Progress in Neuropsychopharmacology & Biological Psychiatry 79 (2017) 206–212

Contents lists available at ScienceDirect

Progress in Neuropsychopharmacology & Biological Psychiatry

journal homepage: www.elsevier.com/locate/pnp

Quantifying dimensional severity of obsessive-compulsive disorder for neurobiological research

Roseli G. Shavitt[a,*], Guaraci Requena[b], Pino Alonso[c,g], Gwyneth Zai[d], Daniel L.C. Costa[a], Carlos Alberto de Bragança Pereira[b], Maria Conceição do Rosário[e], Ivanil Morais[a], Leonardo Fontenelle[f], Carolina Cappi[a], James Kennedy[d], Jose M. Menchon[c,g], Euripedes Miguel[a], Peggy M.A. Richter[d]

[a] Department of Psychiatry, University of São Paulo, School of Medicine, Rua Dr. Ovídio Pires de Campo, 785/3º. andar-sala 7. CEP 01060-970 São Paulo, Brazil
[b] Institute of Mathematics and Statistics, University of Sao Paulo, R. do Matão, 1010 - Vila Universitaria, São Paulo, SP CEP 05508-090, Brazil
[c] OCD Clinical and Research Unit, Department of Psychiatry, Hospital de Bellvitge, Barcelona, Spain
[d] Sunnybrook Health Sciences Centre, Centre for Addiction and Mental Health, 2075 Bayview Avenue, Suite FG42, Toronto, ON M4N 3M5, Canada
[e] Child and Adolescent Psychiatry Unit (UPIA), Department of Psychiatry, Federal University of São Paulo (UNIFESP), Rua Borges Lagoa 570, CEP04038-020 São Paulo, Brazil
[f] Instituto de Psiquiatria, Universidade Federal do Rio de Janeiro (UFRJ), Instituto D'Or de Pesquisa e Ensino (IDOR), Rio de Janeiro, Brazil. Av. Venceslau Braz, 71 fundos. Botafogo, Rio de Janeiro, RJ, 22290-140, Brazil
[g] Bellvitge Biomedical Research Institute (IDIBELL), Centro de Investigación en Red de Salud Mental, Carlos III Health Institute, Department of Clinical Sciences, Bellvitge Campus, University of Barcelona, Feixa Llarga, s/n, 08907 L'Hospitalet de Llobregat, Barcelona, Spain

ARTICLE INFO

Keywords:
Obsessive-compulsive disorder
Phenotype
Algorithm
Dimensional assessment

ABSTRACT

Current research to explore genetic susceptibility factors in obsessive-compulsive disorder (OCD) has resulted in the tentative identification of a small number of genes. However, findings have not been readily replicated. It is now broadly accepted that a major limitation to this work is the heterogeneous nature of this disorder, and that an approach incorporating OCD symptom dimensions in a quantitative manner may be more successful in identifying both common as well as dimension-specific vulnerability genetic factors. As most existing genetic datasets did not collect specific dimensional severity ratings, a specific method to reliably extract dimensional ratings from the most widely used severity rating scale, the Yale-Brown Obsessive Compulsive Scale (YBOCS), for OCD is needed. This project aims to develop and validate a novel algorithm to extrapolate specific dimensional symptom severity ratings in OCD from the existing YBOCS for use in genetics and other neurobiological research. To accomplish this goal, we used a large data set comprising adult subjects from three independent sites: the Brazilian OCD Consortium, the Sunnybrook Health Sciences Centre in Toronto, Canada and the Hospital of Bellvitge, in Barcelona, Spain. A multinomial logistic regression was proposed to model and predict the quantitative phenotype [i.e., the severity of each of the five homogeneous symptom dimensions of the Dimensional YBOCS (DYBOCS)] in subjects who have only YBOCS (categorical) data. YBOCS and DYBOCS data obtained from 1183 subjects were used to build the model, which was tested with the leave-one-out cross-validation method. The model's goodness of fit, accepting a deviation of up to three points in the predicted DYBOCS score, varied from 78% (symmetry/order) to 84% (cleaning/contamination and hoarding dimensions). These results suggest that this algorithm may be a valuable tool for extracting dimensional phenotypic data for neurobiological studies in OCD.

1. Introduction

Obsessive-compulsive disorder (OCD) is a complex neurobiological condition characterized by the presence of obsessions, or intrusive, unwanted thoughts, which cannot be suppressed, and compulsions or repetitive behaviors or mental acts (APA, 2013; Shavitt et al., 2014). The etiology of this disorder is complex, with a strong genetic element based on heritability estimates of approximately 40–65%, depending on early versus post-adolescent age of onset (Mataix-Cols et al., 2013; Pauls, 1992, 2010), presence of tics (Pauls et al., 1995), and sporadic or

* Corresponding author at: Rua Dr. Ovídio Pires de Campos, 785/3º. andar-sala 7. CEP 01060-970, São Paulo, Brazil.
E-mail addresses: roseli.gshavitt@hc.fm.usp.br (R.G. Shavitt), mpalonso@bellvitgehospital.cat (P. Alonso), gwyneth.zai@camh.ca (G. Zai), jim.kennedy@camh.ca (J. Kennedy), jmmenchon@bellvitgehospital.cat (J.M. Menchon), peggy.richter@sunnybrook.ca (P.M.A. Richter).

http://dx.doi.org/10.1016/j.pnpbp.2017.06.037
Received 29 September 2016; Received in revised form 27 June 2017; Accepted 29 June 2017
Available online 01 July 2017
0278-5846/

familial form (Pauls et al., 2014). In addition, rare variations (Cappi et al., 2016) and epigenetic factors have been reported as relevant to the clinical manifestations of OCD (Yue et al., 2016). Moreover, there is robust evidence that different symptoms of this disorder may have overlapping but distinct neurobiological substrates corresponding to specific genetic features (Alonso et al., 2011; Cavallini et al., 2002; Hasler et al., 2007; Katerberg et al., 2010a; Kohlrausch et al., 2016; Lennertz et al., 2014; Taj et al., 2013).

