Describing Symptom Heterogeneity in Obsessive-Compulsive Disorder: Relationships Between and Stability Within the Overt Symptom and Core Dimensions Models

A Thesis Submitted to the Committee on Graduate Studies
in Partial Fulfillment of the Requirements for the Degree of Master of Science
in the Faculty of Arts and Science

TRENT UNIVERSITY
Peterborough, Ontario, Canada

Psychology M.Sc. Graduate Program
September 2024

Abstract

Describing Symptom Heterogeneity in Obsessive-Compulsive Disorder: Relationships Between and Stability Within the Overt Symptom and Core Dimensions Models

Jordan Till

Two dimension-based models have emerged that capture the high symptom heterogeneity observed in obsessive-compulsive disorder (OCD)—one grouping symptoms by their overt presentation, the other identifying core dimensions of underlying symptom motivations. This paper presents two studies that explore (1) the relationship between these two models, and (2) the stability of the models and dimensions they each represent. Study 1 found that the core dimension motives jointly predicted overall OCD symptom severity, and that each core dimension further predicted unique nonoverlapping overt symptom dimensions. Study 2 demonstrated that despite high instances of item-level symptom change, both models and their respective dimensions were longitudinally stable, with exception to overt symptoms relating to harm/injury/bad luck. Both studies support further consideration of these models as clinical resources and support the candidacy of the core dimensions as endophenotypes for OCD. Results, limitations, clinical implications, and future directions are discussed.

Keywords: Obsessive-Compulsive Disorder, OCD, Overt Symptoms, Core Dimensions, Harm Avoidance, Incompleteness, Symptom Models, Stability, Endophenotypes

Acknowledgements

I would like to thank my supervisor, Dr. Laura J. Summerfeldt, for her constant support and encouragement, and for the autonomy I was granted in designing and carrying out this project. Thank you for your countless hours over these past years, for your continued guidance and expertise, and for your levity, stories, teaching, and passion for research. I owe you so much.

I would also like to thank my second reader, Dr. James D. A. Parker, for his direction and expert oversight, and my good friend and colleague Colin Henning, to whom I promise fewer late-night stats-related questions now that this project has finished. Thank you to my graduate peers Eva Fernandes, Kalyca Hébert, and Thomas Bodrug for always being up for a walk or a chat, and for always being there to brighten my day; I couldn't have done this without you. I would also like to thank my family, for always supporting me, for being such positive role models, and for always providing such a solid base that has allowed for my continued progress.

Most importantly, I would like to thank my wife Shelby, who at this point has read more drafts of my various projects than anyone should have to read. You have helped me at every step. Thank you for your optimism, for our many adventures, for your constant encouragement, and for supporting me in achieving my dreams.

To the many amazing friends, peers, and mentors I am surrounded by

—Thank you!

Table of Contents

Abstract ii

Acknowledgements iii

Chapter 1: An Introduction to Symptom Models in Obsessive-Compulsive Disorder 1

Organizing Symptom Variance in OCD 2

The Overt Symptom and Core Dimension Models 4

Meaningful Associations with Clinically Relevant Variables 10

Symptom Models from the Perspective of Longitudinal Stability 16

Study Overviews 28

Chapter 2: Study 1 – Predicting Overt Symptoms in OCD: Examining the Utility of the Core Dimensions Model 29

Current Study 30

Methods 32

Results 38

Discussion 43

Conclusion 50

Chapter 3: Study 2 – Organizing Symptoms of OCD: Assessing Stability in the Overt Symptom and Core Dimension Models 53

Current Study 55

Methods 59

Results 68

Discussion 72

Conclusion 88

Chapter 4: Summary of Findings, Common Limitations, and Future Directions 92

References 96

Appendix A: Study 1 Tables and Figures 118

Study 1 Sample Demographics 118

Study 1 Measure Descriptives 119

Study 1 Measure Zero-Order Correlations 120

Regression Models: Core Dimensions as Predictors for Overt Symptom Dimensions 121

Study 1 Regression Model β Comparisons 122

Study 1 Partial Spearman Correlations 123

Appendix B: Study 2 Tables 124

Study 2 Sample Demographics 124

Study 2 Measure Descriptives 125

Study 1 and 2 Sample Mean Comparison 126

Dimension Mean Stability: Significant Mean Change from Time 1 to Time 2 127

Dimension Mean Stability: Mean Change Scores 127

Dimension Rank-Order Stability 128

Ipsative Stability in the Overt Symptom and Core Dimension Models 128

Significant Exploratory Zero-Order Spearman Correlations 129

Appendix C: The Double-Entry Intraclass Correlation 130

Case Examples, and Modifications 132

Appendix D: Study Advertisement 137

Appendix E: Created and Modified Measures 138

Life Events Questionnaire 138

Treatment History Questionnaire 140

Describing Symptom Heterogeneity in Obsessive-Compulsive Disorder: Relationships Between and Stability Within the Overt Symptom and Core Dimensions Models

Chapter 1: An Introduction to Symptom Models in Obsessive-Compulsive Disorder

Obsessive-compulsive disorder (OCD) is a mental health disorder characterized by obsessions—unwanted recurrent thoughts, urges, or mental images that an individual attempts to ignore or suppress; and compulsions—behaviours that the individual feels driven to perform to diminish subjective distress, usually relating to the obsessions (American Psychiatric Association, 2013). Lifetime clinical prevalence for OCD is estimated between 1–3% (e.g., Ruscio, 2010) and the disorder is a leading cause of disability worldwide (World Health Organization, 2008) with substantial similarities across cultures (Hunt, 2020). When left untreated, OCD tends to demonstrate a chronic symptom trajectory despite symptom severity lessening somewhat with time (Pinto et al., 2006; Skoog & Skoog, 1999). While contemporary treatments are largely effective (e.g., Olatunji et al., 2013), clinical non-response is still significant (e.g., Pallanti et al., 2002) and relapse remains a cogent concern (e.g., Eisen et al., 1999; Eisen et al., 2013), highlighting the need to better understand the disorder, its etiological constructs, maintaining factors, and treatment options.

One point of ongoing contention is a debate over how best to conceptualize the structure of OCD (e.g., American Psychiatric Association, 2013; Leckman et al., 2010; Mataix-Cols et al., 2005; Mataix-Cols et al., 2008; Miguel et al., 2005), particularly regarding whether it is best approached as a unitary disorder or is rather multidimensional or subtyped. Such questions of nosology are largely due to the

disorder's high between- and within-person symptom heterogeneity (e.g., Miguel et al., 2005; Rettew et al., 1992; Skoog & Skoog, 1999), and from emergent observations of both symptom-specific and shared etiological correlates (e.g., Alsobrook et al., 1999; Mahjani et al., 2021; van den Heuvel et al., 2009). Along with these observations, it has been suggested that a comprehensive model explaining this heterogeneity would allow for better diagnostic specificity, tailored therapeutic support, and ultimately, better clinical outcomes (e.g., Clark, 2006; Matsunaga et al., 2010).

Organizing Symptom Variance in OCD

Despite being moved to its own diagnostic category in the latest major revision of the Diagnostic and Statistical Manual of Mental Disorders (American Psychiatric Association, 2013), OCD remains a unitary diagnosis, with little in the way of criteria to specify variance in symptoms in any standardized manner. Notably high observed symptom heterogeneity has long motivated exploration into whether further categorizing or 'subtyping' the disorder might provide clinical utility (Abramowitz & McKay, 2008; Leckman et al., 2010; McKay & Neziroglu, 2009; Ruscio et al., 2010; Williams et al., 2013). While current literature has established empirical approaches to describing these differences, attempts to explain symptom variance have been found even in early seminal literature, with Psychiatrist Jean-Pierre Falret referencing 'madness of doubt' and 'delirium of touch' (1869, as cited in Leckman et al., 2009). Today, within casual discourse, the adoption and use of colloquial 'subtype' labels to describe symptoms is commonplace. Online forums and select commercial service advertisements (e.g., NOCD—a professional online OCD-specific teletherapy service; McGrath, 2024) are increasingly utilizing subtype labels to describe symptom phenomenology. Because of

the informal nature of the labels, there is no agreed-upon or set list of these specific emergent categories; however, they typically highlight common broad thematic trends in symptoms and tend to focus on the obsession side of the obsession/compulsion dyad. Some examples of these colloquial subtypes are “harm OCD”, “sexual orientation OCD”, and “relationship OCD” (see McGrath, 2024). At the time of this writing, these and similar subtype labels and their acronyms are regularly used in social forums as a type of descriptive symptom shortform. Some issues inherent to this approach of symptom categorization are attributable to the fact that they are neither informed nor structured through empirical data or theory—meaning their reliance on purely descriptive criteria for the creation and justification of symptom categories rests on an arguably precarious foundation of semantic minutia. Given the broad diversity in how symptoms manifest for individuals, using this approach, there are arguably an infinite potential number of ‘subtypes’ that could exist if the only criteria for legitimacy are their own subjective identification. Relatedly, based on observing how OCD symptoms typically manifest, utilizing singular labels to identify ‘your OCD type’ may be a problematically reductive characterization of how symptoms typically present and evolve. As has been shown in the literature, symptoms tend to have broad presentation, rarely load uniformly on single symptom factors, and tend to change over time (e.g., Leckman et al., 2009; Ruscio et al., 2010; Skoog & Skoog, 1999). While it may seem intuitive to label a set of symptoms that coalesce around romantic partners as “relationship OCD”, such a description is vague at best, and does not appear to carry notable clinical utility (Williams et al., 2013). During the time spent interacting within the online forums for this research study alone, I have encountered several instances of members posting in distress due to their OCD symptoms

not aligning with the 'types' they found described online, or similarly, because their care providers refused to diagnose them with an OCD 'type' despite the current diagnostic manual not having type specifiers (American Psychological Association, 2013). While there are clearly perceived benefits of an ability to organize and describe symptoms, as can be seen by the rapid adoption of these colloquial subtype labels, a haphazard nosology built solely on common descriptions of thematic obsessive foci is not likely to benefit research or clinical advancement and may complicate the understanding of symptoms for those trying to seek information about their disorder.

