



Published in final edited form as:

*Eur Child Adolesc Psychiatry*. 2024 November ; 33(11): 3899–3912. doi:10.1007/s00787-024-02431-9.

## Exploring latent clusters in pediatric OCD based on symptoms, severity, age, gender, and comorbidity

Orri Smárason<sup>1,2</sup>, Robert R. Selles<sup>3,4</sup>, Davíð R. M. A. Højgaard<sup>5</sup>, John R. Best<sup>3,4</sup>, Karin Melin<sup>18,6</sup>, Tord Ivarsson<sup>7</sup>, Per Hove Thomsen<sup>5</sup>, Bernhard Weidle<sup>8,9</sup>, Nicole Michelle McBride<sup>10</sup>, Eric A. Storch<sup>11</sup>, Daniel Geller<sup>12,13</sup>, Sabine Wilhelm<sup>12,13</sup>, Lara J. Farrell<sup>14</sup>, Allison M. Waters<sup>14</sup>, Sharna Mathieu<sup>14</sup>, Noam Soreni<sup>15,16</sup>, S. Evelyn Stewart<sup>17,3,4</sup>, Gudmundur Skarphedinsson<sup>1</sup>

<sup>1</sup> Faculty of Psychology, University of Iceland, Reykjavik, Iceland

<sup>2</sup> Department of Child and Adolescent Psychiatry, Landspítali-The National University Hospital of Iceland, Reykjavik, Iceland

<sup>3</sup> University of British Columbia, Vancouver, BC, Canada

✉ Orri Smárason orris@landspitali.is.

**Author Contributions** Drs. Smárason, Skarphedinsson and Højgaard had full access to the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Study concept and design: Selles, Best, Stewart, Højgaard, Smárason & Skarphedinsson. Acquisition, analysis, or interpretation of data: all authors. Drafting of the manuscript: Smárason, Selles, Best, Skarphedinsson and Højgaard. Critical revision of the manuscript for important intellectual content: all authors. Administrative, technical, or material support: Selles, Stewart. Study supervision: Selles, Stewart, Smárason, Højgaard and Skarphedinsson.

**Conflict of interest** Dr. Selles has served as a consultant and Scientific Advisory Committee Member for AnxietyCanada. Dr. Thomsen has served on the Advisory Board for the Tryg Foundation and has received speaking honoraria from Medice and Shire within the last three years. Dr. Storch has received research support from the National Institutes of Health (NIH), the Agency for Healthcare Research and Quality, the International OCD Foundation, the Ream Foundation, Greater Houston Community Foundation, and the Texas Higher Education Coordinating Board. He has received royalties from Elsevier Publications, Springer Publications, American Psychological Association, Wiley, Inc, and Lawrence Erlbaum. He has served on the Speaker's Bureau and Scientific Advisory Board for the International OCD Foundation. He is a consultant for Biohaven and Brainsway. He has received research support from the McIngvale Presidential Endowed Chair. Dr. Geller has received grant support from NIH, book honorarium from the American Academy of Child and Adolescent Psychiatry, and speaking honoraria for Advanced Institute lectures from the American Academy of Child and Adolescent Psychiatry and Massachusetts General Hospital Psychiatry Academy in educational programs supported through independent medical education grants from pharmaceutical companies. He has received funding from Neurocrine Bioscience and Biohaven Pharmaceuticals. Dr. Wilhelm has received grant support from NIH, International OCD Foundation and the Tourette Association of America. She is a presenter for the Massachusetts General Hospital Psychiatry Academy in educational programs supported through independent medical education grants from pharmaceutical companies; she has received royalties from Elsevier Publications, Guilford Publications, New Harbinger Publications, Springer, and Oxford University Press. Dr. Wilhelm has also received speaking honoraria from various academic institutions and foundations, including the International OCD Foundation, Tourette Association of America, and Brattleboro Retreat. In addition, she received payments from the Association for Behavioral and Cognitive Therapies for her role as Associate Editor for the Behavior Therapy journal, as well as from John Wiley & Sons, Inc. for her role as Associate Editor on the journal Depression & Anxiety. Dr. Wilhelm has also received honoraria from One-Mind for her role in PsyberGuide Scientific Advisory Board. Dr. Wilhelm is also on the Scientific Advisory Board for Koa Health and Noom. Dr. Wilhelm has received salary support from Novartis and research support from Koa Health. Dr. Soreni has received grants and research funding from the Canadian Institutes of Health Research, HAHSO Hamilton and the Ontario Brain Institute. Dr. Soreni has been the recipient of an investigator-initiated operating grant from Lundbeck, LLC. He has served on the scientific committee of the Anxiety and Depression Association of America. Dr. Stewart has received grants and research funding from the Anxiety Disorders Association of America, the Canadian Institutes of Health Research, the American Academy of Child and Adolescent Psychiatry, and the Michael Smith Foundation for Health Research. She has received honoraria from Aarhus University for service as an external examiner as well as speaking honoraria from Yale University and from the National Institutes of Mental Health and Neurosciences. She has served as a member of the Scientific and Clinical Advisory Board of the International OCD Foundation and the Scientific Advisory Board of AnxietyCanada. Drs. Skarphedinsson, Højgaard, Ivarsson, Farrell, Waters, Mathieu, Soreni, McBride and Smárason report no biomedical financial interests or potential conflicts of interest.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00787-024-02431-9>.

- <sup>4</sup> BC Children's Hospital, Vancouver, BC, Canada
- <sup>5</sup> Department of Child and Adolescent Psychiatry, Aarhus University Hospital, Psychiatry, Denmark
- <sup>6</sup> Institute of Health and Care Sciences, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
- <sup>7</sup> University of Gothenburg, Gothenburg, Sweden
- <sup>8</sup> Regional Center for Child Mental Health and Child Welfare, Norwegian University of Science and Technology, Trondheim, Norway
- <sup>9</sup> St. Olavs University Hospital, Trondheim, Norway
- <sup>10</sup> Embedded Preventive Behavioral Health Capability, III MEF, United States Marine Corps, Okinawa, Japan
- <sup>11</sup> Baylor College of Medicine, Houston, TX, USA
- <sup>12</sup> Harvard Medical School, Boston, MA, USA
- <sup>13</sup> Massachusetts General Hospital, Boston, MA, USA
- <sup>14</sup> Griffith University, Queensland, Australia
- <sup>15</sup> St. Joseph's Hospital, Hamilton, ON, Canada
- <sup>16</sup> McMaster University, Hamilton, ON, Canada
- <sup>17</sup> BC Mental Health and Substance Use Services, Burnaby, BC, Canada
- <sup>18</sup> Region Västra Götaland, Department of Child and Adolescent Psychiatry, Sahlgrenska University Hospital, Gothenburg, Sweden

## Abstract

Given diverse symptom expression and high rates of comorbid conditions, the present study explored underlying commonalities among OCD-affected children and adolescents to better conceptualize disorder presentation and associated features. Data from 830 OCD-affected participants presenting to OCD specialty centers was aggregated. Dependent mixture modeling was used to examine latent clusters based on their age- and gender adjusted symptom severity (as measured by the Children's Yale-Brown Obsessive-Compulsive Scale; CY-BOCS), symptom type (as measured by factor scores calculated from the CY-BOCS symptom checklist), and comorbid diagnoses (as assessed via diagnostic interviews). Fit statistics favored a four-cluster model with groups distinguished primarily by symptom expression and comorbidity type. Fit indices for 3–7 cluster models were only marginally different and characteristics of the clusters remained largely stable between solutions with small clusters of distinct presentations added in more complex models. Rather than identifying a single classification system, the findings support the utility of integrating dimensional, developmental, and transdiagnostic information in the conceptualization of OCD-affected children and adolescents. Identified clusters point to the centrality of contamination concerns to OCD, relationships between broader symptom expression and higher levels of comorbidity, and the potential for complex/neurodevelopmental presentations.

