Delgado, P., Heninger, G.R., Charney, D.S., 1989a. The Y-BOCS, II: validity. Arch. Gen. Psychiatry 46, 1012–1019.

Goodman, W.K., Price, L.H., Rasmussen, S.A., Mazure, C., Fleischmann, R.L., Hill, C.L., ... & Charney, D.S., 1989b. Yale-brown obsessive compulsive scale (Y-BOCS). Arch. Gen. Psychiatry 46, 1006–1011.

Grados, M.A., Walkup, J., Walford, S., 2003. Genetics of obsessive-compulsive disorders: new findings and challenges. Brain Dev. 25, S55–S61.

Hantouche, E.G., Lancrenon, S., 1996. Modern typology of symptoms and obsessive-compulsive syndromes: results of a large French study of 615 patients. L'Encephale 22, 9–21.

Hasanpour, H., Asadi, S., Meibodi, R.G., Daraeian, A., Ahmadiani, A., Shams, J., Navi, K., 2017. A critical appraisal of heterogeneity in Obsessive-Compulsive Disorder using symptom-based clustering analysis. Asian J. Psychiatry 28, 89–96.

Hasler, G., Pinto, A., Greenberg, B.D., Samuels, J., Fyer, A.J., Pauls, D., ... & Rauch, S.L., 2007. Familiality of factor analysis-derived YBOCS dimensions in OCD-affected sibling pairs from the OCD Collaborative Genetics Study. Biol. Psychiatry 61 (5), 617–625.

Kaiser, H.F., 1970. A second generation little jiffy. Psychometrika 35 (4), 401–415.

Katerberg, H., Delucchi, K.L., Stewart, S.E., Lochner, C., Denys, D.A., Stack, D.E., ... & den Boer, J.A., 2010. Symptom dimensions in OCD: item-level factor analysis and heritability estimates. Behav. Genet. 40 (4), 505–517.

Kim, S.J., Lee, H.S., Kim, C.H., 2005. Obsessive-compulsive disorder, factor-analyzed symptom dimensions and serotonin transporter polymorphism. Neuropsychobiology 52 (4), 176–182.

Leckman, J.F., Grice, D.E., Boardman, J., Zhang, H., Vitale, A., Bondi, C., ... & Goodman, W.K., 1997. Symptoms of obsessive-compulsive disorder. Am. J. Psychiatry 154 (7), 911–917.

Leopold, R., Backenstrass, M., 2015. Neuropsychological differences between obsessive-compulsive washers and checkers: a systematic review and meta-analysis. J. Anxiety Disord. 30, 48–58.

Lochner, C., Hemmings, S.M., Kinnear, C.J., Nel, D., Seedat, S., Moolman-Smook, J.C., Stein, D.J., 2008. Cluster analysis of obsessive-compulsive symptomatology: identifying obsessive-compulsive disorder subtypes. Isr. J. Psychiatry Relat. Sci. 45 (3), 164.

MacCallum, R.C., Widaman, K.F., Zhang, S., Hong, S., 1999. Sample size in factor analysis. Psychol. Methods 4 (1), 84.

Mataix-Cols, D., Marks, I.M., Greist, J.H., Kobak, K.A., Baer, L., 2002. Obsessive-compulsive symptom dimensions as predictors of compliance with and response to behaviour therapy: results from a controlled trial. Psychother. Psychosomatics 71 (5), 255–262.

Mataix-Cols, D., Nakatani, E., Micali, N., Heyman, I., 2008. Structure of obsessive-compulsive symptoms in pediatric OCD. J. Am. Acad. Child Adolesc. Psychiatry 47 (7), 773–778.

Mataix-Cols, D., Rauch, S.L., Manzo, P.A., Jenike, M.A., Baer, L., 1999. Use of factor-analyzed symptom dimensions to predict outcome with serotonin reuptake inhibitors and placebo in the treatment of obsessive-compulsive disorder. Am. J. Psychiatry 156 (9), 1409–1416.

Matsunaga, H., Maebayashi, K., Hayashida, K., Okino, K., Matsui, T., Iketani, T., ... & Stein, D.J., 2008. Symptom structure in Japanese patients with obsessive-compulsive disorder. Am. J. Psychiatry 165 (2), 251–253.

McKay, D., Abramowitz, J.S., Calamari, J.E., Kyrios, M., Radomsky, A., Sookman, D., ... & Wilhelm, S., 2004. A critical evaluation of obsessive–compulsive disorder subtypes: symptoms versus mechanisms. Clin. Psychol. Rev. 24 (3), 283–313.