The Overt Symptom and Core Dimension Models

Parallel to these informal developments, organized attempts to describe and categorize symptoms have progressed within the research literature. The clear inherent benefits of such scientific attention include the prioritization of empirical evidence, iteration, and replication to justify nosological models. Some examples of published approaches to organizing OCD symptom heterogeneity and/or candidate subtype criterion include: age of onset (e.g., Janowitz et al., 2009; Taylor 2011), biological sex (e.g., Torresan et al., 2013), presence of tic-related symptoms (e.g., Grados et al., 2001), present comorbidities (e.g., Garyfallos et al., 2010), observed autoimmune/acute-onset (e.g., PANS/PANDAS; Jaspers-Fayer et al., 2017), autogenic vs. reactive symptoms (e.g., Lee & Kwon, 2003), overt symptom 'themes' (see Rowsell & Francis, 2015), and by associated dysfunctional beliefs (see Bragdon & Coles, 2017). Among these efforts, one of the most replicated emergent nosology models is a dimensional conceptualization that uses statistical data reduction to organize overt symptoms (i.e., the tangible obsessions or compulsion behaviours) as recorded by the Yale-Brown Obsessive-

Compulsive Scale (Y-BOCS; Goodman et al., 1989); the current gold standard measure used to assess OCD symptom severity (Bloch et al., 2008; Williams et al., 2013).

The Y-BOCS symptom checklist from which these overt factor-analyzed models are derived consists of 72 item-level OC symptoms (e.g., fear might harm others) that are organized among 15 a-priori categories (e.g., aggressive obsessions). I have made the distinction of labeling these 15 categories as 'a-priori' because the 72 item-level symptoms they represent were organized using "rational" means—relying solely on theory, observation, and face-valid deduction based on how they are likely to relate/coexist—rather than empirical means. The a-priori categories are split by their representations of obsessions (8 categories) and compulsions (7 categories), and there are two miscellaneous symptom categories among them: one for obsessions, and one for compulsions (Goodman et al., 1989).

In studies that have explored the further reduced 'overt' dimensional approach (henceforth referred to as the overt symptom model), large datasets of symptom checklist answers from the Y-BOCS are analyzed through factor analyses, which typically yield a set of 3–5 factor dimensions, generalizing to: (1) symmetry, (2) unacceptable thoughts, (3) contamination, and (4) hoarding[1] (Bloch et al., 2008). Baer et al. (1994) were the first to publish on this approach, and fairly similar factor-model outcomes have been extensively replicated since (see Bloch et al., 2008). Subsequent studies have suggested that these dimensions are similar across different age cohorts (Mataix-Cols et al., 2008; Stewart et al., 2008) and are present cross-culturally (Matsunaga et al., 2008), suggesting

[1] Hoarding has since become a separate diagnosis from OCD (American Psychiatric Association, 2013).

the possibility of a universal core structure for overt symptom expression under the umbrella diagnosis of OCD.

Despite the popularity and replication of these overt symptom models, there have been questions regarding their empirical validity and the potential limitations of organizing symptoms in this way (Cameron & Streiner et al., 2019; Clark, 2005; Bloch et al., 2008, McKay et al., 2004; Radomsky & Taylor, 2005; Summerfeldt et al., 1999; Summerfeldt et al., 2014). Over-and-above the slight variance in the number of found factor dimensions, which may be attributable to differences in methodology or samples (e.g., Bloch et al., 2008), other cited issues include variables weakly supporting their parent factors (Pinto et al., 2008; Summerfeldt et al., 1999), poor predictive validity from the factor dimensions themselves (Summerfeldt et al., 2004; Storch et al., 2008), symptoms that associate strongly with multiple factor dimensions, and poor integration or omission altogether of the two 'miscellaneous' a-priori categories from the Y-BOCS (which concerningly represent over a quarter of the item-level checklist items; Summerfeldt et al., 1999, 2004). Summerfeldt et al. (2004) suggested these issues may be partially due to the propensity of many teams to rely on the symptom structure provided by the Y-BOCS, by using the scores from each a-priori category for their factor analyses rather than performing item-level analyses with the checklist data. Analysis using the a-priori symptom categories automatically integrates the predetermined groupings of the item-level factors they represent (e.g., 'fear might harm self', 'fear might harm others', and 'fear will act on unwanted impulses' are all included under the a-priori category of 'aggressive obsessions') rather than observing how/if those item-level symptoms might coexist outside of their assumed likeness prescribed through the Y-BOCS structure (for

comparisons of factor outcomes using Y-BOCS category vs. item-level analyses, see Bloch et al., 2008). Furthermore, despite being labeled as 'dimensional' constructs, the scoring of these factor-analyzed dimensions from the Y-BOCS checklist data are limited in their ability to represent truly dimensional scores because the checklist items themselves do not include measures of severity. This typically means that formation of a dimensional representation either relies on a simple count of the number of item-level symptoms endorsed within each a-priori category/dimension (e.g., Delorme et al., 2006), or relies on other workarounds such as identifying the 'primary' vs. 'secondary' experiences of symptoms, then using those assigned values to create dimension scores (e.g., Mataix-Cols et al., 2002). Limitations to these approaches are explored in more depth in a later section; however, suffice it to say that claims of true dimensionality for symptom severity based on binary 'yes/no' reports of item-level symptoms represent methodological and interpretational confounds in this context.

However, another arguably more important point raised by numerous groups (e.g., Clark, 2005; Starcevic et al., 2011; Summerfeldt, 2004; Summerfeldt et al., 1999, 2004), is that grouping symptoms by overt features may carry inherent limitations altogether. Consider the example of an individual who washes excessively: it may be that they are doing so to eliminate germs to alleviate fears of contamination or sickness; conversely, it may be that they feel compelled to continue because anything less than a certain number of consecutive washes feels subjectively 'wrong', or 'incomplete'. In this example, the overt symptom is identical; however, the underlying motive(s) may be meaningfully disparate. Summerfeldt et al. argue these lower-level motives—the reasons or drivers underlying symptoms—may represent a more cogent and informative criteria

for categorization and may have the potential to explain some of the empirical issues seen from the factor-analytic approaches to overt symptoms (Summerfeldt, 2004; Summerfeldt et al., 1999, 2004).

> Our findings provide grounds for reservation about the soundness of relying upon overt symptom similarities to identify a taxonomy of OCD. This practice overlooks the fact that overtly similar behaviors may have highly dissimilar underlying causes and functions… [the overt symptom models] simply identify the reliable co-occurrence of symptoms in the same individual. (Summerfeldt et al., 2004, pp. 1463–1464)

These limitations have supported the development and exploration of a newer "core dimensions" model (Summerfeldt, 2004; Summerfeldt et al., 2014). The core dimensions model contextualizes OCD variance through dimensional measures of two underlying symptom motivators—i.e., the reasons people feel driven to utilize OC behaviour—rather than a focus on the overt symptoms or behaviours themselves. The core dimensions model describes: (1) harm avoidance (HA)—the heightened sensitivity to threat, avoidance of feared negative outcome(s), and anxious apprehension; and (2) incompleteness (INC)—the subjective perception of things being unfinished, unbalanced, imperfect, or not "just right" (Summerfeldt, 2004; Summerfeldt et al., 2014). Both facets represent trait-like dimensions that extend across clinical and non-clinical cohorts (Summerfeldt et al., 2014), and while typically positively correlated, are recognized independently as distinct co-occurring constructs (Mathes et al., 2019; Pietrefesa & Coles, 2008; Summerfeldt et al., 2014). Summerfeldt et al. (2014) have created measures to assess the core dimensions at both the trait level (i.e., how people are most likely to

react), and at the state level (i.e., assessing the subjective motives underlying specifically identified symptoms).

Despite a rapid adoption in the literature, the relative recency of the core dimensions model means that it lacks the same extensive replication as seen from the overt symptoms model. This is especially true for the construct of incompleteness, as trait harm avoidance has been explored in different contexts as a transdiagnostic construct of psychopathology more broadly (e.g., Cloninger, 1986; Cloninger & Svrakic, 1997; Pelissolo & Corruble, 2002; Wiborg et al., 2005). Until more recently, harm avoidance was considered a requisite component (i.e., as a driver of anxiety) in the formation and perpetuation of the OCD symptom cycle (e.g., Taylor et al., 2007; Salkovskis, 1985). The inclusion of incompleteness has been an important development in understanding OC symptomatology and the underlying mechanisms of the disorder; especially when considering certain historically counterintuitive findings such as observing that approximately 40% of clinical patients commonly reported not having a feared consequence associated with their symptoms (e.g., Foa et al., 1995, as cited in Pietrefesa & Coles, 2008) which stood in direct contrast to the prevailing conceptualization of OCD as an anxiety-dependant disorder. Despite its comparative recency, incompleteness has gained considerable interest over the last two decades (e.g., Belloch et al., 2016; Horncastle et al., 2020; Lee & Wu. 2019; Summerfeldt et al., 2014; Summers et al., 2020; Taylor et al., 2014), and has been found to be closely related (and, consequently, sometimes used synonymously) with the construct of not-just-right experiences (NJREs; e.g., Belloch et al., 2016; Coles et al., 2005; Melli et al., 2020).