## Keywords

Obsessive–compulsive disorder; Assessment; Symptomatology; Comorbidity

---

## Background

Obsessive compulsive disorder (OCD) is an impairing psychiatric condition affecting approximately 1% of individuals across the lifespan [1], with onset commonly occurring between childhood and late adolescence [2]. Rather than topically-specific, OCD is defined by a psychopathological relationship (i.e., unwanted/distressing internal experiences that are avoided, reduced, or escaped via repetitive action [3], resulting in diverse symptom expression that appears influenced by developmental factors [4–6], and interactions with common comorbid concerns [7–9], such as anxiety, tic, and attention deficit hyperactivity disorders (ADHD) [10].

Identifying clusters of OCD-affected children and adolescents based on symptom expression, comorbid psychopathology, and demographic factors is consistent with recommendations towards dimensional and psychometrically-informed conceptualizations of psychopathology [11], and may shed additional light on common clinical presentations, offer insights regarding potential underlying transdiagnostic processes, and lead to improved specificity of care [12]. Various methods have previously been utilized to explore shared characteristics among OC-presentations. Factor analysis of the symptom checklist of the Children’s Yale-Brown Obsessive Compulsive Scale (CY-BOCS) [13] has suggested symptoms coalesce around: (1) contamination obsessions and cleaning/washing compulsions; (2) harm/sexual obsessions and checking compulsions; and (3) symmetry obsessions and compulsions, with hoarding symptoms sometimes included or identified as their own factor [14–17]. Unfortunately, frequent within-individual endorsement of symptoms across multiple dimensions limits the utility of these factors for informing individual-level classification of OCD-affected children and adolescents [14, 17].

Benefiting from a person- as compared to item-oriented approach, cluster-based analytic methods (e.g., dependent mixture modelling and latent class analysis [LCA]) have demonstrated utility in producing meaningful subgroups among youth with mental health concerns [18]. LCA is a structural equation modeling technique similar to confirmatory factor analyses (CFA); both measure a latent variable while accounting for measurement error [19]. However, LCA assumes the presence of a categorical instead of continuous latent variable and groups observations, or participants, into classes (rather than grouping variables into latent constructs as CFA) [20]. While utilized to explore either OC-symptom expression or comorbid disorder presentations among OCD-affected individuals, these methods have not yet been utilized to examine both domains simultaneously. In OCD-affected adults, analyses based on symptom expression have identified frequency-[21] and etiologically-based clusters [22], while analyses of comorbidity among OCD-affected adults has suggested three- or four-class solutions based on number and type of comorbid disorders [8, 23]. Among pediatric samples, cluster analyses examining symptom expression have found similar groupings to factor-analytic studies [24, 25]. Højgaard et al. [9] examined

comorbid profiles among OCD-affected children and adolescents and identified a three-class solution, characterized by those with no comorbidity, those with neurological/behavioral conditions (e.g., ADHD, tics, ODD), and those with comorbid anxiety disorders. Although OCD symptom types were not included in the model, secondary analyses suggested symmetry/hoarding symptoms were associated with both comorbid classes, harm/sexual symptoms were associated with the comorbid anxiety class, and comorbidity was less common among those with contamination/cleaning symptoms.

The present study seeks to replicate and expand on the work of Højgaard et al. [9, 16] by exploring latent clusters among a large, aggregated, international sample of OCD-affected children and adolescents.

### **Aim 1.**

Confirm a factor structure to develop factor loadings for use in the analysis. We hypothesized that a three-factor solution as in Højgaard et al. [16] (i.e., contamination/cleaning; harm/sexual; symmetry/hoarding) would provide an acceptable model fit. If a three-factor solution does not provide an acceptable fit in this sample, an exploratory factor analysis will be carried out to extract a usable factor structure.

### **Aim 2.**

Identify latent clusters representing sub-groups of OCD-affected children and adolescents based upon symptom factor scores, symptom severity, and the presence/absence of comorbid disorders. We hypothesized a four-class solution would emerge characterized by: (A) contamination/cleaning symptoms and low comorbidity; (B) harm/sexual symptoms and anxious comorbidity; (C) symmetry/hoarding symptoms and neurobehavioral comorbidity; and (D) multi-dimensional symptoms and multiple comorbid conditions.

### **Aim 3.**

Explore relationships of cluster membership with other available clinically relevant variables, including age of symptom onset, avoidance, impairment, and family accommodation.

## **Methods**

### **Ethical considerations**

All sites had ethics approval and obtained participant assent and family consent for participation in research and data sharing.

### **Participants and procedures**

The present study includes data from 830 OCD-affected children and adolescents. Data was obtained by aggregating participants with completed CY-BOCS checklist data from seven international pediatric OCD programs (see Table S1) (NordLOTS [26]; GBG [27]; DCS [28]; UBC POP [29]; Griffith [30]). All participants had a confirmed OCD diagnosis. Average OCD-Severity was in the moderate-severe range (mean = 24.5; SD = 5.9). The

sample was 54% female and 5–19 years of age (mean = 12.9; SD = 2.9). Participant characteristics for individual programs and the combined sample are presented in Table 1.

## Measures

**Diagnostic Assessments.**—Baseline diagnoses were assessed and diagnosed using either the Schedule for Affective Disorders and Schizophrenia for School-Age Children—Present and Lifetime (n = 567; 68.3%) [31] or the Anxiety Disorder Interview Schedule for Children—Parent Version (n = 263; 31.2%) [32]. *OCD Symptoms and Severity.* OCD symptoms were assessed across all sites using the CY-BOCS, a measure with well-established psychometric properties [13, 33]. The CY-BOCS symptom checklist was utilized to assess OCD symptom types. Symptom severity was assessed via the CY-BOCS total score, which is comprised of ten-items rated 0 to 4, with higher scores indicating more severe symptoms.

**Additional Variables.**—The six CY-BOCS extension items, which evaluate, on a 0 to 4 scale, additional domains related to OCD (e.g., insight, avoidance, doubt) were also completed with many participants, although this varied between sites (n = 402–728; 49–88%, Table S7 provides more information about missingness of extension items). While exact measure usage varied across sites (see Table S1), child- and parent-ratings of OCD-related impairment was assessed using variations of the Child Obsessive-Compulsive Impact Scale [34, 35] and family accommodation was assessed using variations of the Family Accommodation Scale [36–38]. While of interest to profile composition, less consistent measure completion across sites limited the feasibility of their inclusion. Therefore, relationships between these variables and profiles were examined in a secondary analysis.

## Analytic plan

Data processing and statistical analyses were conducted using Mplus 8.6[39] and R version 4.0.3 ([r-project.org](http://r-project.org)). For *Aim 1*, an acceptable factor structure of the CY-BOCS symptom checklist variables was established. Factor structures derived from previously identified solutions by Højgaard et.al [16], and from an exploratory factor analysis of the current data were tested in a confirmatory factor analysis (CFA). Models were directly compared using the scaled chi-squared difference test [40]. Latent factor scores for each participant were generated based on the best fitting CFA model, using the R *lavaan* package (version 0.6–7) [41]. Checklist items were modelled as ordinal variables with diagonally weighted least squares to estimate model parameters and robust standard errors. Model fit statistics, including the comparative fit index (CFI), the Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean square residual (RMSR) were calculated and compared across models, seeking the most parsimonious acceptable fit. Acceptable fit was defined as CFI and TLI values of at least 0.9.