Addressing the OCD phenotype for genetic studies has been a challenge for researchers in the field and many studies have emphasized the relevance of using a dimensional approach (Aleman et al., 2016; Lecrubier, 2008; Waszczuk et al., 2017). To this date, most studies have emphasized a four-factor model comprising: (I) repugnant/harm obsessions (i.e. sexual, religious, harm-related, somatic) and checking compulsions; (II) symmetry obsessions and repeating, counting, and ordering compulsions; (III) contamination obsessions and cleaning compulsions; and (IV) hoarding obsessions and compulsions (Bloch et al., 2008). This biological heterogeneity based on primary symptom dimensions has been supported by functional neuroimaging, family history, age of onset, and response to pharmacotherapy (De Luca et al., 2011; Harrison et al., 2013; Jhung et al., 2014; Mataix-Cols et al., 2004; Pertusa et al., 2010; Rosario-Campos et al., 2001; Via et al., 2014). Additional support is conferred from reports of individuals' symptoms staying within the same symptom groups over time (shown in adults and children) (Delorme et al., 2006; Mataix-Cols et al., 2002; Rufer et al., 2005), despite the observation in clinical practice that specific symptom types may change over the course of the disease. Despite the delineation of distinct OCD subgroups obtained by exploratory factor analyses (EFA) of data from the Yale-Brown Obsessive-Compulsive Scale [YBOCS] (Goodman et al., 1989a; Goodman et al., 1989b); (Mataix-Cols et al., 2004; Leckman et al., 1997), one issue that remains unresolved with this methodology is to what extent the severity of each symptom dimension contributes to the observed phenotype. Furthermore, within a given individual, symptoms may coexist from two or more factors simultaneously. Therefore, individuals cannot easily be assigned to one predominant symptom "class". Consequently, there is a clear need for a novel, practical, readily available, and standardized way to quantitatively assess OCD symptoms across the differing dimensions present in a given individual, particularly for any exploration of genetic vulnerability.

Factorial analyses allow for the characterization of the phenotype based on the presence/absence of symptoms pertaining to each category (Schooler et al., 2008; Katerberg et al., 2010b), but not on the contribution of the severity of each symptom dimension to the overall clinical severity. In this way, OCD genetic studies have potentially been hampered by the heterogeneity of this illness, and it has been proposed that analyses based on the quantitative measures of specific symptom dimensions may thus be a powerful way to explore more genetically homogeneous subgroups of OCD. To the best of our knowledge, no study in the OCD field has ever tried to extract dimensional data (i.e., how severe?) from large datasets containing only categorical data (present/absent). The novel approach proposed in the present study aims to enable, for the first time, the determination of the severity of individual symptom dimensions for a better delineation of the OCD phenotypes. The approach based on separating the different types of symptoms is a necessary first step in refining the OCD phenotype, but the need to determine which group of symptoms is more relevant to the observed clinical picture remains unattended. It is becoming increasingly clear that all efforts to investigate the genetic basis of this condition should take a symptom dimensional approach, as the few studies that specifically examined genetic risk support both shared and unique genetic vulnerability across these dimensions (Alonso et al., 2011; Cavallini et al., 2002; Katerberg et al., 2010a; Kohlrausch et al., 2016; Taj et al., 2013; Iervolino et al., 2011).

The phenomenology of OCD can be captured by structured interviews that can be self-reported or clinician administered, like the YBOCS (Goodman et al., 1989a; Goodman et al., 1989b) and the Dimensional YBOCS (DYBOCS) (Rosario-Campos et al., 2006). Although the YBOCS is the most widely used instrument in OCD studies internationally, it does not allow for the collection of quantitative data by symptom type. By contrast, the DYBOCS was developed to assess the presence and severity of five individual symptom dimensions and their respective severity in patients with OCD, plus one miscellaneous dimension comprising symptoms of the OCD-related conditions. The DYBOCS enables determination of the clinical relevance and severity of each symptom dimension, as well as an overall OCD severity rating. This dimensional approach seems particularly pertinent to the biological investigation of a complex condition such as OCD. Thus, broad consensus has emerged in the field for the need to explore OCD not as a homogeneous diagnosis, but rather utilizing quantitative assessments of these symptom factors.

Efforts to elucidate genetic risk factors in OCD have been underway by several international centers. The recent genome-wide association studies that searched for common DNA sequence variations predisposing individuals to OCD have not yielded genome-wide significant results, but these datasets did not historically include any dimensional measures to permit analysis based on symptom dimensions (Mattheisen et al., 2014; Stewart et al., 2013). Future exploration with increased attention to phenotypes, especially the consideration of subtypes of this disorder, could result in greater success (Burmeister et al., 2008). If there was a reliable way to extrapolate valid quantitative dimensional data from the most widely used OCD scale (the YBOCS), the existing international datasets could be re-explored in a more refined and symptom-specific fashion.

The aim of this study was to develop and validate a novel, statistical algorithm for the extraction of quantitative, symptom-dimension specific data for all symptoms in a given individual from the most commonly used OCD rating scale, the YBOCS. We used data from 1183 subjects from three independent international samples. We postulate that this algorithm will allow for a more successful way in identifying the neurobiological underpinnings of OCD, such as genetic vulnerability factors associated with specific OCD symptom dimensions.

2. Methods

This work was done with DYBOCS and YBOCS data obtained from 1183 adult patients with primary OCD, diagnosed according to DSM-IV criteria, from three independent groups: the Brazilian OCD Research Consortium (Miguel et al., 2008) (n = 912), the Anxiety Disorders Clinic at the Centre for Addiction and Mental Health and the Frederick W. Thompson Anxiety Disorders Centre at the Sunnybrook Health Sciences Centre, Canada (n = 36) and the Hospital of Bellvitge, Barcelona, Spain (n = 235). Data from the YBOCS and DYBOCS were obtained by trained clinicians at the same point in time. All the work was developed with de-identified data sets built over the years using data from different research projects approved by the local Ethics Committee at each participant Institution. The funding for this study came from a joint grant of the University of Sao Paulo and University of Toronto, process number 13.1.13252.1.6, 2012.