While the overt symptom and core dimension models have typically developed independently within the literature, both core dimensions of INC and HA are thought to be strong candidate endophenotypes for OCD and related behaviour. Taylor (2012) summarizes endophenotypes as intermediary variables in a causal chain that link genetic and/or environmentally etiological variables to disorders, which, if identified, can facilitate understandings of the disorder's causal factors (p. 259). This suggests HA and INC may be describing constructs that represent a middle ground—and, by extension, may be possible causal mediators—between genetic predispositions (i.e., genotype) and overt symptoms (i.e., phenotype; Belloch et al., 2016; Bey et al., 2020; Kloosterman et al., 2013; Sica et al., 2016; Summerfeldt, 2004; Summerfeldt et al., 2014). As possible endophenotype variables, it is proposed that these core dimensions underlie or cut across overt symptoms (Summerfeldt, 2004; Summerfeldt et al., 2014), and thus the core dimensions model may be better conceptualized on a different level or plane than the overt symptoms model—existing between genotype variables and observed phenotype symptoms, rather than as a competing model altogether. This theoretical relationship notwithstanding, each models' fitness as clinical resources or diagnostic specifiers is arguably more reliant on how well they each (or together) describe organization(s) of symptom variance in ways that convey or lead to meaningful (i.e. practical or actionable) clinical/research differences outside of the models themselves.

Meaningful Associations with Clinically Relevant Variables

Psychologist Dr. David Clark made the claim that "if individuals can be categorized into homogenous symptom types, it may be possible to look for a common etiology and develop specialized treatment protocols for particular OCD [groups]"

(Clark, 2006, p. 17). While overt symptom models and the core dimensions model are interesting from a purely theoretical and observational standpoint, as Dr. Clark suggests, their primary utility arguably rests on an ability to communicate functional or clinical differences and/or demonstrate meaningful associations with other useful variables. This line of approach in research—the exploration and identification of unique correlates for these models of symptom variance—has already identified numerous promising findings spanning the biopsychosocial spectrum. I will briefly highlight some examples of these correlates—both for the overt symptoms model, and then for the core dimensions model.

The Overt Symptoms Model

Due to the relative recency of the core dimensions model, most literature exploring important correlates with differing categories/dimensions of symptoms have done so using overt symptoms. Associations have been established between organizations of symptoms and many clinically relevant variables. From a biological approach, it has been demonstrated that dimensions from the overt symptom model have both shared and unique genetic risk and heritability potential (e.g., Alemany-Navarro et al., 2020; Alsobrook et al., 1999; Hasler et al., 2007; Leckman et al., 2003; Pinto et al., 2008; for recent reviews, see Mahjani et al., 2021; Strom et al., 2021) suggesting that variance observed in specific symptomatology (e.g., the overt symptom dimensions), and the heritability of such symptoms, might be due—at least in part—to unique (poly)genetic differences and heritability. Put simply, certain genetic predispositions may lead to unique symptom outcomes.

From another biological perspective, when looking at neural function using magnetic resonance imaging, Mataix-Cols et al. (2004) found evidence for associations

with unique brain regions among the different overt symptom dimensions, and van den Heuvel et al. (2009) found differences in grey matter volume from different brain regions that were uniquely associated with the overt symptom dimensions. Relatedly, evidence is slowly building to show potential differences in neuropsychological performance specific to symptom clusters (see Cameron et al., 2020; McGuire et al., 2014); together, such findings demonstrate that differences in overt symptoms may be related to unique brain structure and/or (dys)function.

From an approach that considers possible causal factors, a recent study found evidence that specific types of experienced trauma may inform and shape symptom profiles (Ojalehto, 2023). Grisham & Roemer (2004) found support for childhood influences, including environmental, developmental, and behavioural, to be uniquely predictive of later OC-symptom outcomes. Differences in clinical trajectory have also been identified, with Kichuk et al. (2013) demonstrating that symptoms within the symmetry dimension were associated with earlier age of onset (a finding with a rich and consistent history of related replication, e.g., Rasmussen & Eisen, 1992), whereas unacceptable thoughts were associated with greater waxing-and-waning of symptoms over time.

Finally, treatment outcomes also seem to differ among groupings of like symptoms. Specific to pharmacotherapy, Landeros-Weisenberger et al. (2010) found that patients with harm, religious, and sexual obsessions, and checking-related compulsions, responded more favourably to serotonergic medications (SRIs), whereas other groups found that symptoms marked by sexual/religious-based obsessions are especially resistant to behavioural therapy and SRIs (Alonso et al., 2001; Ferrão et al., 2006). Stein

et al., found that symptoms relating to symmetry seem to be markedly resistant to first-line pharmacotherapies (Stein et al., 2007; Stein et al., 2008). In psychotherapy, greater adherence to treatment homework (e.g., between-session exposure therapy) predicted more favourable outcomes for patients experiencing symptoms from the overt symptom dimensions of harm, unacceptable thoughts, and symmetry, but not for those with symptoms relating to contamination (Ojalehto, 2020). Conversely, other groups found no overall differences in treatment outcomes across symptom dimensions (Chase et al., 2015). For more on therapy response, see Sookman et al. (2005) and Williams et al. (2013).

The Core Dimensions Model

Emergent research has also begun to explore important correlates as organized through the core dimension model's two factors of HA and INC. Even before the official conceptualization of the core dimensions model, the concepts of HA and INC in relation to OCD have been enduringly evident. For example, it has been a long-standing belief that anxiety (i.e., a future-based fear of negative outcomes, i.e., HA) was a fundamental feature of the OCD symptom cycle (e.g., Taylor et al., 2007; Salkovskis, 1985) and thus many benchmark therapies were built around addressing the role of anxiety and those feared outcomes (e.g., Foa & Kozak, 1986; Tolin, 2009). HA has also been implicated across a number of psychopathologies (e.g., Cloninger & Svrakic, 1997; Pelissolo & Corruble, 2002; Wiborg et al., 2005), has established associations with the personality trait neuroticism (Weyers et al., 1995), and may be associated with serotonergic activity (e.g., Hansenne & Ansseau, 1999; Melke et al., 2003).

HA and INC both predict OC psychopathology; however, it may be that INC is a unique requisite to OC-related disorders, whereas HA may be more transdiagnostic (Belloch et al., 2016; Lee & Wu, 2019). A meta-analysis by Taylor et al. (2014) supported this idea by finding that INC was more uniquely related to OC-symptoms, whereas HA was equally related to OC-symptoms and general distress.[2] Relatedly, it has been found and replicated across numerous independent studies that INC has a strong unique association with overall OCD symptom severity (e.g., Sibrava et al., 2016); and has a stronger association with overall symptom severity than HA when directly compared (e.g., Belloch et al., 2016; Ecker & Gönner, 2008; Lee & Wu, 2019). HA and INC also seem to be differentially associated with categories of overt symptoms. For example, using different measures in both clinical and non-clinical samples, it was found that symptoms relating to symmetry, ordering, and personally-prescribed perfectionism were associated with INC, but not HA (Ecker & Gönner, 2008; Pietrefesa & Coles., 2008) that obsessional thoughts predicted HA but not INC (Ecker & Gönner 2008; Pietrefesa & Coles., 2008); and that checking (Ecker & Gönner, 2008), contamination, washing, and mental neutralizing (Lee & Wu, 2019; Pietrefesa & Coles, 2008) were associated with both INC and HA. HA and INC may also be differentially associated with facets of neuropsychological performance. Cameron & Summerfeldt et al. (2019) found that participants high in HA demonstrated greater deficits in verbal memory than those high in INC, whereas the INC group was more associated with deficits in certain measures of executive function such as problem-solving, planning, and set-shifting abilities. These findings together have prompted some to question whether INC and HA

[2] The second follow-up study within Taylor et al. (2014) using a non-clinical sample did not replicate the finding from their meta-analysis.

might play different mechanistic roles in the OC-cycle—that HA may initiate symptoms and/or prompt compulsive action, whereas INC might play the role of signaling when the utilization of compulsions is complete or 'successful' (e.g., Ecker & Gönner., 2008; Summerfeldt et al., 2014).

In light of these shifts in understanding, contemporary perspectives are beginning to explore the unique etiological role of INC (see Summerfeldt, 2004, pp. 1158–1160) and are demonstrating that those who lack explicit feared consequences (e.g., those with symptoms primarily driven by INC) may require different or modified approaches to treatment (Foa et al., 1999; Mathes et al., 2019; McKay 2020; Summerfeldt, 2004; Summerfeldt, 2007). Supporting this belief, Cervin & Perrin, (2021) found that youth high in INC and disgust symptoms tend to fare less favourably in current treatment paradigms, and a meta-analysis by Schwartz found early evidence for benefits of tailored treatments that address incompleteness, with increased efficacy seen from those explicitly focused on INC and/or that used measures designed to assess the core dimensions (Schwartz, 2018). This is an important consideration as there is evidence to suggest that those high in INC-related symptoms may be uniquely associated with an earlier age of onset (Rasmussen & Eisen, 1992; Summerfeldt et al., 2002), and the importance of effective early interventions for those with OCD is well established (e.g., Fineberg et al., 2019).