For *Aim 2* the analysis entailed dependent mixture modelling with the R-package *depmixS4* (version 1.4–2) [42] to identify distinct latent groups underlying the observed data. LCA with continuous indicators is also known as a *mixture model*, reflecting the assumption that the observed data is a mixture of distributions from multiple underlying subpopulations [20]. Each participant' factor scores extracted from the best-fitting CFA, total CY-BOCS

score, and 9 comorbidities were included as the observed variables in this analysis. Child age and gender were treated as covariates using a stepwise approach, where factors were enumerated first using age and gender adjusted indicators and differences in factors then tested based on relevant covariates (i.e., age of symptom onset, avoidance, impairment, and family accommodation). The continuous measures—the CFA-derived factors, CY-BOCS total, and child age—were normalized such that the minimum score was transformed to zero and the maximum score to one. This was done separately by study site to account for variation across site. Some checklist items were missing from specific sites (“*No concern with consequences of contamination other than how it might feel*”, and “*Mental Compulsions*” were missing in Griffith; “*Checking that did not/will not harm others*”, and “*Need to do things until it feels just right*” were missing in ATRC) and were excluded from the CFA analysis to maximize numbers of participants. Participants with CFA derived factor scores as well as other indicators (e.g., comorbidities) were included in the mixture model. In the model, continuous variables were modeled as a Gaussian distribution while categorical variables were modeled with a multinomial. Parameters were estimated by the expectation-maximization algorithm and posterior state classification was determined by the Viterbi algorithm.

Mixture solutions ranging from 1 to 9 distinct groups were evaluated; Akaike Information Criterion [AIC] and Bayesian Information Criterion [BIC] were computed for each solution, with BIC providing a stronger penalty of model complexity [43, 44]. Additionally, entropy and average posterior probability for each observation’s most likely group measures were computed to evaluate the quality of group separation, with values closer to 1 reflecting superior delineation of the groups [43, 45]. The likelihood ratio tests directly compared models using the difference in  $\chi^2$  and degrees of freedom; a larger  $\chi^2$  value indicates a larger difference in the models.

For Aim 3 a set of analyses examined between-cluster differences in secondary outcomes of interest. Prior to conducting between-cluster tests, children were assigned to the cluster to which they had the maximum posterior probability value of belonging. Omnibus cluster differences were identified with a  $p$  value threshold  $< 0.05$  and were adjusted for study site. Follow-up pairwise comparisons used the Tukey correction for multiple comparisons.

## Results

Descriptive statistics for the overall sample and each study site are provided in Table 1.

### Factor analysis of OCD symptoms

The previously identified three-factor model [16] was initially tested and did not meet requirements of acceptable fit in the current sample (see Table S3). Therefore, an exploratory factor analysis was conducted to suggest the composition of usable factor solutions. A parallel analysis suggested a ten-factor model. A visual scree plot inspection indicated four to six factors (Figure S1), and the 0.7 Kaiser criterion indicated six factors. Fit measures for three- to eleven-factor EFA models were examined further and all models with four factors or more had adequate fit, with fit values improving with each added factor

(see Table S2) The CY-BOCS checklist dataset was screened for multivariate assumptions (normality, linearity, homogeneity, and homoscedasticity), which were all met [46].

A six-factor solution provided an acceptable fit (see Table S3) in a CFA. Although EFA results potentially suggest factor solutions with more factors, models with more than six factors failed to converge on an optimal solution in a CFA. The current six-factor model suggests CY-BOCS symptom checklist items can be attributed to contamination/cleaning, sexual/religious/embarrassment, somatic/illness, harm/checking, superstition/repetition, and hoarding/perfectionism latent factors. Factor loadings for the checklist items and correlations among the six factors are provided in Table S4 and Table S5, respectively.

### **Cluster analysis of OCD symptom factors, severity, comorbidities and demographics**

Nine possible solutions were generated by the dependent mixture model. Based on model fit criteria and visual interpretability, solutions involving 3–7 clusters were explored in further detail (See Fig. 1). Table 2 provides model fit and descriptive outcomes of the solutions and identified clusters, while Table 3 describes features associated with the identified clusters. Fit indices favored the four-cluster solution; however, all examined models demonstrated acceptable participant classification/entropy (0.8 or higher, see Table 2).

As observed in Fig. 1 and described in Table 3, variation across clusters regarding contamination/cleaning symptoms, symptom severity, and gender distribution was mostly minimal except in more complex models. Clusters were most notably distinguishable by loadings on the other symptom factors and comorbidity patterns (variable specific entropy values for 3–7 cluster models are presented in Table S6). The overall characteristics of the first three identified clusters remained observable and somewhat consistent across subsequent models. In the four-cluster model, two distinct groups emerge; C2 (pink, see Fig. 1) with a mixed symptom presentation, relatively lower contamination/cleaning scores compared to other symptom types, and high levels of comorbid disorders compared to other clusters; and C4 (green) a smaller group, high on contamination/cleaning scores but low on other symptom types and levels of comorbidity. However, these characteristics became less distinct in later models. C1 (light blue) and C3 (yellow) both present mixed profiles at moderate levels, mainly distinguishable by higher levels of comorbid Tics for C1. In the six-factor model, C6 (dark blue) emerged with a small but distinct subset of children and adolescents, with high rates of ASD, ADHD, tics, GAD, and Social Phobia. In the seven-factor model, C7 (grey) emerged with an additional small but distinct subset of mostly female youth who demonstrate high rates of ASD, low levels of other comorbidities and almost exclusively contamination/cleaning type symptoms.

### **Cluster differences on secondary outcomes**

A final set of analyses compared latent clusters on secondary measures using the four-group solution as it showed the best BIC and had an acceptable entropy value (see Table S7). No significant differences were observed across clusters for CY-BOCS auxiliary items (Tukey-corrected  $p > 0.05$ ) except for slowness where C3 (Yellow) demonstrated higher ratings compared to C4 (green). No differences were observed between clusters on age of

onset, measures of impairment, or family accommodation (Tukey-corrected  $p > 0.05$ ). Table S8 provides descriptive statistics from the 4-cluster solution for variables not included in other analyses (e.g., race, study site).

## Discussion

Using dependent mixture modeling, the present study explored shared characteristics among a global sample of OCD-affected youth based on severity, symptom type, and comorbidity. After confirming a six-factor model for the CY-BOCS symptom checklist, our results suggested that youth could be acceptably classified into three to seven clusters, with a four-cluster solution offering optimal fit statistics, but more complex models identifying unique and potentially clinically relevant sub-groups. Given this, and the limitation that cluster identification is dependent on the variables included (or not included) in the model, we have opted to present, describe, and note differences between clusters in the varying models, rather than highlighting a single model as “the” way that OCD-affected youth are most accurately classified. This open examination is possible in part due to the general stability of cluster characteristics across models.

### Discussion of cluster characteristics

The C1 (light blue, see Fig. 1) cluster appears to feature low to moderate endorsement of symptoms across all six OCD symptom factors, average severity, and generally low frequency of internalizing comorbid disorders. However, Tic disorders and ADHD are quite common in this cluster. The C1 cluster initially comprised the second-largest proportion of the sample (23%), but steadily lost members as additional clusters were added, decreasing to 18% of the sample by the 7-cluster model but remaining stable regarding its characteristics.

The C2 (pink) cluster is the largest cluster throughout all models, although it does shrink in proportion from 42% in the 3-cluster model to 22% in the 7-cluster model. These youth had more severe OCD symptoms, endorsed a varied symptom profile and had high comorbid Tics, ADHD as well as internalizing disorders in most models. These patterns suggest that increased diversity in symptom expression may be indicative of additional underlying fear/anxiety processes and/or cognitive-dysfunction (e.g., maladaptive beliefs, repetitive negative thinking) [47, 48].

Comprised of approximately 20–35% of the sample, cluster C3 (yellow) seems to represent a “median” presentation of OCD in most of the models, featuring average to low (in the five- and six factor models) severity and mixed symptom- and comorbidity profiles throughout.

C4 (green) emerges in the four-factor model as a small (14%) group of youth almost exclusively endorsing contamination/cleaning symptoms and showing low levels of comorbidity. Along with the consistent presence of contamination symptoms across the larger clusters, these findings suggest contamination symptoms are the most specific or central to OCD [16, 49]. In subsequent models however, C4 trends towards more mixed symptom expression and comorbidity profiles while remaining stable in size.