In order to build an algorithm for extraction of a dimensional severity score from the YBOCS, the first step was to recode DYBOCS data into the YBOCS format for the Brazilian sample, since these subjects had the severity ratings but not the symptom checklist from the YBOCS. Table 1 shows the main features of the YBOCS and DYBOCS.

The YBOCS is more general than the DYBOCS in the characterization of symptoms. For example, YBOCS symptom #64 is "I have mental rituals (other than checking/counting)", whereas the DYBOCS has five symptoms related to mental rituals: "I have mental rituals, other than checking, specifically related to: #30-sexual or religious obsessions; obsessions of symmetry, exactness, or just right perceptions (#41); contamination worries (#53); hoarding obsessions (#60) and somatic worries (#64)". Therefore, if a patient scores 2 (present) at YBOCS

Table 1
Main features of the Yale-Brown Obsessive-Compulsive Scale (YBOCS) and the Dimensional Yale-Brown Obsessive-Compulsive Scale (DYBOCS).

YBOCS	DYBOCS
List of 74 OCD symptoms, divided into 8 categories of obsessions and 7 categories of compulsions, that can be rated as absent (0), or present (1) For a better comparison with the DYBCOS and the purpose of the analyses performed in the current study each item was scored as absent (0), past (1) or present (2)	List of 88 OCD symptoms, that can be rated as absent (0), past (1) or present (2), divided into six symptom dimensions: (1) obsessions/compulsions about harm due to aggression/injury/violence/natural disasters (2) obsessions/compulsions of sexual/moral/religious content (3) obsessions/compulsions concerning symmetry/counting/ordering/arranging (4) contamination obsessions and cleaning compulsions (5) obsessions and compulsions related to hoarding and (6) miscellaneous obsessions and compulsions (including somatic concerns and superstitions)
List of the most severe (target) symptoms (3 to 5 obsessions and 3 to 5 compulsions depending on the version used)	List of 3 target symptoms (obsessions or compulsions)
Global severity scale based on time, interference, distress (ranging from 0 - none - to 4 - extreme), resistance (from 0 - always resists - to 4 - completely yields) and control (from 0 - complete control - to 4 - no control) over obsessions and compulsions	Individual severity scales for each dimension that takes into account time (0–5), distress (0–5) and interference (0–5) caused by symptoms (range 0–15) plus a global severity scale based on symptom severity (time, distress and interference range 0–15) and impairment (range 0–15)

Table 2
Rewriting DYBOCS data in the YBOCS format: an example using DYBOCS symptoms # 30, 41 and 64.

DY-BOCS #30	DY-BOCS #41		Y-BOCS #64
0	0	→	0
0	1	→	1
1	0	→	1
1	1	→	1
0	2	→	2
2	0	→	2
1	2	→	2
2	1	→	2
2	2	→	2

0 = absent; 1 = past; 2 = present.

symptom #64, it is not possible to say they score 2 (present) at DYBOCS #30, #41, #53, #60, or #64. However, if a patient scores 2 (present) at DYBOCS symptom #30, it is possible to say that they score 2 (present) at YBOCS symptom #64. All possible outcomes for scoring are presented in Table 2. There are three additional situations in which one YBOCS symptom is related to more than one DYBOCS symptom: #26-hoarding/saving obsessions (56,55,58); #32-need to know/remember (31,66); #52-checking that will no harm others (Burmeister et al., 2008; Goodman et al., 1989a). Additionally, four YBOCS symptoms have no corresponding items in the DYBOCS (#7-fear will steal things; #19-no concern with consequences of contamination other than how it might feel; #30-obsession with need for symmetry accompanied by magical thinking; #38-bothered by certain sounds/noises). Finally, the questions that address avoidance in the DYBOCS are not captured in the YBOCS and are not represented in the present algorithm.

Proceeding as above, we rewrote the DYBOCS data in the same format as the YBOCS. Of note, a list of target symptoms and a global severity score are available in both the YBOCS and DYBOCS. Thus, the dimensional information present in the rewritten DYBOCS data is presented as in the YBOCS. To extract this information, multinomial logistic regressions have been fitted (for each homogeneous symptom dimension: aggression, sexual/religious, symmetry/ordering, contamination/cleaning and hoarding). The miscellaneous dimension did not receive a model in this proposal due to its heterogeneous content, which is non-informative for the purpose of refining the phenotype for neurobiological research.

The multinomial regressions were built considering the DYBOCS scores, ranging from 0 to 15, as the response variable, multinomially distributed (Table 3).

The predictor variables were: (1) The weighted average number of symptoms in each dimension (never = 0, past = 1, current = 2). In spite of the fact that the miscellaneous dimension was not modeled, the symptoms in the YBOCS checklist identified as belonging to this dimension were used to compose a predictor variable; (2) an indicator variable of the presence of the YBOCS checklist target symptom in the dimension being modeled (yes/no); (3) a YBOCS obsessions severity score composed of the time, interference and distress sub-scores (range 0–12); (4) a YBOCS compulsions severity score composed of the time, interference and distress sub-scores (range 0–12). The YBOCS questions regarding resistance and control of obsessions and compulsions were not included in the model because there are no corresponding questions in the DYBOCS; (5) an indicator variable of the presence of each symptom confirmed in the YBOCS checklist in the dimension being modeled. The multinomial regressions were built avoiding redundancy in the considered predictors. In addition, the stepwise method for variable selection was used in all steps of modeling, further avoiding collinearity.

The multinomial regressions, for each dimension, were built nesting in a tree- structure binomial logistic model (Basu & Pereira, 1982; Pereira & Stern, 2008). Fig. 1 illustrates the tree-structure binomial logistic regression using the aggression dimension as an example.