As a final mention, transcranial magnetic stimulation may have unique benefits for those higher in INC (Mantovani et al., 2013) and, while not explicitly indicating INC, Jenike et al. (1997) found unique benefits from the monoamine oxidase inhibitor phenelzine for those with symmetry-related symptoms, but not for those with anxiety.

Together, this suggests the possibility of unique biological or neurological bases for symptoms associated with INC. Furthermore, genetic research by Smit et al. (2020) found that only compulsions were associated with genetic bases, whereas obsessions were not. While this alone does not speak to specific genetic underpinnings for the core dimensions, when presented alongside the emerging research suggesting obsessions and compulsions may be differentially related to HA and INC respectively (e.g., Ecker & Gönner, 2008; Pietrefesa & Coles, 2008; Summerfeldt et al., 2014; Taylor et al., 2014), this may prove to be a valuable direction for further research.

The above examples broadly highlight the breadth of important associations between each model and variables that are important to clinical practice and research. While insufficient at this stage to infer causal or directional relationships, such emergent findings demonstrate the value of more detailed conceptualizations of the disorder than is reflected by the current diagnostic nosology. If conceptual symptom models such as the overt symptom or core dimension models can reliably account for variance that explains unique biological correlates, predict symptom trajectory, and/or inform treatment efficacy, then their value in continued research and clinical use becomes compellingly apparent. One common limiting factor in these studies is their cross-sectional nature, meaning there are inherent limitations to inferring causal relationships, or in knowing how the relationships fare longitudinally.

Symptom Models from the Perspective of Longitudinal Stability

This question of progression, or the longitudinal stability of these structure models, is an important characteristic to explore, because the models' clinical utility and the theory that informs their very structure are largely predicated on an assumption that

they capture, describe, and/or are outcomes of, otherwise stable constructs (e.g., genotype, endophenotypes, and/or biological/neural structures). To illustrate this, consider the observations that symptoms strongly motivated by INC may uniquely benefit from purposeful adaptations to current leading psychotherapies (e.g., Schwartz, 2018; Summerfeldt, 2004). While a core dimension measure could easily be utilized as a screening measure prior to therapy, if it was found that within-person symptom motivations (i.e., the core dimensions) were highly variable, random, and unstable over time, the implementation of testing for them as further diagnostic specifiers might have limited clinical benefit. Conversely, if these symptom dimensions remain stable over time, or evolve, but in predictable ways, then an assessment for those meaningful symptom specifiers (be it the core dimensions model, overt symptoms model, or otherwise) may prove diagnostically informative. Second, as alluded to above, many of the associations found for these symptom dimensions include neural or genetic correlates, which may suggest an assumed causal relationship toward the symptoms themselves. For example, as mentioned previously, it might be that heritable genetic factors confer unique etiological predisposition to specific symptomatology (e.g., Mahjani et al., 2021; Strom et al., 2021). Similarly, despite what is known about neuroplastic adaptation, it might be that certain neural correlates precede or predispose unique symptom development. If such assumptions are correct—where the genotype uniquely anchors symptom-specific outcomes at the phenotype—then it might be expected that symptoms would reflect stability similar to that of the associated biological precedents. Third, stability helps to inform how these variables fit together, and the relationships that exist between them. For example, HA and INC have been

conceptualized as candidate endophenotypes (Bey et al., 2020; Kloosterman et al., 2013), suggesting a hierarchy placement between the genotype and phenotype. Common criterion for endophenotypes include an assumed causal relationship from the endophenotype to phenotype, and 'trait-like' temporal stability (e.g., Gottesman & Gould, 2003; Taylor, 2012), so exploring stability among these variables provides clues to their possible relationships and meta-structures. Finally, stability can help to evaluate the fit and validity of chosen model criteria. For example, findings of poor stability may be attributable to actual change among symptoms, or it may be more indicative of a poorly designed model, where the latent structures or criteria chosen to represent categories or dimensions are not capturing stable constructs. In this way, while not always clearly interpretable, model (in)stability can help to incrementally refine model validity and development.

Measuring Stability

Stability can be conceptualized in numerous ways. At a very basic level, stability can be observed at the individual overt symptom itself. In the studies within this paper, I refer to this level of change as "qualitative" change because it refers to changes within the more subjective experiences of symptoms, despite remaining consistent within larger symptom organization (i.e., dimensions). For example, if someone historically compulsively checked the stove for fear of their house burning down, and over time changed to checking the doors for fear of being robbed, their qualitative symptom experience has evolved (i.e., changed) despite their symptoms remaining associated with checking and motives of HA.

A second way to conceptualize change is at the level of each symptom category/dimension and takes into consideration scores across participants. Once symptom data have been organized into their respective dimensions, the stability of each dimension can be assessed individually using the scores from the study sample. For example, it might be that one symptom dimension (e.g., harm/injury/bad luck) is likely to remain stable over time, whereas another (e.g., unacceptable thoughts) may be more variable. This is typically how stability has been assessed for this topic thus far in the literature and dimensional stability is typically explored using three main analytical approaches. The first analysis examines changes in mean dimension scores (also referred to as absolute stability), which indicates if average levels of that dimension have meaningfully changed over time (i.e., if severity or number of symptoms has meaningfully increased or decreased). The second analysis is through rank order or 'test–retest' correlations of all participant scores within a dimension from time 1 to time 2 (also called relative stability). The resulting coefficient from this analysis indicates how stable participants' positions relative to other participants remain over time; higher positive correlations indicate the dimension demonstrates more universal trajectories of stability/change, whereas lower correlations indicate the dimension is more variable from person to person over time. The third dimension-level measure of stability typically utilizes regression analyses to see if the time 1 scores on a dimension uniquely predict that same dimension at time 2 while controlling for the other symptom dimensions as additional independent variables. This indicates whether there are sample-level trends in change from one dimension to another over time, or whether symptom severity within a dimension at time 1 is indeed the best (i.e., a unique) predictor of that same dimension at

time 2. Put more simply, this approach explores whether participants, on average, 'change' from one dimension to another over time.

A third way to conceptualize change looks at how symptom dimension profiles—or the unique personal configuration of dimension scores within each participant—evolve over time. This is referred to as ipsative stability (e.g., Atherton et al., 2022; de Fruyt et al., 2006) and it has yet to be explored in OCD symptomatology. Despite its novel nature, it is a unique and important conceptualization of stability to consider, especially if a central tenet of the theory behind these models is that overt symptoms and the core dimensions are partially determined by stable constructs like the genotype. If this tenet is true, one could expect that the symptom profiles for each individual should remain relatively stable over time—as the underlying causes for each dimension are themselves stable constructs. This suggests ipsative stability might be one of the most directly informative measures of stability in this context.

Stability of OCD Symptoms in the Literature

Thus far, few studies have explored symptom stability, and fewer still using the overt symptom or core dimension models. Seminal work by Rettew et al. (1992) attempted to track symptom stability in a group of children and adolescents over an average of 8 years. In this study, they used the a-priori Y-BOCS categories (8 obsession categories and 7 compulsion categories) and tracked average number of categories endorsed across the study. Their findings suggested that OCD symptoms changed both in content and severity over time. More specifically, the authors noted that everybody from their study changed in their symptom "constellations" at follow-up, that overall number of obsessions lessened over time whereas number of compulsions stayed the same, and

that there were no statistical differences (i.e., notable patterns) in how symptoms changed between their four most common-found symptom types. Because the content change analyses were primarily conducted using group-average data, there was low empirical resolution for person-level change. It was also unclear exactly how the authors managed individual (i.e., item-level) symptoms. Regardless, the authors noted that many people reported within-category changes (what I refer to as qualitative change) that were not officially recorded due to the changed symptom still being part of the same higher-order category. This means their study design left for a significant amount of change at the qualitative level to go unrecorded. Another broad look at symptom progression was published by Skoog and Skoog (1999) who performed an impressive 40-year longitudinal study. They observed that participants typically saw improvements (i.e., lessening) in symptoms despite largely maintaining clinical or sub-clinical levels over time. While generally focused on more prognostic measures, the authors anecdotally noted that symptom breadth seemed to be stable (i.e., those loading uniformly on narrow symptoms, what they called "monosymptomatic", were not likely to become "polysymptomatic"). They further reported those with both obsessions and compulsions were more likely to maintain their disorder than those with just obsessions, and that two-thirds of their cohort reported obvious symptom changes across the study.

Beyond these broader findings, there are two notable studies that focused on exploring overt symptom stability by observing change using models other than the overt-symptom or core dimensions models. Fullana et al. (2007) utilized the Revised Obsessive-Compulsive Inventory (OCI-R; Foa et al., 2002) with non-clinical undergraduates over two years. At the group level, no significant differences were seen in

any of the six OCI-R subscales from time 1 to time 2 except for obsessions, which dropped over the two-year period; suggesting symptoms, as measured, generally remained stable. Furthermore, when assessing the predictive potential of each subscale, the only significant predictors for time 2 scores were the same subscales' time 1 scores, prompting the authors to conclude that the content of symptoms remained stable over time (i.e., that there were no major group level patterns of change from one measure subscale to another over time; p. 822). The second related study is interesting because it highlights the importance of how model organization can confer differences in study outcome. Besiroglu et al. (2007) studied symptom stability in clinical patients across an average of three years both at the level of the a-priori Y-BOCS categories, and then after dividing obsessions into three higher-order categories: (1) autogenous (i.e., occurring on their own accord); (2) reactive (i.e., in response to a state-specific precipitant); and (3) mixed. The study found significant within-category changes for most of the a-priori categories from time 1 to time 2 (similar to findings from Rettew et al., 1992). However, when observing obsessions as categorized by autogenous, reactive, or mixed, symptoms remained stable. The authors conceded that by assigning overt obsessions to larger constructs, the claim of stability still allowed for within-category change. Together, these findings demonstrate the effects (for better or for worse) of different model architecture on findings of stability.