C5 (orange) initially identified 9% of the sample, favored males slightly and featured high severity, relatively high levels of superstition/repetition and harm/checking symptoms in the five and six-cluster models, but more predominant contamination/cleaning symptoms in the seven-cluster model. C5 increases slightly in size and trends toward higher levels of anxiety comorbidities in the last two models.

Two small but distinct clusters emerged in the final two models. Distinguished by significantly higher frequencies of comorbid disorders and making up 9% of the sample, C6 (dark blue), showed average severity, high levels of perfectionism/hoarding and the highest levels of all clusters on ASD, ADHD and tics, indicative of a neuropsychological profile and/or difficulties in inhibitory control. The presence of this class might lend partial support to the concept of a tic-related OCD subtype associated with hoarding symptoms, previously identified in some studies [50–52]. In the seven-factor model, however, C6 is distinguished by very low levels of contamination/cleaning symptoms, low severity and low internalizing comorbidity indicating that this is not a stable cluster between models. Making up 6% of the sample, C7 (grey) consisted mostly of females exclusively endorsing contamination/cleaning symptoms, showing overall low levels of comorbidity but highest levels of ASD of all classes in the model. Although such small subgroups might be unstable, they could, if replicated, provide important information for future studies examining the etiology of OCD or the impact of cluster membership on treatment outcomes.

The present results are not directly comparable to previous findings as we used a new six-factor structure for OCD-symptom dimensions. We, however, note similarities with the current best fitting four-cluster model and previous findings; with the almost exclusive endorsement of contamination/cleaning symptoms being associated with a low level of comorbidities [9], and a trend of more diverse OCD-symptom expressions being associated with higher comorbidity levels [8, 9]

The identification of specific clusters of pediatric OCD patients may lead the way for more research regarding dimensional conceptualization of OCD. OCD is a highly heterogeneous disorders with a wide variety of symptoms [15, 16]. However, the present results also indicate that specific dimensions beyond OCD symptoms may exist, such as profiles based on symptoms, comorbidity, and other related factors. Many experienced clinicians treating OCD might be able to recognize and identify frequently coexisting patterns of symptoms and characteristics of behaviors or personality in children with OCD. Examples may be the typically different presentation of OCD symptoms with comorbid tic disorder, or the tendency for low motivation for OCD treatment in children with comorbid ASD. These symptom clusters based on experience and limited studies about how comorbidity and personality may modify OCD-symptoms could obtain a better scientific base by identifying more homogenous subsets of pediatric OCD, based on factors including, but not limited to, symptom profiles. At least, the present study has provided a grouping pattern that might potentially be valuable for specific targeted studies of neurocircuitry, neurocognitive function and dysfunction with the overall goal of more personalized and effective treatment [53].

Studies of latent clusters may open ways to investigate treatment effects for different subsets of patients and, additionally, to identify potential treatment modifications that could be made to tailor treatments for different subsets. Some clusters (e.g., cluster C4, with the majority endorsing contamination and cleaning with low comorbidity) might suggest main E/RP focus on contamination and cleaning with minimum modification needed for most patients [53, 54], while C2, with its diverse symptom profile and high level of comorbidity, might require a more comprehensive approach, possibly involving a larger focus on dysfunctional cognitive beliefs [55]. Previous studies have also found hoarding symptoms to be associated with higher levels of comorbidity. In line with this finding, we find a large class (C2) which shows the highest level of hoarding symptoms, a highly diverse symptom profile, and the highest levels of overall comorbidity. As both high levels of hoarding symptoms [56, 57] and comorbidity [58] are associated with attenuated treatment benefits, this group might be of particular clinical interest and might require tailored treatment adjustments. Overall, the current results are in line with family-genetic studies indicating that pediatric-onset OCD is etiologically heterogeneous [59].

### Limitations

There are several limitations to note. First, while the study provides indications of cluster characteristics at the time of assessment, the study design does not inform the extent to which cluster distribution changes over time (e.g., if participants migrate from one group to another as they age), nor does it allow for direct testing of underlying mechanisms. Second, the clusters provide average scores of those included within that class, rather than clear and definable rules regarding class membership. As a result, applying these categorizations to any individual would be premature at this time. Third, secondary variables were compared between classes after making concrete class assignments (as opposed to allowing for partial class membership). This process overestimates confidence in class assignment. Entropy was strong but not perfect, meaning that there was good but not perfect confidence in identifying which participant was assigned to each cluster. Fourth, site-based variance might have been enhanced by lack of standardization of procedures. Fifth, ASD diagnoses were based on the K-SADS and ADIS which are primarily symptom screeners of ASD caseness, and not adequate for establishing ASD diagnosis on their own. ASD diagnosis was an exclusion criterion in the NordLOTS study, resulting in the underrepresentation of comorbid ASD in the current sample. Sixth, the 5–7 factor models have fairly unequal class prevalence and simulation studies have indicated that in such models smaller classes may be difficult to recover at small sample sizes [45, 60]. Seventh, the enumeration of factors through the mixture modelling process is sensitive to how covariates are treated and the assumption of different population models to ours, regarding the effects of age and gender, might yield different results [60]. Eighth, although the sample was large and combined from global centers, these centers were specialized, age-specific, located within countries with high development indexes and primarily Caucasian populations, and utilized dichotomized, rather than continuous, assessments of comorbidity. Given that results are dependent on sample characteristics and included variables, a different formation of clusters could occur if other variables (e.g., avoidance, other comorbidities) were included and the generalizability of results to other OCD-affected populations may be limited.

## Conclusions

The present study provides an explorative overview of potential latent sub-groups among OCD-affected children and adolescents. The identified clusters support efforts towards dimensional or transdiagnostic conceptualizations of psychopathology and, by identifying more homogenous subsets of OCD, based on relevant shared characteristics, may serve as translational markers for further investigations of relevant neurocircuitry and mechanisms of dysfunction, such as neurocognition, cognitive flexibility, inhibitory control, and fear extinction processes [61–64]. In informing clinical care, clusters provide an indication of potentially relevant adaptations to treatment, such as supplemental emphasis on cognitive processes and maladaptive beliefs, executive skills and impulse control, distress tolerance and emotion regulation, and/or parental roles and involvement.

Replication of clusters in more diverse populations (e.g., additional countries, socio-economic groups, ages, and comorbidities) and among individuals with other primary disorders is warranted.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Funding

This work was supported by fellowships to Dr. Selles from the BC Children's Hospital Research Institute and the Michael Smith Foundation for Health Research (MSFHR: #17821). In addition, the original studies from which data were aggregated in the present study report the following funding: The DCS-CBT study was supported by grants from the National Institute of Mental Health (NIMH) to Drs. Storch (1R01MH093381) and Geller (5R01MH093402). The Griffith program received funding from the National Health and Medical Research Council (NHMRC), Foundation for Children, and Rotary Foundation. Nord-LOTS was supported by the TrygFonden, Danish Council for Strategic Research, Pulje til styrkelse af psykiatrisk forskning i Region Midtjylland, Center for Child and Adolescent Mental Health, Eastern and Southern Norway, Stiftelsen Clas Groschinsky's Minnesfond, and Norwegian Research Council, Helse and Rehabilitering, Norge. UBC-POP was supported by MSFHR and the British Columbia Provincial Health Services Administration. Sahlgrenska University Hospital OCD Outpatient Clinic – GBG study was supported by the Agreement concerning research and education of doctors – Region Vestra Götaland, the Claes Groschinsky Memorial Fund and the Iris Jonzén-Sandblom and Greta Jonzé's Foundation. The funding organizations were not involved in the design or conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, and approval of the manuscript. The opinions expressed within the present study are those of the authors and do not necessarily reflect those of the funding organizations.