For each dimension, 15 dichotomizations were performed using a partitioning cut-off given by the respective binomial regression, through the maximization of the accuracy measures: sensitivity, specificity, positive and negative predictive value, obtained by the leave-one-out cross validation. After determining the cut-offs, in each "node" of this tree structure - i.e. in each bifurcation where the "ancestor" subset is partitioned – there is a fitted binomial logistic regression in which the predictors are selected by the stepwise method, avoiding redundancy. Of note, for different binomial regressions, different predictors were selected, taking into consideration the number of available observations. A minimum of 10 observations per predictor was considered, as recommended by Hosmer et al. (2013), in order to enable an adequate sample size to be analyzed.

Once the multinomial regression was fitted, we performed, for each dimension, a leave-one-out cross validation. Each subject was left out of the sample that fitted the binomial regressions. For each binomial regression we calculated an optimal probability cut-off for classification based in ROC curves. Then, we classified the left-out subject sequentially, based on the built tree structure, passing from each respective node ("ancestor" subset) to one of two possibilities ("descendant" subsets), until reaching one of the terminal nodes, which corresponds to one of the possible values of the DYBOCS score (0 to 15).

Table 3
Multinomial distribution of the DYBOCS scores.

DYBOCS dimension	DYBOCS score															Total
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
Aggression	415	8	15	28	33	49	53	62	58	109	77	79	89	47	35	26	
Sexual/religious	627	8	8	41	28	28	41	50	46	70	46	56	71	27	19	17	
Symmetry/ordering	295	8	26	36	37	53	63	55	71	106	94	87	116	66	38	32	
Cleaning/contamination	414	5	8	41	25	37	51	48	58	94	65	78	115	70	47	27	
Hoarding	660	13	40	58	49	38	70	49	41	44	36	23	33	17	9	3	1183

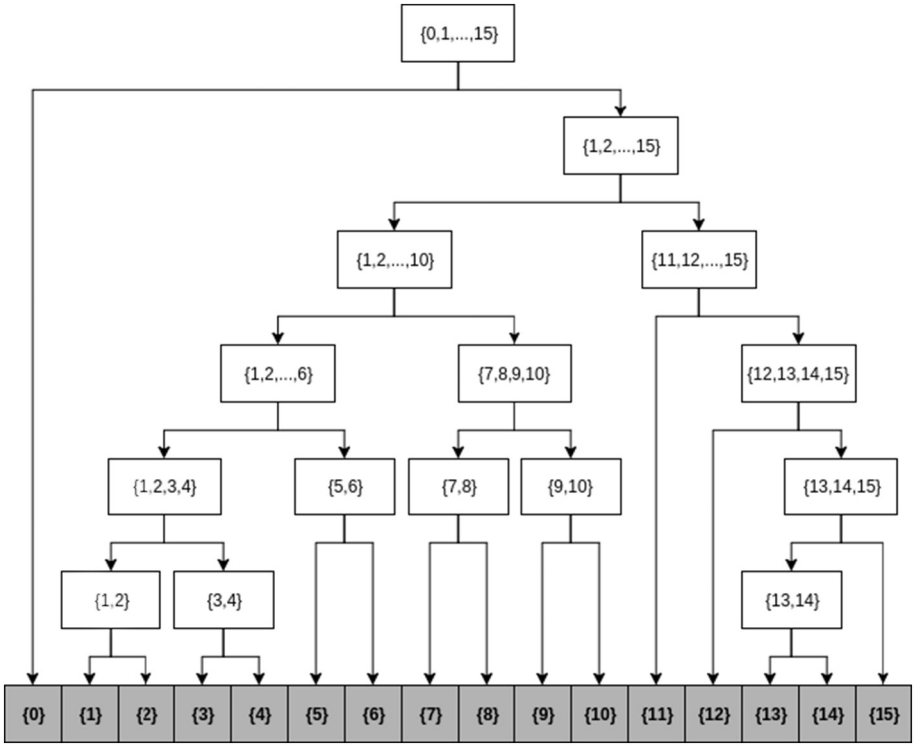

Fig. 1. Binomial regressions structured for the aggression dimension.

The validation was carried out twice: first, the variable “site” was considered as a predictor in the model; next, the validation without considering the subject's site of origin was performed and there were no significant differences in the results of the two analyses. The latter are presented in the Results section, in the format of observed versus predicted cross tables and the respective accuracies. Such accuracies validate the proposed algorithm.

All analyses were conducted in R 3.4.0 (R Core Team, 2014). The package “MASS” was used for performing the stepwise method (Venables & Ripley, 2002).

3. Results

The results of the cross validation for the aggression dimension are shown, as an example, in Table 4. The other dimensions were similarly tested and their results are available in the supplementary material section. The main diagonal line, marked with the darkest shade, has the number of subjects for whom the predicted DYBOCS values were equal to the directly observed DYBOCS values (in this diagonal, the model is 100% right). At both sides of the main diagonal there are diagonals in lighter shades showing the number of subjects for whom the predicted DYBOCS values differed from the observed values by 1, 2, or 3 points. The darker shade is where we would like to see the largest number of subjects.

Using this approach, we would consider a model to have a good predictive capacity if the majority of subjects stayed close to the main diagonal. We calculated the proportions of subjects in which the model predicted correctly (accuracy) for 0, 1, 2, 3 and 4 points of deviation. Considering our sample size and the number of covariates, a deviation of up to 3 points in each direction in the predicted DYBOCS score revealed the best compromise between the predictive capacity of our model and the phenotype to be studied. According to this criterion, 81.91% of subjects stayed in the shaded regions of aggression. Table 5 summarizes the model's goodness of fit for all symptom dimensions.

A package containing the algorithm's equation and the instructions for using the R software is available upon request.

4. Discussion

This study describes the development of a statistical algorithm for

Table 4
Results of the leave-one-out cross validation for the aggression dimension in 1183 patients with Obsessive-Compulsive Disorder.