Stability in the Overt Symptoms Model

Only four studies were found that explored stability utilizing the overt-symptoms dimensional model. These are the literature most related to the studies within in this paper. The first was published by Mataix-Cols et al. (2002) with a population of clinical

patients across a two-year study window. Here, the authors coded the main thirteen a-priori Y-BOCS symptom categories from 0–2 based on symptoms within those categories being not present (0), present but secondary symptoms (1), or present and primary symptoms (2). The scored a-priori categories were then further combined to form dimensions representing the overt symptoms model and the scores from the combined categories were summed to represent each dimension score. Analyses such as the McNemar's test, ANOVA, and multiple regressions were used to explore stability based on changes in the a-priori and overt symptom dimension mean scores, and the ability for each dimension to uniquely predict the same dimension over time. At the level of the a-priori Y-BOCS categories, most participants were likely to retain symptoms over time (i.e., the scores for the categories remained statistically unchanged). At the level of the overt symptom dimensions, based on the summed three-point scoring metric, the authors reported changes within the aggressive, symmetry, and contamination dimensions. Finally, similar to the findings from Fullana et al. (2007) the authors found that time 1 scores for the overt symptom dimensions were the best predictors of time 2 scores after controlling for the other dimensions. These overall findings were later replicated using comparable analytical approaches by Rufer et al. (2005) with an adult clinical cohort over an average of six years; Delorme et al. (2006) with a youth cohort for an average of nearly four years; and de la Cruz et al, (2013) with a cohort of youth across an average of five-years. Across these studies, if dimension changes occurred, they were typically related to contamination/cleaning (de la Cruz et al., 2013; Delorme et al., 2006; Mataix-Cols, 2002; Rufer et al., 2005); symmetry/ordering; and aggressive/sexual symptoms (Mataix-Cols, 2002; Rufer et al., 2005). Based on these findings, the

respective authors broadly converge on the conclusion that despite change observed within dimensions over time, overt symptom content stability—or the propensity for dimensions at time 1 to uniquely predict the same dimension at time 2—appeared to be significant and replicable.[3]

Due to the designs used in the above studies, there are notable limitations in their findings. One of the common methods used to assess stability (Besiroglu et al., 2007; de la Cruz et al., 2013; Mataix-Cols et al., 2002; Rufer et al., 2005) builds on the process mentioned prior relating to assigning dimensional scores based on categorical measures. The process starts by assigning severity ranking scores (0–2) to the a-priori Y-BOCS categories based on identification of any primary or secondary item-level symptoms experienced within them. For example, if one of the 10 item-level symptoms (e.g., fear might harm others) under the a-priori category label of 'aggressive obsessions' was reported as present and identified as a 'primary' symptom, the entire category of aggressive obsessions would be given a score of two. Through this approach, each category is reduced and scored with very little sensitivity to what else may be occurring within the item level of overt symptoms over time. In these studies, the related a-priori categories were then further combined to represent the dimensions forming the overt-symptoms model; their 0–2 scores summed to calculate the total scores for each dimension, and stability assessed through analyses using those summed scores over time. This means that the final analyses were using latent dimension score values up to three

[3] In instances where there were observed cross-dimension correlations using the regression analyses (e.g., de la Cruz et al., 2013; Rufer et al., 2005), the changes were associated with the dimension of hoarding, which is now considered a separate diagnosis and thus the findings were not addressed within this review. Mataix-cols et al. (2002, see table 2) observed cross-correlations within truncated sections of their study window, but not from baseline to the end of the study.

steps removed from specific overt symptoms. This approach would allow for change within the item-level symptoms (i.e., qualitative changes in fears about harming others), change within the a-priori categories (i.e., among any of the other eight item-level symptoms under ‘aggressive obsessions’, so long as they retain the same primary/secondary rating from time 1), and dimension total scores (i.e., changes among the related a-priori categories within the final dimension, so long as their summed scores remained relatively unchanged) to go unrecorded. A second but related limitation was seen in studies that quantified the a-priori symptom categories by summing the number of endorsed item-level symptoms they represented (e.g., de la Cruz et al., 2013; Delorme et al., 2006; Rettew et al., 1992). In these cases, instead of the 0–2 categorical scoring system mentioned above, this approach recorded the number of item-level symptoms that were reported within each a-priori category and used those summed scores as proxies for symptom severity within that category. The issue with this approach is that changes in item-level symptoms can still go unrecorded. Using the previous example, if two of the eight symptoms within the category of aggressive obsessions are reported at time 1, and two totally different symptoms within the same category reported at time 2, the category is still considered perfectly stable, as the category score remains the same despite significant item-level change. All the above studies were also limited in their ability to represent severity within each dimension, as reliance on number of symptoms experienced within categories is not necessarily equivalent to levels of subjective distress or impairment (e.g., Abramowitz et al., 2010; Summerfeldt et al., 2004), and the coding of symptoms as not present/secondary/primary severely limits the range of measurement.

Together, these design limitations mean that an even higher amount of within-dimension change may be occurring than was already captured and recorded in the prior studies.

A third potential issue worth mentioning was the use of prospective windows for data collection. While prospective studies are arguably superior to other methods such as retrospective designs for many well-established reasons, one counterpoint is to question whether the study windows were long enough, or timed appropriately, to adequately capture symptom development or change. While these studies ranged in their longitudinal timelines, given the chronic nature of OCD, even generous testing windows might be limited as they may not capture or span unique transitional periods in the participants' disorder trajectory. Coles et al. (2011) found that OCD symptoms tend to manifest in distinct developmental phases. Thus, it is likely that as OCD progresses, symptoms may present differing stages of stability over the lifespan as is seen with other trait-like constructs such as those related to personality (e.g., Borghuis et al., 2017). As most of the participants in these previous studies were already diagnosed clinical patients with established symptoms and pathology, the reports of stability might not be generalizable to a broader lifetime symptom course.

Finally, in the above studies, stability was only ever observed at the level of each dimension or category (i.e., across participants) and not as a connected or interrelated profile of symptoms as it manifests for the individual (i.e., across dimensions). While these observations are important and informative, if the prevailing theory is that overt symptoms might be determined by unique genetic, neural, or developmental antecedents, then the within-person (i.e., ipsative) dimension profile stability would be equally, if not more, important to observe. To illustrate this point, because these models' dimensions

show unique clinical correlates, to answer the question of how useful the models might be as diagnostic specifiers, the most important measure of stability is to observe how the dimensions might change relative to each other over time; not how the dimensions together or in isolation may become more, or less, severe.

Stability in the Core Dimensions Model

Limited findings for stability using the overt symptoms models notwithstanding, less still is known about longitudinal stability of the core dimensions. The construct of trait HA outside of the context of the core dimensions model, and as operationalized by other measures, has shown to be stable. For example, Josefsson et al. (2013) observed high test–retest rank-order stability (r = .70–.82) and no significant mean-level changes in a sample of 3596 non-clinical patients across 27 years. There is limited published research exploring the longitudinal stability of INC or NJREs, with longitudinal analyses typically limited to the context of psychometric assessments of specific measures. For example, Ghisi et al. (2010) reported a one-month test–retest window for NJREs showing a reliability (i.e., rank-order stability) of r = .76, and Sica et al. (2012) explored six month and one year test–retest windows for NJREs using the same measure and found rank-order stabilities of r = .70 and .75 respectively. No published literature to date has examined symptom stability using the two-factor core dimensions model at either trait or state levels. Unpublished longitudinal data (Till et al., 2024) suggest the individual core dimensions when measured at the trait level may show similar stability to other well-established personality traits over a 15-year period, with significant semi-partial correlations found for both HA (r = .23) and INC (r = .38). No published research has examined stability of the core dimensions when measured at the state level.

Study Overviews

This paper presents two studies designed to explore questions presented in this introduction, namely: (1) the possible associations between the overt symptom and core dimension models, and (2) the question of stability for each. Both studies used the same derivation sample and were hosted online. Study 1 used data from self-report measures, and Study 2 used data from a semi-structured one-on-one interview conducted via video teleconferencing. Study 1 explores the cross-sectional associations and predictive potential of the two core dimensions of trait HA and INC compared to the contemporary four-dimension overt symptoms model comprised of symptoms related to: (1) germs/contamination; (2) harm/injury/bad luck; (3) unacceptable thoughts; and (4) symmetry/completeness/'just right' feelings. Study 2 extrapolates longitudinal data about the overt symptom and state-measured core dimension models through retrospective questioning and examines the progression of each model using various conceptualizations of stability. Implications, limitations, and future directions for each study are discussed.