## References

1. Fawcett EJ, Power H, Fawcett JM (2020) Women are at greater risk of OCD than men: a meta-analytic review of OCD prevalence worldwide. *J Clin Psychiatr.* 10.4088/JCP.19r13085
2. Anholt GE et al. (2014) Age of onset in obsessive-compulsive disorder: admixture analysis with a large sample. *Psychol Med* 44(1):185–194 [PubMed: 23517651]
3. American Psychiatric Association (2013) Diagnostic and statistical manual of mental disorders, fifth edition (DSM-5(TM)). American Psychiatric Publishing, Arlington
4. Nakatani E et al. (2011) Children with very early onset obsessive-compulsive disorder: clinical features and treatment outcome. *J Child Psychol Psychiatr* 52(12):1261–1268
5. Selles RR, Storch EA, Lewin AB (2014) Variations in symptom prevalence and clinical correlates in younger versus older youth with obsessive-compulsive disorder. *Child Psychiatr Hum Dev* 45(6):666–674

6. Smarason O et al. (2022) Age differences in children with obsessive-compulsive disorder: symptoms, comorbidity, severity and impairment. *Nord J Psychiatr*. 10.1080/08039488.2021.2019917
7. Højgaard DRMA et al. (2018) Symptom Dimensions and Clinical Correlates of OCD in a Large Sample of Children and Adolescents. In: Højgaard DRMA (ed) 65th annual meeting. The American Academy of Child and Adolescent Psychiatry, Seattle
8. Nestadt G et al. (2009) Obsessive-compulsive disorder: subclassification based on co-morbidity. *Psychol Med* 39(9):1491–1501 [PubMed: 19046474]
9. Højgaard DRMA et al. (2018) Obsessive-compulsive symptom dimensions: Association with comorbidity profiles and cognitive-behavioral therapy outcome in pediatric obsessive-compulsive disorder. *Psychiatr Res* 270:317–323
10. Selles RR et al. (2018) Symptom insight in pediatric obsessive-compulsive disorder: outcomes of an international aggregated cross-sectional sample. *J Am Acad Child Adolesc Psychiatr* 57(8):615–619
11. Krueger RF, Markon KE (2006) Understanding psychopathology: melding behavior genetics, personality, and quantitative psychology to develop an empirically based model. *Curr Dir Psychol Sci* 15(3):113–117 [PubMed: 18392116]
12. Gillan CM, Fineberg NA, Robbins TW (2017) A trans-diagnostic perspective on obsessive-compulsive disorder. *Psychol Med* 47(9):1528–1548 [PubMed: 28343453]
13. Scahill L et al. (1997) Children's yale-brown obsessive compulsive scale: reliability and validity. *J Am Acad Child Adolesc Psychiatr* 36(6):844–852
14. Bernstein GA et al. (2013) Pediatric obsessive-compulsive disorder: Symptom patterns and confirmatory factor analysis. *J Obsess Compuls Relat Disord* 2(3):299–305
15. Bloch MH et al. (2008) Meta-analysis of the symptom structure of obsessive-compulsive disorder. *Am J Psychiatr* 165(12):1532–1542 [PubMed: 18923068]
16. Højgaard DRMA et al. (2017) Structure and clinical correlates of obsessive-compulsive symptoms in a large sample of children and adolescents: a factor analytic study across five nations. *Eur Child Adolesc Psychiatr* 26(3):281–291
17. Stewart SE et al. (2008) Four-factor structure of obsessive-compulsive disorder symptoms in children, adolescents, and adults. *J Am Acad Child Adolesc Psychiatr* 47(7):763–772
18. Petersen KJ, Qualter P, Humphrey N (2019) The application of latent class analysis for investigating population child mental health: a systematic review. *Front Psychol* 10:1214 [PubMed: 31191405]
19. Molenaar PCM, von Eye A (1994) On the arbitrary nature of latent variables. In: von Eye A, Clogg CC (eds) *Latent variables analysis: Applications for developmental research*. Sage Publications, Inc., Thousand Oaks, CA, pp 226–242
20. Van Lissa CJ, Garnier-Villarreal M, Anadria D (2023) Recommended practices in latent class analysis using the open-source R-package tidySEM. *Struct Equat Model Multidiscip J*. 10.1080/10705511.2023.2250920
21. Delucchi KL et al. (2011) Latent class analysis of the yale-brown obsessive-compulsive scale symptoms in obsessive-compulsive disorder. *Compr Psychiatr* 52(3):334–341
22. Atli A et al. (2014) Latent class analysis of obsessive-compulsive symptoms in a clinical sample. *Compr Psychiatr* 55(3):604–612
23. Nestadt G et al. (2003) The identification of OCD-related subgroups based on comorbidity. *Biol Psychiatr* 53(10):914–920
24. Hybel KA et al. (2017) Symptom profiles and executive function in childhood obsessive-compulsive disorder. *J Obsess-Compu Relat Disord* 14:36–46
25. Ivarsson T, Valderhaug R (2006) Symptom patterns in children and adolescents with obsessive-compulsive disorder (OCD). *Behav Res Ther* 44(8):1105–1116 [PubMed: 16207484]
26. Torp NC et al. (2015) Effectiveness of cognitive behavior treatment for pediatric obsessive-compulsive disorder: acute outcomes from the nordic long-term OCD treatment study (Nord-LOTS). *Behav Res Ther* 64:15–23 [PubMed: 25463245]
27. Melin K et al. (2015) Treatment and 12 month outcome of children and adolescents with obsessive-compulsive disorder: a naturalistic study. *J Obsess-Compu Relat Disord* 6:1–6

28. Storch EA et al. (2016) Efficacy of augmentation of cognitive behavior therapy with weight-adjusted d-cycloserine vs placebo in pediatric obsessive-compulsive disorder: a randomized clinical trial. *JAMA Psychiat* 73(8):779–788
29. Stewart SE et al. (2017) A multisite study of family functioning impairment in pediatric obsessive-compulsive disorder. *J Am Acad Child Adolesc Psychiatr* 56(3):241–249
30. Farrell L et al. (2022) Efficacy of D-cycloserine augmented brief intensive cognitive-behavioural therapy for paediatric obsessive-compulsive disorder: a randomised clinical trial. *Depress Anxiety* 39(6):461–473 [PubMed: 35084071]
31. Kaufman J et al. (1997) Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatr* 36(7):980–988
32. Silverman WK, Albano AM (1996) The anxiety disorders interview schedule for children and parents— DSM-IV version. Graywind, New York
33. Storch EA et al. (2004) Psychometric evaluation of the children's yale-brown obsessive-compulsive scale. *Psychiatr Res* 129(1):91–98
34. Piacentini J et al. (2003) Functional impairment in children and adolescents with obsessive-compulsive disorder. *J Child Adolesc Psychopharmacol* 13(Suppl 1):S61–S69 [PubMed: 12880501]
35. Piacentini J et al. (2007) Functional impairment in childhood OCD: development and psychometrics properties of the child obsessive-compulsive impact scale-revised (COIS-R). *J Clin Child Adolesc Psychol* 36(4):645–653 [PubMed: 18088221]
36. Calvocoressi L et al. (1999) Family accommodation of obsessive-compulsive symptoms: instrument development and assessment of family behavior. *J Nerv Ment Dis* 187(10):636–642 [PubMed: 10535658]
37. Flessner CA et al. (2011) Predictors of parental accommodation in pediatric obsessive-compulsive disorder: findings from the pediatric obsessive-compulsive disorder treatment study (POTS) trial. *J Am Acad Child Adolesc Psychiatr* 50(7):716–725
38. Pinto A, Van Noppen B, Calvocoressi L (2013) Development and preliminary psychometric evaluation of a self-rated version of the family accommodation scale for obsessive-compulsive disorder. *J Obsess Compuls Relat Disord* 2(4):457–465
39. Muthén LK, Muthén BO (1998) *Mplus use s guide*, 8th edn. Muthén & Muthén, Los Angeles
40. Satorra A, Bentler PM (2001) A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika* 66(4):507–514
41. Rosseel Y (2012) lavaan: an R package for structural equation modeling. *J Stat Softw* 48(2):1–36
42. Visser I, Speekenbrink M (2010) depmixS4: an R package for hidden markov models. *J Stat Softw* 36(7):1–21
43. Masyn KE (2013) Latent class analysis and finite mixture modeling. In: Little TD (ed) *The oxford handbook of quantitative methods*. Oxford University Press, pp 551–611
44. Nylund KL, Asparoutiov T, Muthen BO (2007) Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Struct Equ Model Multidiscip J* 14(4):535–569
45. Nylund-Gibson K (2018) Ten frequently asked questions about latent class analysis. *Transl Issu Psychol Sci*. 10.1037/tps0000176
46. Howard MC (2016) A review of exploratory factor analysis decisions and overview of current practices: what we are doing and how can we improve? *Int J Hum-Comput Interact* 32(1):51–62
47. Cervin M et al. (2020) Incompleteness, harm avoidance, and disgust: a comparison of youth with OCD, anxiety disorders, and no psychiatric disorder. *J Anxiety Disord* 69:102175 [PubMed: 31896022]
48. Cervin M et al. (2021) Involvement of fear, incompleteness, and disgust during symptoms of pediatric obsessive-compulsive disorder. *Eur Child Adolesc Psychiatr* 30(2):271–281
49. Torres AR et al. (2013) Clinical features of pure obsessive-compulsive disorder. *Compr Psychiatr* 54(7):1042–1052