OBSERVED \ PREDICTED	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	TOTAL
0	398	2	1	2	2	4	1	0	1	1	0	0	0	3	0	0	415
1	0	2	1	1	1	1	0	0	1	0	0	0	1	0	0	0	8
2	1	0	1	2	1	1	2	1	2	0	2	1	0	1	0	0	15
3	3	0	0	4	3	5	5	1	1	2	1	2	0	1	0	0	28
4	1	0	1	5	2	5	3	6	4	0	3	2	0	1	0	0	33
5	0	1	1	5	2	10	5	6	2	3	3	4	3	3	1	0	49
6	1	0	2	1	1	5	7	9	4	4	5	7	4	1	0	2	53
7	1	0	0	5	1	3	9	16	6	9	5	3	2	2	0	0	62
8	0	0	0	4	0	2	8	16	7	8	6	2	2	3	0	0	58
9	2	0	1	6	1	4	8	24	9	8	17	13	6	5	5	0	109
10	1	1	1	2	0	5	3	10	5	10	19	7	7	3	3	0	77
11	0	0	1	0	1	3	6	10	5	11	6	20	6	5	4	1	79
12	0	0	0	2	0	1	2	8	2	8	10	20	25	7	0	4	89
13	0	0	1	0	0	1	2	3	1	5	2	9	10	10	2	1	47
14	0	0	1	0	0	0	1	1	0	1	4	9	8	4	4	2	35
15	0	1	0	0	0	0	0	1	0	1	0	8	3	1	6	5	26
TOTAL	408	7	12	39	15	50	62	112	50	71	83	107	77	50	25	15	1183

Table 5
Goodness of fit of the statistical model for each symptom dimension in 1183 subjects with Obsessive-Compulsive Disorder.

$Y_k - \hat{Y}_k \vee \leq$	0	1	2	3	4
Dimension 1 (Aggression)	45%	61%	74%	82%	88%
Dimension 2 (Sexual/Religious)	58%	66%	74%	81%	86%
Dimension 3 (Symmetry)	37%	54%	67%	78%	85%
Dimension 4 (Contamination)	47%	62%	75%	84%	91%
Dimension 5 (Hoarding)	61%	70%	78%	84%	89%

Yk = observed DYBOCS severity score. $\hat{Y}k$ = respective predicted DYBOCS severity score. The first column (0) shows the proportion of subjects whose predicted values were equal to the observed values. The second, third and fourth columns show, respectively, the proportions of subjects whose observed values had a deviation from the predicted values of 1, 2, 3 or 4 points.

the extraction of quantitative, specific symptom-dimension data from the most commonly used OCD rating scale, the YBOCS. We utilized a large data set derived from three independent international centers, which specialize in OCD, to create and validate the model. The model's goodness of fit, accepting a deviation of up to three points in the predicted DYBOCS score for each individual symptom dimension, varied between 78% (symmetry/order dimension) to 84% (cleaning/contamination and hoarding dimensions). The main objective of this algorithm was to enable the extraction of dimensional data from information obtained with a categorical interview. Such dimensional data would be more suitable for OCD genetic and biological studies. If, despite a deviation of up to three points in the predicted DYBOCS scores, this algorithm allowed us to classify subjects with OCD who differ in terms of their most severe symptom dimensions, we believe it has served the intended purpose.

We could argue that the concordance between predicted and observed values would be expected to be higher for behaviors that are more homogeneous and derived from fewer questions in the checklists (i.e., hoarding and contamination/cleaning) than for dimensions composed of numerous and diverse symptoms (i.e., sexual/religious obsessions and mental rituals, or need for symmetry and ordering behaviors). Nevertheless, the observation of high concordance rates for all dimensions suggests that this algorithm may be used to extract quantitative dimensional data from the YBOCS in order to refine analyses of neurobiological data. Our results also support the relevance of the DYBOCS (Rosario-Campos et al., 2006) as a standard instrument to evaluate OCD symptom dimensions for the purposes of neurobiological research.

This algorithm may prove particularly useful in analyses of genetics data. While a number of international sites have amassed large OCD genetic databases, most of these sites did not collect dimensional data for subjects. Thus, despite recognition of the importance of utilizing a symptom dimension approach, this type of analysis could not have been done currently in these existing datasets. Collection of new and larger datasets can be prohibitively expensive. Our results yield a working statistical algorithm, which could be used in these worldwide available resources for further genetic and neurobiological interrogation. It seems likely that this would translate into clearer identification of susceptibility genes and/or biological markers and substrates for this complex and severe condition.

The objective of our study was to build and test an algorithm that allows the extraction of quantitative information on symptom dimensions from Y-BOCS categorical data. Previous findings suggest that D-YBOCS scores are useful to detect genetic differences between symptom dimensions in OCD patients (Alonso et al., 2012). Still, it is unclear whether this new approach actually yields superior genetic findings than the former YBOCS algorithm method. Therefore, a convergence study between the former YBOCS method and the algorithm proposed in this study is warranted in the near future in order to confirm the utility of our algorithm in genetic studies.

Some limitations of this study merit consideration. The first concerns our sample size. Due to the lack of similar studies conducted so far, we could not perform a classical power calculation. On the other hand, we worked with the largest available sample in the world with data collected with both the DYBOCS and YBOCS, comprising 1183 subjects. In addition, we complied with the recommendation of the literature to run each regression with at least 10 observations per predictor. Second, the proposed algorithm did not predict severity scores to symptoms pertaining to the miscellaneous dimension, resulting in a certain loss of information in the conversion of YBOCS into DYBOCS data. However, we need to consider that this was due to the nature of this study, focused on defining more homogeneous OCD phenotypes for neurobiological research. Importantly, this algorithm could be used with regard to any symptom dimension, including the miscellaneous, in future studies with different objectives. Another aspect that merits consideration is the subjective choice of methodology.