Chapter 2: Study 1 – Predicting Overt Symptoms in OCD: Examining the Utility of the Core Dimensions Model

High heterogeneity of symptom variance in obsessive-compulsive disorder has motivated the development of an empirically supported symptom nosology. From this effort, two primary models have emerged. The first is an overt symptoms model (see Bloch et al., 2008) which describes four dimensions of observable symptoms (i.e., the obsessions and compulsions) based on their propensity to coexist. The contemporary labels for these dimensions used for this study are (1) germs and contamination; (2) harm, injury, and bad luck; (3) unacceptable thoughts; and (4) symmetry, balance, and 'just right' feelings (Abramowitz et al., 2010). The second is the core dimensions model (Summerfeldt et al., 2014) which describes two dimensions of underlying motivations for symptoms: (1) harm avoidance (HA)—describing motivations relating to staving off harm, fear, or other perceived aversive consequences, and (2) incompleteness (INC)—describing motivations relating to achieving or maintaining symmetry, balance, or 'just right' feelings. Despite these models developing independently in the literature, the core dimensions are theorized to be endophenotypes for OCD—cutting across models of overt symptoms (e.g., Bey et al., 2020; Kloosterman et al., 2013; Summerfeldt et al., 2004), and thus may demonstrate predictive, causal, and/or unique relationships with overt symptom dimensions (see Taylor et al., 2012).

If it is the case that the core dimensions are endophenotypes and meaningfully predict or are associated with outcomes in overt OCD symptoms, then perhaps the core dimension model could be utilized not only as an important research development, but perhaps more directly to support clinical or diagnostic efforts (e.g., as screening

measures, as further diagnostic specifiers, or perhaps to inform more personally curated clinical approaches). While studies have explored associations between the core dimensions and specific overt symptoms (i.e., when organized by other means; Ecker & Gonner, 2008; Lee & Wu, 2019; Pietrefesa & Coles, 2008), none have yet explored associations between the core dimension and overt symptom models. Given the core dimension model's potential to address or mitigate limitations observed from the overt-symptoms model (see Summerfeldt et al., 2004), and the many noted associations between both models and other known and important clinical correlates, there is potential utility in understanding how these two models relate to each other; namely if the trait-like core dimensions predict differential outcomes in overt symptom expression. A finding of association between these two models would support the conceptualization of the core dimensions as candidate endophenotypes, would support the potential clinical utility of assessing or screening for the core dimensions, and may support the eventual integration of these models as further diagnostic specifiers.

Current Study

The current study aims to explore the relationship between trait measures of the core dimensions as measured by the Obsessive-Compulsive Core Dimensions Questionnaire–Trait (OC-CDQ-T; Summerfeldt et al., 2014) and the four-factor overt symptoms model as presented by the Dimensional Obsessive-Compulsive Scale (DOCS; Abramowitz et al., 2010). The DOCS is unique as it is the first measure designed to purposefully allow for truly dimensional measures of symptom severity for each of the four contemporary overt symptom dimensions. The DOCS was chosen for this study because: (a) it purposefully organizes symptoms by the four categories already prevalent

in the literature from the overt symptom models; (b) it does not rely on unwieldy symptom checklists, making it comparatively quick to administer and flexible enough to capture more abstract symptoms; (c) it does not confound the number or types of symptoms identified with symptom severity; and (d) it utilizes more nuanced multi-faceted items to represent symptom severity and dimensionality over the more typical 'yes/no' advocation of symptoms. Additionally, unlike the a-priori categories from the Y-BOCS where there are an inconsistent number of item-level symptoms, each dimension from the DOCS is derived through a consistent number of lower-level item questions. This means that each dimension is measured using the same standardized criteria, which makes direct comparisons between dimensions more consistent. Furthermore, the DOCS' separate and truly dimensional scoring structure, along with the accompanying total score, facilitates various approaches to data analysis. The OC-CDQ-T was chosen for this study because it represents the current gold standard approach for assessing HA and INC, especially within the context of OCD. The measure allows both dimensions to be assessed in tandem and through comparable means (i.e., HA and INC have the same number of items per dimension), and is an efficient and concise self-report measure, making it ideal for the online questionnaire format of this study.

Following the conceptualization of the core dimensions as endophenotypes, HA and INC will be examined as predictors of overall symptom severity, as well as for the four overt symptom dimensions. As no prior studies have explored relationships between the core dimensions model and the overt symptoms model—especially when considering the overt symptoms model as measured by the DOCS—hypotheses of the two models' respective associations were informed through previous studies examining overt

symptoms using other measurement tools. For this study, based on reviews of the existing literature, it is hypothesized that (1) trait HA and INC will both be significant predictors of overall symptom severity; (2) that trait HA and INC will uniquely predict the DOCS dimensions of harm and symmetry respectively; (3) that HA and INC will both significantly predict severity in the DOCS dimension of germs/contamination; and (4) that HA will be more uniquely predictive of the DOCS dimension of unacceptable thoughts.

Methods

Participants

Recruitment posters (Appendix D) and a short video describing the study (Till, 2023) were posted across several social media platforms including Facebook, Reddit, YouTube, and Instagram. Posts were shared, typically bi-weekly, within peer-run groups dedicated to supporting people with OCD. Furthermore, the Orchard OCD research organization (Orchard, 2024) shared the study advertisements with their online participant registry, along with posting on their website and social media channels. Data collection was open and active from February 2022 to February 2023.

Participants were internationally recruited, and inclusion criteria required participants to be at least 18 years of age and to have an existing formal diagnosis for OCD. Both inclusion criteria relied fully on self-disclosure. A total of 176 potential participants were recorded; removal of cases with incomplete data resulted in a final *N* of 90. This final sample was primarily female (*n* = 67, 82%), 88% of participants self-identified as White/Caucasian, and 80% reported being educated at a post-secondary level or higher (for a full overview of the demographics, see Table A1). Participants were

on average 33 years old (*SD* = 11.47) and the average age of onset of OCD (i.e., time-point 1 [TP1]) was 16 years (*SD* = 6.48). Participants had been experiencing symptoms for an average of 23 years (*SD* = 10.87) and the average time elapsed between their first symptoms and their diagnoses was 14 years (*SD* = 8.54).

Procedure

Study I focused on how participants' symptoms were currently experienced. Data were collected online using the Qualtrics XM Survey software. Upon clicking the link included in the advertisements, participants were directed to the study introduction page. Here, they were provided further information about the study process, screened for eligibility based on age and self-disclosure of a formal diagnosis for OCD, and asked to provide consent to participate. The first questions addressed basic demographic queries, along with more specific questions about notable age-related anchor points throughout their disorder. After that, participants were asked questions about how their symptoms were currently experienced using the DOCS. In the final section, participants were asked questions about trait levels of the core dimensions HA and INC using the OC-CDQ-T, along with questions designed to assess facets of emotional regulation (DERS-16), and alexithymia (TAS-20). Data from the DERS-16 and the TAS-20 were primarily reserved for the follow-up study (i.e., Study 2); however, answers from these two analyses were utilized in this sample to further validate their consideration as a clinical cohort. Descriptions for these two measures can be seen in the Study 2 methods section.

Upon completion of Study 1, participants were given the option to volunteer for the second stage of the study (Study 2). If not, participants were provided debriefing

information and thanked for their time. Participants were also given the opportunity to opt in to being contacted for further studies from the lab if they chose to do so.

Due to the often-distressing nature of OCD symptoms, care was taken to ensure participants felt as comfortable and autonomous in their participation as possible. Questions were purposefully broad (e.g., participants were never required to provide details about their symptom content), and it was expressly communicated that participants could choose to answer questions or not as they wished. Furthermore, the debriefing form provided an overview of the study purpose, recommendations for prior literature, contact information for the research lab and ethics board, and a link to the International OCD Foundation website (International OCD Foundation, 2024) for further access to trusted OCD-specific resources.

Measures

Dimensional Obsessive-Compulsive Scale. The Dimensional Obsessive-Compulsive Scale (DOCS; Abramowitz et al., 2010) is a 20-item self-report questionnaire designed to assess obsessive-compulsive symptom presence and severity. Unlike other OC symptom measures which typically take the form of symptom checklists, the DOCS outlines and describes four broad thematic symptom dimensions to the participant and allows them to self-identify and report symptoms (both obsessions and compulsions) they experience from each. The four DOCS dimensions are: (1) concerns about germs and contamination; (2) concerns about being responsible for harm, injury, or bad luck; (3) unacceptable thoughts; and (4) concerns about symmetry, completeness, and the need for things to be “just right”. These dimensions were designed

to represent the replicated overt symptoms model from the factor-analyzed Y-BOCS symptom checklist items in prior literature, excluding the hoarding factor.

The DOCS is the first known measure specifically designed to provide truly dimensional severity scores for the established overt-symptoms dimensions. Unlike previous attempts to provide dimensional representation based on number of item-level symptoms reported or summed categorical coding of primary/secondary symptoms, the DOCS synthesizes dimensional severity based on five facet measures of subjective distress within each dimension. Within the measure, participants are given examples of OC symptoms typical of each dimension and asked to think of their own experiences. Once they have identified whether they experience symptoms relevant to the dimension in question, participants are presented with five Likert-scale items designed to assess dimensional symptom severity. These questions assess: (1) time occupied by OC symptoms; (2) avoidance behaviours; (3) associated distress; (4) functional impairment; and (5) difficulty disregarding OC symptoms and behaviour. Scores for each dimension are summed (0–20) and a total score is created (0–80) to represent overall OCD symptom severity.