McKay, D., Piacentini, J., Greisberg, S., Graae, F., Jaffer, M., Miller, J., 2006. The structure of childhood obsessions and compulsions: dimensions in an outpatient sample. Behav. Res. Ther. 44 (1), 137–146.

Milligan, G.W., Sokal, L.M., 1980. A two-stage clustering algorithm with robust recovery characteristics. Educ. Psychol. Measur. 40, 755–759.

Norman, G., Streiner, D., 2008. Principal components and factor analysis. Biostatistics: The Bare Essentials. BC Decker Inc, Hamilton, ON, pp. 208–225.

O'connor, B.P., 2000. SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. Behav. Res. Methods Instrum. Comput. 32 (3), 396–402.

Olsson, U., 1979. Maximum likelihood estimation of the polychoric correlation coefficient. Psychometrika 44 (4), 443–460.

Pinto, A., Eisen, J.L., Mancebo, M.C., Greenberg, B.D., Stout, R.L., Rasmussen, S.A., 2007. Taboo thoughts and doubt/checking: a refinement of the factor structure for obsessive–compulsive disorder symptoms. Psychiatry Res. 151 (3), 255–258.

Pinto, A., Greenberg, B.D., Grados, M.A., Bienvenu, O.J., Samuels, J.F., Murphy, D.L., ... & Pauls, D.L., 2008. Further development of YBOCS dimensions in the OCD Collaborative Genetics study: symptoms vs. categories. Psychiatry Res. 160 (1), 83–93.

Radomsky, A.S., Taylor, S., 2005. Subtyping OCD: prospects and problems. Behav. Ther. 36 (4), 371–379.

Rosario-Campos, M.C., Miguel, E.C., Quatrano, S., Chacon, P., Ferrao, Y., Findley, D., ... & Tolin, D., 2006. The Dimensional Yale–Brown Obsessive–Compulsive Scale (DY-BOCS): an instrument for assessing obsessive–compulsive symptom dimensions. Mol. Psychiatry 11 (5), 495.

Starcevic, V., Brakoulias, V., 2008. Symptom subtypes of obsessive–compulsive disorder: are they relevant for treatment? Aust. N. Z. J. Psychiatry 42 (8), 651–661.

Stevens, J.P., 2001. Applied Multivariate Statistics for the Social Sciences, 4th edition. Lawrence Erlbaum Associates, Mahwah, NJ.

Stewart, S.E., Rosario, M.C., Baer, L., Carter, A.S., Brown, T.A., Scharf, J.M., ... & Rasmussen, S., 2008. Four-factor structure of obsessive-compulsive disorder symptoms in children, adolescents, and adults. J. Am. Acad. Child Adolesc. Psychiatry 47 (7), 763–772.

Stewart, S.E., Rosario, M.C., Brown, T.A., Carter, A.S., Leckman, J.F., Sukhodolsky, D., ... & Pauls, D.L., 2007. Principal components analysis of obsessive–compulsive disorder symptoms in children and adolescents. Biol. psychiatry 61 (3), 285–291.

Storch, E.A., Nadeau, J.M., De Nadai, A.S., Cepeda, S.L., Riemann, B.C., Seibell, P., Kay, B., 2017. Symptom correspondence between clinicians and patients on the Yale–Brown Obsessive Compulsive Scale. Compr. Psychiatry 73, 105–110.

Summerfeldt, L.J., Kloosterman, P.H., Antony, M.M., Richter, M.A., Swinson, R.P., 2004. The relationship between miscellaneous symptoms and major symptom factors in obsessive-compulsive disorder. Behav. Res. Ther. 42 (12), 1453–1467.

Summerfeldt, L.J., Kloosterman, P.H., Antony, M.M., Swinson, R.P., 2014. Examining an obsessive-compulsive core dimensions model: structural validity of harm avoidance and incompleteness. J. Obsessive-Compulsive Related Disord. 3 (2), 83–94.

Summerfeldt, L.J., Richter, M.A., Antony, M.M., Swinson, R.P., 1999. Symptom structure in obsessive-compulsive disorder: a confirmatory factor-analytic study. Behav. Res. Ther. 37 (4), 297–311.

Taylor, S., 2011. Early versus late onset obsessive–compulsive disorder: evidence for distinct subtypes. Clin. Psychol. Rev. 31 (7), 1083–1100.

Tek, C., Ulug, B., 2001. Religiosity and religious obsessions in obsessive–compulsive disorder. Psychiatry Res. 104 (2), 99–108.

Velicer, W.F., 1976. Determining the number of components from the matrix of partial correlations. Psychometrika 41, 321–327.