Other methods might have provided better and/or different results. However, to our knowledge, this is the first study that attempts to address what is now recognized as a limitation of the YBOCS to neurobiological researchers, namely the lack of quantitative, specific symptom-dimension data. Although our results were less than perfect, until now, researchers have typically derived their own formulas ad hoc, without testing or validation, using a weighted average of the number of checklist symptoms per dimension. However, the lack of standardized approach has made comparisons between studies utilizing different methods difficult. Our proposed approach represents a better and validated alternative.

5. Conclusions and future directions

The main contribution of the algorithm developed in this study is to allow for the generation of more refined dimensional phenotypes in datasets that have already been collected with older instruments without symptom subtype scales. The statistical procedure proposed in the current study seems useful for generating dimensional severity ratings from existing YBOCS severity scores. Since our knowledge of the etiopathological mechanisms that lead to OCD is limited, the identification of genetic risk factors for this disorder may provide a better understanding of its etiology. In addition, genetic findings may lead to the identification of targeted and specific treatments that would improve the long-term outcome for subjects suffering from this condition. Finally, the development of this algorithm may have implications that go well beyond the genetics field. Other neurobiological studies such as those focusing on neuroimaging, phenomenology or treatment may benefit from the alternative to analyze YBOCS data in a dimensional fashion.

Ethical statement

This work was developed exclusively with de-identified data sets built over the years using data from different research projects approved by the local Ethics Committee at each participant Institution.

References

Aleman, A., Lincoln, T.M., Bruggeman, R., Melle, I., Arends, J., Arango, C., Knegtering, H., 2016. Treatment of negative symptoms: Where do we stand, and where do we go? Schizophr. Res.. http://dx.doi.org/10.1016/j.schres.2016.05.015. S0920-9964, 16, 30239-0.

Alonso, P., Gratacòs, M., Segalàs, C., Escaramís, G., Real, E., Bayés, M., Labad, J., Pertusa, A., Vallejo, J., Estivill, X., Menchón, J.M., 2011. Variants in estrogen receptor alpha gene are associated with phenotypical expression of obsessive-compulsive disorder. Psychoneuroendocrinology 36 (4), 473–483.

Alonso, P., Gratacòs, M., Segalàs, C., Escaramís, G., Real, E., Bayés, M., Labad, J., López-Solà, C., Estivill, X., Menchón, J.M., 2012. Association between the NMDA glutamate receptor GRIN2B gene and obsessive–compulsive disorder. J. Psychiatry Neurosci. 37 (4), 273–281.

American Psychiatric Association, 2013. Diagnostic and Statistical Manual of Mental Disorders, fifth edition. Washington, DC.

Basu, D., Pereira, C.A.B., 1982. On the Bayesian analysis of categorical data: The problem of non response. J. Stat. Plan. Infer. 6 (4), 345–362.

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.

Burmeister, M., McInnis, M.G., Zöllner, S., 2008. Psychiatric genetics: progress amid controversy. Nat. Rev. Genet. 9 (7), 527–540.

Cappi, C., Brentani, H., Lima, L., Sanders, S.J., Zai, G., Diniz, B.J., Reis, V.N., Hounie, A.G., Conceição do Rosário, M., Mariani, D., Requena, G.L., Puga, R., Souza-Duran, F.L., Shavitt, R.G., Pauls, D.L., Miguel, E.C., Fernandez, T.V., 2016. Whole-exome sequencing in obsessive-compulsive disorder identifies rare mutations in immunological and neurodevelopmental pathways. Transl. Psychiatry 6, e764.

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. 114 (3), 347–353.

R Core Team, 2014. R: A language and environment for statistical computing. In: R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/..

De Luca, V., Gershenzon, V., Burroughs, E., Javaid, N., Richter, M.A., 2011. Age at onset in Canadian OCD patients: mixture analysis and systematic comparison with other studies. J. Affect. Disord. 133 (1–2), 300–304.

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.

Goodman, W.K., Price, L.H., Rasmussen, S.A., Mazure, C., Delgado, P., Heninger, G.R., Charney, D.S., 1989a. The Yale-Brown obsessive compulsive scale. II. Validity. Arch. Gen. Psychiatry 46 (11), 1012–1016.

Goodman, W.K., Price, L.H., Rasmussen, S.A., Mazure, C., Fleischmann, R.L., Hill, C.L., Heninger, G.R., Charney, D.S., 1989b. The Yale-Brown obsessive compulsive scale. I. Development, use, and reliability. Arch. Gen. Psychiatry 46 (11), 1006–1011.

Harrison, B.J., Pujol, J., Cardoner, N., Deus, J., Alonso, P., López-Solà, M., Contreras-Rodríguez, O., Real, E., Segalàs, C., Blanco-Hinojo, L., Menchon, J.M., Soriano-Mas, C., 2013. Brain corticostriatal systems and the major clinical symptom dimensions of obsessive-compulsive disorder. Biol. Psychiatry 73 (4), 321–328.

Hasler, G., Pinto, A., Greenberg, B.D., Samuels, J., Fyer, A.J., Pauls, D., Knowles, J.A., McCracken, J.T., Piacentini, J., Riddle, M.A., Rauch, S.L., Rasmussen, S.A., Willour, V.L., Grados, M.A., Cullen, B., Bienvenu, O.J., Shugart, Y.Y., Liang, K.Y., Hoehn-Saric, R., Wang, Y., Ronquillo, J., Nestadt, G., Murphy, D.L., Study, O.C.G., 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.

Hosmer, D., Lemeshow, S., Sturdivant, R., 2013. Applied Logistic Regression, third ed. John Wiley & Sons, Hoboken, NJ.

Iervolino, A.C., Rijsdijk, F.V., Cherkas, L., Fullana, M.A., Mataix-Cols, D., 2011. A multivariate twin study of obsessive-compulsive symptom dimensions. Arch. Gen. Psychiatry 68 (6), 637–644.

Jhung, K., Ku, J., Kim, S.J., Lee, H., Kim, K.R., An, S.K., Kim, S.I., Yoon, K.J., Lee, E., 2014. Distinct functional connectivity of limbic network in the washing type obsessive-compulsive disorder. Prog. Neuro-Psychopharmacol. Biol. Psychiatry 53, 149–155.