Abramowitz et al. (2010) concluded that reliability for the DOCS was at least equal to other widely used symptom measures such as the Obsessive-Compulsive Inventory (OCI-R; Foa et al., 2002). Subsequent studies have found test–retest coefficients of total score to be high in clinical samples (r = .81; López-Solà et al., 2014), with Cronbach's alpha values ranging from .87 to .95 for total scores (Eilertsen et al., 2017). The DOCS total scores have been shown to correlate strongly with total OCI-R scores (r = .86; López-Solà et al., 2014) and correlated as strongly as the OCI-R with

scores on the Y-BOCS. Validity of the DOCS dimensions were also supported through their convergence with the corresponding subscales of the OCI-R (Abramowitz et al., 2010). Furthermore, DOCS scores demonstrated sensitivity to change comparable to that of the Y-BOCS and the OCI-R, and the DOCS also demonstrated greater diagnostic accuracy than the OCI-R when distinguishing between clinical and non-clinical patients, and between OCD and other anxiety disorders (Abramowitz et al., 2010). Based on Abramowitz' analyses, a cut-off total score of 18 was suggested for differentiating OCD patients from non-clinical patients. The sample for this study returned excellent internal validity across the DOCS dimensions: germs/contamination (α = .96); harm/injury/bad luck (α = .95); unacceptable thoughts (α = .96); symmetry/order/'just right' (α = .95); and total score (α = .88).

Obsessive-Compulsive Trait Core-Dimensions Questionnaire. The Obsessive-Compulsive Core-Dimensions Questionnaire (OC-CDQ; Summerfeldt et al., 2014) is a 20-item self-report questionnaire designed to measure two underlying motivational core dimensions associated with OCD: incompleteness (INC)—describing motivations relating to achieving or maintaining symmetry, balance, or 'just right' feelings; and harm avoidance (HA)—describing motivations relating to staving off harm, fear, or other perceived aversive consequences. The OC-CDQ can take the form of either a state- or trait-based measure; the difference being the preface statement presented to the participant which either contextualizes the questions to pertain to how they would typically act (i.e., trait) or directs the questions toward specific symptoms (i.e., state). This study utilized the trait version of the questionnaire. Ten items of the measure relate to INC, ten relate to HA, and all items take the form of self-statements (e.g., "I must do

things in a certain way, or I will not feel right"). Answers are presented across a five-point scale ranging from "never" (0) to "always" (4), and scores are summed within each dimension to calculate the dimension score (0–40).

Prior studies demonstrated high internal consistency for each dimension in both clinical and non-clinical cohorts, with Cronbach's Alpha for HA ranging from .89 to .92, and INC ranging from .88 to .91 (Cervin & Perrin, 2019; Summerfeldt et al., 2014). Structural validity of the OC-CDQ (state-based version) was assessed and supported using a clinical sample, which demonstrated support for the use of a self-report questionnaire in this context (Summerfeldt et al., 2014). Relatedly, structural validity of the trait-based version was also supported using non-clinical samples (Pietrefesa & Coles, 2008; Summerfeldt et al., 2014; Taylor et al., 2014). Internal consistency was recorded within this study sample, with excellent Cronbach's alphas for both HA (α = .91), and INC (α = .89).

Analyses

Select demographic and measure descriptives were compared with extant literature findings to justify generalizability between this collected sample and validated clinical OCD samples. Next, zero-order correlations were generated to examine associations among measure outcomes. Spearman's *rho* was chosen over Pearson's correlations coefficients due to violations of bivariate normality. Finally, the predictive ability of the two core dimensions for dimensions of overt symptom expression was evaluated using multiple linear regressions. Five overall regression models were assessed—one for each of the four DOCS dimensions as dependant variables, and one for the DOCS total score. The repeated regression analyses justified consideration of

adjusted significance values to account for possible familywise error rate. Due to the limited sample *N*, and to allow for the identification of meaningful trends in the data despite the acknowledged limitations in statistical power, I chose to use the more powerful Holm-Bonferroni method (Holm, 1979) over the more conservative (albeit more common) Bonferroni correction (Dunn, 1961). Correlation and regression analyses were conducted using the JASP statistical software (Intel version. 0.18; JASP Team, 2023); power analyses were calculated with G*Power (version 3.1).

Results

Data Preparation

Participants typically either exited the survey after the first few questions or completed the questionnaire in full. In very few instances (*n = 4)* participants missed or chose not to answer individual items despite otherwise completing the study. Due to the relatively low *N*, the four participants with minimal item-level missing data were retained, and imputations based on case median scores from the (sub)scales were utilized (Raymond, 1986; Tsikriktsis, 2005). Median imputation was selected over the mean to minimize undue potential influence of outlying values, which were identified in three of the seven study variables: six were present in trait HA, one was present in trait INC, and one was present in the DOCS total scores. Due to the clinical nature of this sample, the dimensional nature of the measures (i.e., allowing for item-level reports of zero), and the lower sample size, the outliers were assumed to be naturally occurring and potentially meaningful. To test a more conservative approach, post-hoc exploratory imputation of the outlying variables to their closest related normally conforming values did not

meaningfully alter outcomes. Therefore, results reflect analyses with the unmodified (i.e., included) original outlier values.

Participant Representation

To best ensure findings would be relevant to clinical samples, the main inclusion criteria for this study required an existing diagnosis of OCD. However, as this relied on accurate self-disclosure, and no explicit diagnostic interviews were utilized for screening, it was important to determine whether key features of the participant sample were comparable to other validated clinical samples. While the average elapsed time between participant diagnoses and current age was 10 years in our sample, the chronic course of OCD (Bloch et al., 2013; Pinto et al., 2006; Skoog & Skoog, 1999) meant that this sample was still likely to remain clinically significant. To explore this assumption, the suggested clinical cut-off scores from the DOCS were applied: 78 participants (87%) from this sample met the suggested criteria for clinical significance based on their current self-reported symptoms (Abramowitz et al., 2010). Furthermore, two-sample t-tests based on available means, standard deviations, and study *N*s suggested overt symptom DOCS scores from this sample were comparable to, or greater than, other diagnostically validated clinical OCD samples (e.g., Abramowitz et al., 2010; Enander et al., 2012; Melli et al., 2015; for a full summary of measure descriptives, see Table A2).

When considering trait HA, the mean total score from this study (28.3, $SD = 8.5$) was statistically equivalent to the validated clinical OCD sample from Belloch et al. (2016), and significantly higher than non-clinical control cohorts from previous studies (Belloch et al., 2016; Irwin & Jones, 2017; Ólafsson et al., 2020; Summerfeldt et al., 2015; Summers et al., 2020). Trait INC from this study ($M = 24.97$, $SD = 8.25$) was less

clearly comparable with previous studies and seemed to exist somewhere between previously recorded non-clinical and clinical averages. The INC mean from this study was statistically higher than controls in most published literature (Belloch et al., 2016; Irwin & Jones, 2017; Summerfeldt et al., 2015;[4] Summers et al., 2020); however, another study reported control group means comparable or higher than was seen here (Ólafsson et al., 2020). It was discovered that different scoring metrics have been used for the OC-CDQ-T: one ranging from zero to four on the Likert scale, and the other ranging from one to five. Thus, it is likely that the study with higher control group means had used the different scoring metric; however, the Likert scoring values were not specified in the publication, and the authors could not be reached to confirm this assumption. Clinical OCD averages were similarly varied, with some verified clinical samples statistically comparable (Summerfeldt, 2014), and others significantly higher than what was found in this group (Belloch et al., 2016).

While not included in the main analyses for Study 1, descriptives from the TAS-20 and the DERS-16 were utilized here for their capacity to support validation of the sample's clinical status. Specific to Alexithymia, scores on the TAS-20 differed mostly from non-clinical controls in the difficulty with identifying feelings subscale (e.g., Aluja et al., 2020; Stivaleti Colombarolli et al., 2019) which were notably higher in this group. Total scores in this sample (M = 53.61, SD = 13.52) were most comparable to other psychiatric and medical patients (Taylor et al., 2003). When compared to other clinical OCD cohorts, this sample had statistically similar total alexithymia scores, but deviated

[4] Irwin & Jones (2017) and Summerfeldt et al. (2015) were confirmed to use the higher (1–5) Likert-scale coding. Statistical comparisons to the findings from these studies utilized modified mean scores to account for the difference in scoring metrics.

from past studies with notably higher difficulties identifying feelings, and notably lower difficulties with external oriented thinking (e.g., De Berardis et al., 2005; Roh et al., 2011). According to the measure's suggested cut-off scores, 36 (40%) of the study participants had no alexithymia, 22 (24%) had possible alexithymia, and 32 (36%) had alexithymia (Bagby, Parker & Taylor, 1994). Specific to difficulties in emotion regulation, this study sample scored statistically higher across all subscales than was elsewhere recorded for non-clinical populations (e.g., Bjureberg et al., 2016; Shahabi et al., 2020; Yiğit et al., 2019). Total scores on the DERS-16 (M = 56.55, SD = 13.52) were comparable to populations with identified anxiety disorders, borderline personality disorders, and 'severe' depression (Bjureberg et al., 2016; Burton et al., 2022).