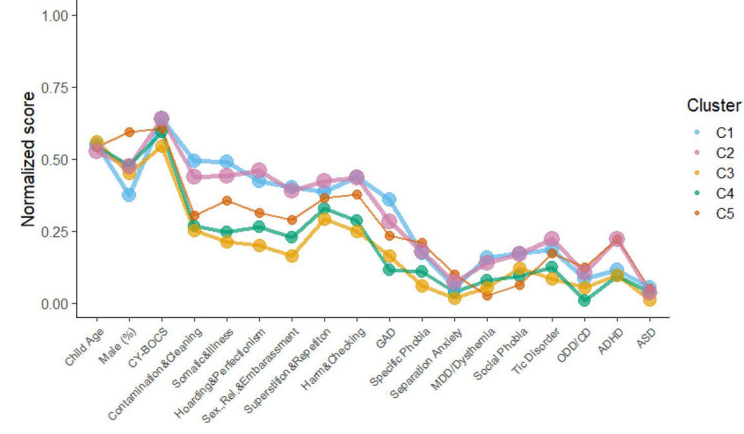
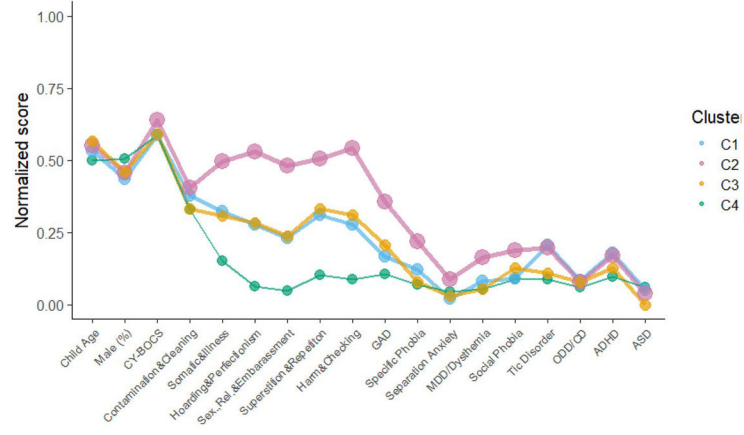
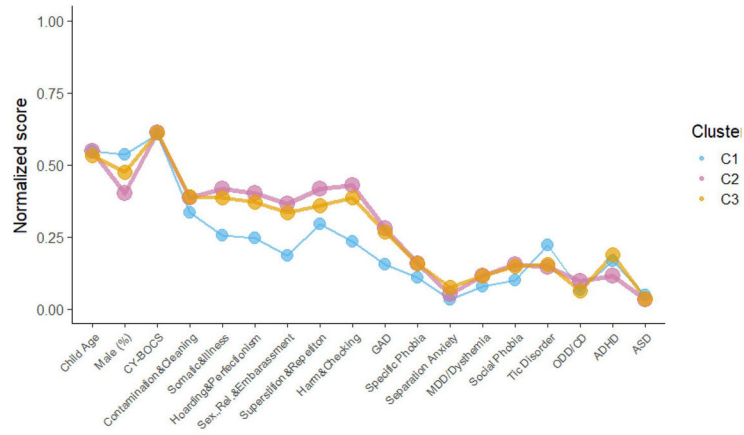
50. Højgaard DRMA et al. (2017) Pediatric obsessive-compulsive disorder with tic symptoms: clinical presentation and treatment outcome. *Eur Child Adolesc Psychiatr* 26(6):681–689
51. Leckman JF et al. (1997) Symptoms of obsessive-compulsive disorder. *Am J Psychiatr* 154(7):911–917 [PubMed: 9210740]
52. Storch EA et al. (2008) Obsessive-compulsive disorder in youth with and without a chronic tic disorder. *Depress Anxiety* 25(9):761–767 [PubMed: 17345600]
53. Caporino NE, Storch EA (2016) Personalizing the treatment of pediatric obsessive-compulsive disorder: evidence for predictors and moderators of treatment outcomes. *Curr Behav Neurosci Rep* 3(1):73–85
54. Ginsburg GS et al. (2008) Predictors of treatment response in pediatric obsessive-compulsive disorder. *J Am Acad Child Adolesc Psychiatr* 47(8):868–878
55. Cervin M et al. (2022) Cognitive beliefs across the symptom dimensions of pediatric obsessive-compulsive disorder: type of symptom matters. *Behav Ther* 53(2):240–254 [PubMed: 35227401]
56. Abramowitz JS et al. (2003) Symptom presentation and outcome of cognitive-behavioral therapy for obsessive-compulsive disorder. *J Consult Clin Psychol* 71(6):1049–1057 [PubMed: 14622080]
57. Mataix-Cols D et al. (2002) Obsessive-compulsive symptom dimensions as predictors of compliance with and response to behaviour therapy: results from a controlled trial. *Psychother Psychosom* 71(5):255–262 [PubMed: 12207105]
58. Storch EA et al. (2008) Impact of comorbidity on cognitive-behavioral therapy response in pediatric obsessive-compulsive disorder. *J Am Acad Child Adolesc Psychiatr* 47(5):583–592
59. Leckman JF, Bloch MH, King RA (2009) Symptom dimensions and subtypes of obsessive-compulsive disorder: a developmental perspective. *Dialog Clin Neurosci* 11(1):21–33
60. Nylund-Gibson K, Masyn KE (2016) Covariates and mixture modeling: results of a simulation study exploring the impact of misspecified effects on class enumeration. *Struct Equ Model Multidiscip J* 23(6):782–797
61. Abramovitch A, De Nadai AS, Geller DA (2021) Neurocognitive endophenotypes in pediatric OCD probands, their unaffected parents and siblings. *Prog Neuropsychopharmacol Biol Psychiatr* 110:110283
62. Geller DA et al. (2018) Neurocognitive function in paediatric obsessive-compulsive disorder. *W J Biol Psychiatr* 19(2):142–151
63. McGuire JF et al. (2016) Fear conditioning and extinction in youth with obsessive-compulsive disorder. *Depress Anxiety* 33(3):229–237 [PubMed: 26799264]
64. Negreiros J et al. (2020) Neurocognitive risk markers in pediatric obsessive-compulsive disorder. *J Child Psychol Psychiatr* 61(5):605–613