Katerberg, H., Cath, D.C., Denys, D.A., Heutink, P., Polman, A., van Nieuwerburgh, F.C., Deforce, D.L., Bochdanovits, Z., van Balkom, A.J., den Boer, J.A., 2010a. The role of the COMT Val(158) Met polymorphism in the phenotypic expression of obsessive-compulsive disorder. Am. J. Med. Genet. B Neuropsychiatr. Genet. 153B (1), 167–176.

Katerberg, H., Delucchi, K.L., Stewart, S.E., Lochner, C., Denys, D.A., Stack, D.E., Andresen, J.M., Grant, J.E., Kim, S.W., Williams, K.A., den Boer, J.A., van Balkom, A.J., Smit, J.H., van Oppen, P., Polman, A., Jenike, M.A., Stein, D.J., Mathews, C.A., Cath, D.C., 2010b. Symptom dimensions in OCD: item-level factor analysis and heritability estimates. Behav. Genet. 40 (4), 505–517.

Kohlrausch, F.B., Giori, I.G., Melo-Felippe, F.B., Vieira-Fonseca, T., Velarde, L.G., de Salles Andrade, J.B., Fontenelle, L.F., 2016. Association of GRIN2B gene polymorphism and obsessive compulsive disorder and symptom dimensions: a pilot study. Psychiatry Res. 243, 152–155.

Leckman, J.F., Grice, D.E., Boardman, J., Zhang, H., Vitale, A., Bondi, C., Alsobrook, J., Peterson, B.S., Cohen, D.J., Rasmussen, S.A., Goodman, W.K., McDougle, C.J., Pauls, D.L., 1997. Symptoms of obsessive compulsive disorder. Am. J. Psychiatry 154 (7), 911–917.

Lecrubier, Y., 2008. Refinement of diagnosis and disease classification in psychiatry. Eur. Arch. Psychiatry Clin. Neurosci. 258 (Suppl. 1), 6–11. http://dx.doi.org/10.1007/s00406-007-1003-0.

Lennertz, L., Wagner, M., Grabe, H.J., Franke, P.E., Guttenthaler, V., Rampacher, F., Schulze-Rauschenbach, S., Vogeley, A., Benninghoff, J., Ruhrmann, S., Pukrop, R., Klosterkötter, J., Falkai, P., Maier, W., Mössner, R., 2014. 5-HT3 receptor influences the washing phenotype and visual organization in obsessive-compulsive disorder supporting 5-HT3 receptor antagonists as novel treatment option. Eur. Neuropsychopharmacol. 24 (1), 86–94.

Mataix-Cols, D., Rauch, S.L., Baer, L., Eisen, J.L., Shera, D.M., Goodman, W.K., Rasmussen, S.A., Jenike, M.A., 2002. Symptom stability in adult obsessive-compulsive disorder: data from a naturalistic two-year follow-up study. Am. J. Psychiatry 159 (2), 263–268.

Mataix-Cols, D., Wooderson, S., Lawrence, N., Brammer, M.J., Speckens, A., Phillips, M.L., 2004. Distinct neural correlates of washing, checking, and hoarding symptom dimensions in obsessive-compulsive disorder. Arch. Gen. Psychiatry 61 (6), 564–576.

Mataix-Cols, D., Boman, M., Monzani, B., Rück, C., Serlachius, E., Långström, N., Lichtenstein, P., 2013. Population-based, multigenerational family clustering study of obsessive-compulsive disorder. JAMA Psychiatry 70 (7), 709–717.

Mattheisen, M., Samuels, J.F., Wang, Y., Greenberg, B.D., Fyer, A.J., McCracken, J.T., Geller, D.A., Murphy, D.L., Knowles, J.A., Grados, M.A., Riddle, M.A., Rasmussen, S.A., McLaughlin, N.C., Nurmi, E.L., Askland, K.D., Qin, H.D., Cullen, B.A., Piacentini, J., Pauls, D.L., Bienvenu, O.J., Stewart, S.E., Liang, K.Y., Goes, F.S., Maher, B., Pulver, A.E., Shugart, Y.Y., Valle, D., Lange, C., Nestadt, G., 2014. Genome-wide association study in obsessive-compulsive disorder: results from the OCGAS. Mol. Psychiatry 20 (3), 337–344.

Miguel, E.C., Ferrão, Y.A., Rosário, M.C., Mathis, M.A., Torres, A.R., Fontenelle, L.F., Hounie, A.G., Shavitt, R.G., Cordioli, A.V., Gonzalez, C.H., Petribú, K., Diniz, J.B., Malavazzi, D.M., Torresan, R.C., Raffin, A.L., Meyer, E., Braga, D.T., Borcato, S., Valério, C., Gropo, L.N., Prado, H.a.S., Perin, E.A., Santos, S.I., Copque, H., Borges, M.C., Lopes, A.P., Silva, E.D., Disorders, B.R.C.o.O.-C.S., 2008. The Brazilian research consortium on obsessive-compulsive Spectrum disorders: recruitment, assessment instruments, methods for the development of multicenter collaborative studies and preliminary results. Rev. Bras. Psiquiatr. 30 (3), 185–196.

Pauls, D.L., 1992. The genetics of obsessive compulsive disorder and Gilles de la Tourette's syndrome. Psychiatr. Clin. N. Am. 15 (4), 759–766.

Pauls, D.L., 2010. The genetics of obsessive-compulsive disorder: a review. Dialogues Clin. Neurosci. 12 (2), 149–163.

Pauls, D.L., Alsobrook, J.P., Goodman, W., Rasmussen, S., Leckman, J.F., 1995. A family study of obsessive-compulsive disorder. Am. J. Psychiatry 152 (1), 76–84.