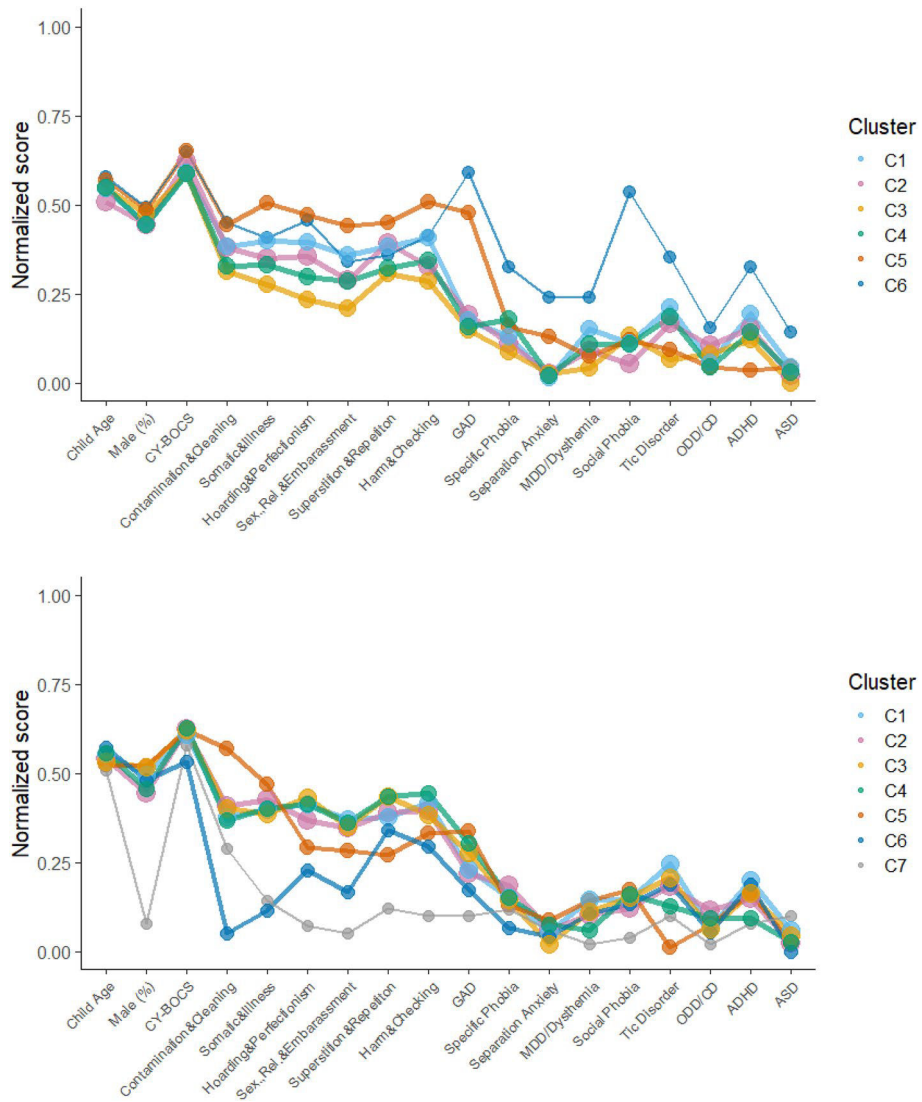
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript





**Fig. 1.**  
Plots of variable scores across 3–7 cluster models



Table 1

Participant characteristics for individual programs and the combined sample

Variable	Overall, N = 830 <sup>f</sup>	GBG, N = 91 <sup>f</sup>	AFRC, N = 56 <sup>f</sup>	DCS-CBT, N = 153 <sup>f</sup>	Griffith, N = 114 <sup>f</sup>	Iceland, N = 4 <sup>f</sup>	Nord-LOTS, N = 263 <sup>f</sup>	UBC-POP, N = 148 <sup>f</sup>	Site difference p value <sup>2</sup>
Child age	12.9 (2.9)	12.9 (2.6)	13.5 (2.8)	12.4 (3.0)	12.2 (2.6)	12.5 (2.4)	12.8 (2.8)	13.7 (3.1)	< 0.001
Age of onset	8.7 (3.4)	9.8 (2.9)	-	8.2 (3.4)	8.6 (3.2)	-	8.8 (3.5)	8.5 (3.5)	0.018
< Missing >	< 103 >	< 26 >	< 56 >	< 8 >	< 9 >	< 4 >	< 0 >	< 0 >	
Age of diagnosis	10.8 (4.2)	-	-	-	11.1 (2.9)	-	11.6 (3.0)	10.0 (5.2)	0.056
< Missing >	< 360 >	< 91 >	< 56 >	< 153 >	< 21 >	< 4 >	< 35 >	< 0 >	
OCD severity	24.5 (5.9)	23.0 (5.9)	22.7 (5.7)	25.1 (6.1)	26.8 (4.0)	24.2 (3.0)	24.6 (5.1)	23.2 (7.4)	< 0.001
Insight	1.3 (1.0)	-	1.3 (0.6)	1.1 (1.1)	1.6 (1.0)	-	1.1 (1.0)	1.6 (0.9)	< 0.001
< Missing >	< 102 >	< 91 >	< 0 >	< 0 >	< 1 >	< 4 >	< 0 >	< 6 >	
Avoidance	1.8 (1.2)	-	-	1.5 (1.2)	2.2 (0.9)	-	1.8 (1.1)	2.0 (1.3)	< 0.001
< Missing >	< 158 >	< 91 >	< 56 >	< 0 >	< 1 >	< 4 >	< 0 >	< 6 >	
Indecision	1.1 (1.1)	< 91 >	-	1.0 (1.2)	1.5 (1.1)	-	1.0 (1.1)	1.1 (1.1)	< 0.001
< Missing >	< 154 >	< 91 >	< 56 >	< 0 >	< 1 >	< 4 >	< 0 >	< 2 >	
Responsibility	0.8 (1.0)	< 91 >	-	0.6 (0.9)	1.5 (1.1)	-	0.8 (1.0)	0.6 (0.8)	< 0.001
< Missing >	< 157 >	< 91 >	< 56 >	< 0 >	< 1 >	< 4 >	< 0 >	< 5 >	
Slowness	1.2 (1.1)	< 91 >	-	1.2 (1.1)	1.5 (1.0)	-	1.1 (1.1)	1.3 (1.0)	0.010
< Missing >	< 159 >	< 91 >	< 56 >	< 0 >	< 2 >	< 4 >	< 0 >	< 6 >	
Doubt	1.0 (1.0)	< 91 >	-	0.9 (1.0)	1.7 (1.0)	-	-	0.9 (1.0)	< 0.001
< Missing >	< 418 >	< 91 >	< 56 >	< 0 >	< 1 >	< 4 >	< 263 >	< 3 >	
Impairment (parent)	31.1 (24.9)	56.3 (33.8)	-	15.6 (10.9)	41.0 (24.6)	-	25.7 (18.7)	35.7 (20.3)	< 0.001
< Missing >	< 181 >	< 7 >	< 56 >	< 8 >	< 7 >	< 4 >	< 92 >	< 6 >	
Impairment (child)	25.5 (2284)	41.7 (31.3)	-	14.0 (10.4)	40.4 (30.0)	-	19.9 (14.4)	25.5 (19.1)	< 0.001
< Missing >	< 185 >	< 17 >	< 56 >	< 16 >	< 9 >	< 4 >	< 60 >	< 23 >	
Family accommodation	13.3 (12.0)	-	18.0 (10.6)	15.1 (8.4)	26.1 (10.1)	-	16.7 (12.0)	0.4(0.3)	< 0.001
< Missing >	< 221 >	< 91 >	< 1 >	< 0 >	< 66 >	< 4 >	< 53 >	< 6 >	
Gender, male	382 (46%)	38 (42%)	19 (34%)	78 (51%)	50 (44%)	1 (25%)	129 (49%)	67 (45%)	< 0.001

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Variable	Overall, N = 830 <sup>1</sup>	GRG, N = 91 <sup>1</sup>	ATRC, N = 56 <sup>1</sup>	DCS-CBT, N = 153 <sup>1</sup>	Griffith, N = 114 <sup>1</sup>	Iceland, N = 4 <sup>1</sup>	Nord-LOTS, N = 263 <sup>1</sup>	UBC-POP, N = 148 <sup>1</sup>	Site difference p value <sup>2</sup>
Child race									< 0.001
American Indian	5 (0.6%)	-	-	0 (0%)	0 (0%)	0 (0%)	-	5 (3.3%)	
Asian	27 (3.3%)	-	-	2 (1.3%)	2 (2.1%)	0 (0%)	-	23 (15.5%)	
Black	8 (1.0%)	-	-	6 (3.9%)	0 (0%)	0 (0%)	-	2 (1.4%)	
Native Hawaiian/ Pacific islander	0 (0%)	-	-	0 (0%)	0 (0%)	0 (0%)	-	0 (0%)	
White	347 (41.8%)	-	-	142 (93%)	92 (98%)	4 (100%)	-	109 (73.6%)	
Mixed race	2 (0.4%)	-	-	2 (1.3%)	0 (0%)	0 (0%)	-	0 (0%)	
Other	10 (1.2%)	-	-	1 (0.7%)	0 (0%)	0 (0%)	-	9 (6.1%)	
< Missing >	< 339 >	< 91 >	< 56 >	< 0 >	< 20 >	< 0 >	< 263 >	< 0 >	
Ethnicity, hispanic	28 (3.4%)	-	0 (0%)	24 (16%)	0 (0%)	0 (0%)	-	4 (2.7%)	< 0.001
< Missing >	< 283 >	< 91 >	< 0 >	< 0 >	< 20 >	< 0 >	< 263 >	< 0 >	
Any prior treatment	351 (74%)	-	21 (38%)	99 (65%)	55 (79%)	-	-	102 (70%)	< 0.001
< Missing >	< 439 >	< 91 >	< 0 >	< 0 >	< 44 >	< 4 >	< 263 >	< 37 >	
Current SSRIs	219 (26.5%)	53 (58.2%)	14 (25%)	44 (29%)	36 (32%)	-	0 (0%)	72 (48.6%)	0.041
< Missing >	< 4 >	< 1 >	< 0 >	< 0 >	< 0 >	< 4 >	< 0 >	< 0 >	
GAD	204 (24.6%)	11 (12%)	35 (62%)	34 (22%)	63 (55%)	1 (25%)	19 (7.2%)	40 (27%)	< 0.001
Social phobia	115 (13.9%)	8 (8.8%)	26 (46%)	19 (12%)	27 (24%)	4 (100%)	12 (4.6%)	19 (12.8%)	< 0.001
Sep anxiety	45 (5.4%)	3 (3.3%)	6 (11%)	9 (5.9%)	17 (15%)	0 (0%)	6 (2.3%)	4 (2.7%)	< 0.001
Specific phobia	123 (14.8%)	25 (27.5%)	17 (30%)	16 (10%)	39 (34%)	1 (25%)	20 (7.6%)	5 (3.4%)	< 0.001
ADHD	127 (15.3%)	15 (16.5%)	8 (14%)	30 (20%)	14 (12%)	0 (0%)	23 (8.7%)	37 (25%)	< 0.001
ODD/CD	64 (7.7%)	8 (8.8%)	7 (12%)	18 (12%)	12 (11%)	0 (0%)	10 (3.8%)	9 (6.1%)	< 0.001
MDD/dys-thymia	88 (10.6%)	27 (29.7%)	11 (20%)	18 (12%)	11 (9.6%)	1 (25%)	8 (3.0%)	12 (8.1%)	< 0.001
Tic disorder	137 (16.5%)	26 (28.6%)	9 (16%)	7 (4.6%)	15 (13%)	0 (0%)	49 (19%)	31 (20.9%)	< 0.001
ASD	29 (3.5%)	5 (5.5%)	0 (0%)	0 (0%)	18 (16%)	0 (0%)	1 (0.4%)	6 (4.1%)	< 0.001

<sup>1</sup> Mean (SD); n (%)

<sup>2</sup> Kruskal-Wallis rank sum test; Pearson's Chi-squared test

Table 2

## Outcomes of the dependent mixture modeling

	3 Cluster	4 Cluster	5 Cluster	6 Cluster	7 Cluster
AIC	16375.24	15736.28	15938.22	15908.63	15424.94*
BIC	17163.72	16789.16*	17255.50	17490.31	17271.01
Entropy <sup>a</sup>	0.86	0.90*	0.86	0.86	0.89
LRT $\chi^2$ (df) <sup>b</sup>	316.8 (21)	150.9 (21)	107.8 (21)	116.9 (21)	

Factors	Fit c	n	%	Fit c	n	%	Fit c	n	%	Fit c	n	%
C1 (Light Blue)	0.91	194	23	0.93	200	24	0.90	190	23	0.90	160	19
C2 (Pink)	0.95	345	42	0.96	350	42	0.93	265	32	0.90	200	24
C3 (Yellow)	0.94	291	35	0.92	165	20	0.89	164	20	0.92	150	18
C4 (Green)	-	-	-	0.98	115	14	0.91	130	16	0.94	140	17
C5 (Orange)	-	-	-	-	-	-	0.91	81	9	0.87	109	13
C6 (D. Blue)	-	-	-	-	-	-	-	-	-	0.87	71	9
C7 (Grey)	-	-	-	-	-	-	-	-	-	-	-	-

AIC Akaike information criterion, BIC Bayesian information criterion, LRT log-likelihood ratio test

<sup>a</sup>Entropy values range from 0 to 1 (better separation between factors)

<sup>b</sup>The log-likelihood ratio test compares two adjacent models (e.g., 3-factor versus 4-factor). Larger values represent a bigger improvement in model fit from the simpler to more complex model. Across all comparisons,  $p < .001$

<sup>c</sup>Fit is defined as the average posterior probability of belonging to this particular group after assigning individuals to groups based on their maximum posterior probability. Optimally = 1.0

\* indicates best fit

Table 3

## Descriptive interpretation of identified clusters

ID (color)	Proportion of sample	Age distribution <sup>a</sup>	Gender distribution	Severity	Symptoms	Comorbidities	Stability
C1 (L Blue)	Medium (18–24%)	Mean	Females = males	Mean	Mix at low to moderate frequency	Low internalizing comorbidity, high on Tics and ADHD	Gradually shrinks
C2 (pink)	Medium to large (22–42%)	Mean	Females slightly > males	Slightly above Mean	Mixed at moderate frequency	Higher rates of GAD, Social Phobia, Tics and ADHD	Shrinks after 4-factor model
C3 (yellow)	Medium (19–35%)	Mean	Females = males	Mean	Mixed at moderate frequency	Low comorbidity	Becomes smaller in 4-factor model; stable in subsequent models
C4 (green)	Small (14–17%)	Mean	Males slightly > females	Mean	Predominant contamination with others at low frequency initially, becomes mixed in later models	Low comorbidity initially, moderate comorbidity in six and seven cluster models	Remains stable in size but gradually moves towards more mixed symptom profiles and higher comorbidity
C5 (orange)	Small (9–13%)	Slightly older	Females = males	Above mean	Moderate frequency, favors superstition/repetition and harm/checking in five and six-cluster models, predominant contamination in seven cluster model	Higher rates of GAD and Social Phobia in six and seven-cluster models	Moves towards more contamination symptoms and higher GAD and Social Phobia comorbidities
C6 (D blue)	Small (9%)	Mean	Males = females	Mean	Mix at high frequency, high on contamination/cleaning and hoarding/perfectionism in 6-factor solution; low frequency in the 7-factor solution with very low contamination/cleaning	Highest comorbidity overall in 6-factor model, especially high in GAD, Social Phobia, Tics, ADHD and ASD; much lower in 7-factor model although ADHD is still prevalent	Highly unstable
C7 (grey)	Very small (6%)	Mean	Females greatly > males	Below mean	Mix at low frequency, favors contamination/cleaning	Low overall; Highest cluster on ASD in model	n/a

<sup>a</sup>By normalized age: slightly older (0.51–0.63), much older (0.63–0.75), very much older (> 0.75), slightly younger (0.38–0.49), much younger (0.25–0.38), very much younger (< 0.25)