Pauls, D.L., Abramovitch, A., Rauch, S.L., Geller, D.A., 2014. Obsessive-compulsive

disorder: an integrative genetic and neurobiological perspective. Nat. Rev. Neurosci. 15 (6), 410–424.
Pereira, C.A.B., Stern, J.M., 2008. Special characterization of standard discrete models. RevStat Stat. J. 6, 199–230.
Pertusa, A., Frost, R.O., Fullana, M.A., Samuels, J., Steketee, G., Tolin, D., Saxena, S., Leckman, J.F., Mataix-Cols, D., 2010. Refining the diagnostic boundaries of compulsive hoarding: a critical review. Clin. Psychol. Rev. 30 (4), 371–386.
Rosario-Campos, M.C., Leckman, J.F., Mercadante, M.T., Shavitt, R.G., Prado, H.S., Sada, P., Zamignani, D., Miguel, E.C., 2001. Adults with early-onset obsessive-compulsive disorder. Am. J. Psychiatry 158 (11), 1899–1903.
Rosario-Campos, M.C., Miguel, E.C., Quatrano, S., Chacon, P., Ferrao, Y., Findley, D., Katsovich, L., Scahill, L., King, R.A., Woody, S.R., Tolin, D., Hollander, E., Kano, Y., Leckman, J.F., 2006. The dimensional Yale-Brown obsessive-compulsive scale (DY-BOCS): an instrument for assessing obsessive-compulsive symptom dimensions. Mol. Psychiatry 11 (5), 495–504.
Rufer, M., Grothusen, A., Mass, R., Peter, H., Hand, I., 2005. Temporal stability of symptom dimensions in adult patients with obsessive-compulsive disorder. J. Affect. Disord. 88 (1), 99–102.
Schooler, C., Revell, A.J., Timpano, K.R., Wheaton, M., Murphy, D.L., 2008. Predicting genetic loading from symptom patterns in obsessive-compulsive disorder: a latent variable analysis. Depress Anxiety 25 (8), 680–688. http://dx.doi.org/10.1002/da.20444.
Shavitt, R.G., de Mathis, M.A., Oki, F., Ferrao, Y.A., Fontenelle, L.F., Torres, A.R., Diniz, J.B., Costa, D.L., do Rosário, M.C., Hoexter, M.Q., Miguel, E.C., Simpson, H.B., 2014. Phenomenology of OCD: Lessons from a large multicenter study and implications for ICD-11. J. Psychiatr. Res. 57, 141–148.
Stewart, S.E., Yu, D., Scharf, J.M., Neale, B.M., Fagerness, J.A., Mathews, C.A., Arnold, P.D., Evans, P.D., Gamazon, E.R., Osiecki, L., McGrath, L., Haddad, S., Crane, J., Hezel, D., Illman, C., Mayerfeld, C., Konkashbaev, A., Liu, C., Pluzhnikov, A., Tikhomirov, A., Edlund, C.K., Rauch, S.L., Moessner, R., Falkai, P., Maier, W., Ruhrmann, S., Grabe, H.J., Lennertz, L., Wagner, M., Bellodi, L., Cavallini, M.C., Richter, M.A., Cook, E.H., Kennedy, J.L., Rosenberg, D., Stein, D.J., Hemmings, S.M., Lochner, C., Azzam, A., Chavira, D.A., Fournier, E., Garrido, H., Sheppard, B., Umaña, P., Murphy, D.L., Wendland, J.R., Veenstra-Vander Weele, J., Denys, D., Blom, R., Deforce, D., Van Nieuwerburgh, F., Westenberg, H.G., Walitza, S., Egberts, K., Renner, T., Miguel, E.C., Cappi, C., Hounie, A.G., Conceição do Rosário, M., Sampaio, A.S., Vallada, H., Nicolini, H., Lanzagorta, N., Camarena, B., Delorme, R., Leboyer, M., Pato, C.N., Pato, M.T., Voyiaziakis, E., Heutink, P., Cath, D.C., Posthuma, D., Smit, J.H., Samuels, J., Bienvenu, O.J., Cullen, B., Fyer, A.J., Grados, M.A., Greenberg, B.D., McCracken, J.T., Riddle, M.A., Wang, Y., Coric, V., Leckman, J.F., Bloch, M., Pittenger, C., Eapen, V., Black, D.W., Ophoff, R.A., Strengman, E., Cusi, D., Turiel, M., Frau, F., Macciardi, F., Gibbs, J.R., Cookson, M.R., Singleton, A., Hardy, J., Crenshaw, A.T., Parkin, M.A., et al., 2013. Genome-wide association study of obsessive-compulsive disorder. Mol. Psychiatry 18 (7), 788–798.
Taj, M.J., Viswanath, B., Purushottam, M., Kandavel, T., Janardhan Reddy, Y.C., Jain, S., 2013. DRD4 gene and obsessive compulsive disorder: do symptom dimensions have specific genetic correlates? Prog. Neuro-Psychopharmacol. Biol. Psychiatry 41, 18–23.
Venables, W.N., Ripley, B.D., 2002. Modern Applied Statistics With S. Springer-Verlag, New York. http://dx.doi.org/10.1007/978-0-387-21706-2.
Via, E., Cardoner, N., Pujol, J., Alonso, P., López-Solà, M., Real, E., Contreras-Rodríguez, O., Deus, J., Segalàs, C., Menchón, J.M., Soriano-Mas, C., Harrison, B.J., 2014. Amygdala activation and symptom dimensions in obsessive-compulsive disorder. Br. J. Psychiatry 204 (1), 61–68.
Waszczuk, M.A., Kotov, R., Ruggero, C., Gamez, W., Watson, D., 2017. Hierarchical structure of emotional disorders: from individual symptoms to the spectrum. J. Abnorm. Psychol. http://dx.doi.org/10.1037/abn0000264.
Yue, W., Cheng, W., Liu, Z., Tang, Y., Lu, T., Zhang, D., Tang, M., Huang, Y., 2016. Genome-wide DNA methylation analysis in obsessive-compulsive disorder patients. Sci Rep 6, 31333.