Zero-Order Correlations

To overview the relationships within the dataset, an inclusive correlation table was generated (Table A3). Due to numerous observed violations of bivariate normality, Spearman's *rho* was used (Bishara & Hittner 2015; Calkins 1974). The first point of interest from the dataset was the significant moderate correlation between trait HA and INC (r_s = .42, p <.001). Given past literature, this relationship was expected; however, it raised potential concerns of collinearity for the planned subsequent regression analyses. Second, both HA and INC correlated significantly with total DOCS scores (r_s = .47, p <.001 and r_s = .57, p = <.001 respectively). Additionally, INC was weakly but negatively correlated with age at first symptoms (r_s = -.21, p = .05), and HA was weakly but negatively correlated with the time elapsed between first symptoms and current day (r_s = -.25, p = .019).

Multiple Linear Regression and Partial Correlations

Following best practice guidelines outlined in Walker et al. (2021), all parametric assumptions relevant to the planned regression analyses were explored. Assessments of skewness and kurtosis, linear relationships, heteroscedasticity, collinearity, normality of residuals, and independence of residuals were all found to be within acceptable ranges. Outliers were retained due to the clinical nature of this sample, the dimensional nature of the measures, and the lower sample size. Multiple linear regressions were performed using trait HA and INC as independent variables, with each of the DOCS dimensions and DOCS total score as dependant variables (Table A4; see Figure A1 for a visual comparison of regression β values). As an accompanying conservative approach due to previous observations of bivariate non-normality, Spearman partial correlations were also utilized to further confirm the outcome relationships.

When observing the regression model for DOCS total scores, HA and INC together significantly predicted 43.7% of the variance in overall symptom severity ($F_{2,87} = 33.77$, $p < .001$). HA and INC both positively correlated to the DOCS total scores, and both remained significant after controlling for the other, with INC demonstrating a slightly higher beta value than HA. Partial Spearman correlations for HA and INC in relation to the DOCS total scores were $r_s = .31$ and .47 respectively. The model for the DOCS dimension of germs and contamination predicted a significant 19.2% of the variance ($F_{2,87} = 10.33$, $p < .001$). In this model, INC was positively related to contamination symptoms ($r_s = .37$) and was the only significant predictor after controlling for HA. The model for the DOCS dimension of harm, injury, and bad luck, predicted 34.9% of the sample variance ($F_{2,87} = 23.33$, $p < .001$) which was the highest

explained variance of all the DOCS dimensions. HA was positively related to these symptoms (r_s = .46) and was the only significant predictor after controlling for INC. The model for the DOCS dimension of unacceptable thoughts predicted 7.7% of the variance in this sample ($F_{2,87}$ = 3.65, p = .03) which was still significant despite being the lowest of the regression tests. HA was positively related to unacceptable thoughts (r_s = .22) and was the only significant predictor after controlling for INC. Conversely, INC was negatively associated with unacceptable thoughts, but not significantly so. For the final DOCS dimension model, HA and INC significantly predicted 29.4% of the variance in symptoms relating to symmetry and 'just right' feelings ($F_{2,87}$ = 18.13, p <.001). INC was positively related to these symptoms (r_s = .48) and was the only significant predictor after controlling for HA. HA was negatively associated with these symptoms after controlling for INC, but not significantly so. The above reported accompanying Spearman partial correlations supported the findings of all models, with significance patterns directly matching those of the regressions and their associated partial Pearson correlations (see Figure A2).

Discussion

This study sought to explore relationships between: (a) the core dimensions of HA and INC; and (b) severity levels of overt symptom dimensions relating to germs/contamination, harm/injury/bad luck, unacceptable thoughts, and symmetry/completeness/'just right' feelings. Based on prior research and theory that HA and INC might be (1) lower-level constructs that cut across dimensions of overt symptoms, (2) possible endophenotypes for OC symptomatology, and thus (3) possible causally associated with overt symptoms, the core dimensions were oriented as

independent variables in regression models for each of the overt symptom dimensions, as well as for total overt symptom scores.

Prior to running the main analyses, measure descriptives were compared to previous published findings to establish whether claims of clinical parity would be appropriate in this sample. The age-related demographic data in this sample relevant to important symptom milestones (i.e., age of onset, age first diagnosed), were comparable to those from previous literature (e.g., Hezel et al., 2022; Ziegler et al., 2021). Furthermore, 87% of the study sample met the clinical cutoffs provided by the DOCS (Abramowitz, 2010). Mean scores from this sample across the DOCS, OC-CDQ-T, TAS-20, and DERS-16 typically exceeded established control scores, and were typically comparable to established means for other OCD and clinical samples. These findings support previous literature citing associations between emotion regulation difficulties and psychopathology (e.g., Compass et al. 2017; Lincoln et al., 2022; Sheppes et al., 2015), and for those with OCD (e.g., See et al., 2022; Yap et al., 2018). Notable deviations were seen in the TAS-20 subscales of identifying feelings, where this sample displayed scores notably higher than established OCD samples; and in externally oriented thinking, where this sample scored significantly lower (De Berardis et al., 2005; Roh et al., 2011). Another notable difference was observed in the core dimension of INC, where this sample was equal to some validated OCD samples (Summerfeldt, LPA data), but lower than others (Belloch et al., 2016). Despite these small deviations, descriptive scores from this sample suggested they were meaningfully different than controls, and comparable to other OCD and clinical cohorts.

Next, a zero-order correlation matrix was created to establish a broad overview of the relationships present within the data. INC and HA were significantly correlated in this sample (r_s = .42), which was higher than seen in previous clinical findings (e.g., r = .23; Summerfeldt et al., 2014), but much lower than was seen in non-clinical and/or student samples (cf., r = .76 and r = .93 from Pietrefesa & Coles, 2008 and Taylor et al., 2014 respectively). This finding may highlight how differentiation between these two variables is more likely to emerge as overall OC-symptom severity increases, which could be an important area for future study. Furthermore, correlations between INC and HA compared to participants' age at first symptoms, and the time elapsed to current day, support previous literature suggesting that INC and related overt symptoms are associated with earlier age of onset (e.g., Kichuk et al., 2013; Rasmussen & Eisen, 1992). Finally, the relationships between the core dimensions model and each of the DOCS dimensions and total scores were assessed using multiple regression analyses.

Hypothesis 1

Hypothesis one was supported, with HA and INC both significantly predicting overall symptom severity scores. This aligns with established literature regarding the joint role of both HA and INC in the manifestation of symptoms (Pietrefesa & Coles, 2008, Summerfeldt, 2004; Taylor et al., 2014). Relatedly, above both variables being significant predictors, it should be noted that INC was a stronger predictor than HA, which is also consistent with previous research highlighting a potentially unique and important association between INC and overall OCD symptom severity (e.g., Belloch et al., 2016; Ecker & Gönner, 2008; Lee & Wu, 2019; Sibrava et al., 2016).

Hypothesis 2

The second hypothesis was also supported, with HA and INC uniquely predicting symptoms relating to harm/injury/bad luck and symmetry/completeness/'just right' feelings respectively. This aligns with past research linking HA to other measures of symptoms relating to reducing harm, obsessing, and checking; and for INC with symptoms relating to symmetry and order (e.g., Ecker & Gönner, 2008; Pietrefesa & Coles, 2008; Pietrefesa & Coles, 2009; Summerfeldt et al., 2015). This result was further expected given the face-valid overlap of those specific DOCS and core dimension domains.

Hypothesis 3

The third hypothesis, that the DOCS dimension of germs/contamination would be equally predicted by HA and INC, was not supported. Instead, in this sample, contamination symptoms were only significantly predicted by INC (β = .42) and not by HA (β = .04). This was highly unexpected based on previous research which found them both to be significantly associated with contamination/washing symptoms (e.g., Lee & Wu, 2019; Pietrefesa & Coles, 2008)[5] or for washing/cleaning activities to be associated with or to elicit higher responses in HA (Ecker and Gönner, 2008; Pietrefesa & Coles, 2009). These results might indicate that different symptoms were subjectively attributed to the DOCS dimension of germs and contamination in this sample than were included in the contamination/washing/cleaning-related groupings from prior studies. If previous speculation is correct—where HA may be more associated with obsession symptoms,

[5] In both Lee & Wu, (2019) and Pietrefesa & Coles, (2008), zero-order correlations for contamination and washing respectively were higher for INC than HA. However, in both studies, HA and INC were still both significantly correlated with overt symptoms in those domains.

whereas INC is more associated with compulsions and/or signalling when they are complete (e.g., Ecker & Gönner, 2008; Summerfeldt et al., 2014; Taylor et al., 2014)—then it might have been that when asked to think about symptoms related to germs and contamination, participants were more likely to elicit cleaning- or avoidance-related compulsions than contamination-related fears or obsessions.

Hypothesis 4

The fourth and final hypothesis, that HA alone would be predictive of unacceptable thoughts, was supported. This was congruent with extant findings that obsessions were more strongly associated with HA (Ecker & Gönner, 2008; Pietrefesa & Coles, 2008). While symptoms relating to unacceptable thoughts do not exclude compulsions, the label of the dimension specifically referencing thoughts may have meant that participants were more likely to think about obsessions. In acknowledging the possibly unique associations between obsessions and HA (and compulsions with INC respectively; Ecker & Gönner, 2008; Summerfeldt et al., 2014), it may have been that HA was more prevalently associated with this dimension due to an overrepresentation of obsession-specific symptoms evoked through the DOCS. Even when compulsions may be present in symptoms relating to unacceptable thoughts (e.g., reassurance seeking/thought neutralization; Williams et al., 2011) it may be that participants are still more likely to align those symptoms with HA over INC, due to various maladaptive cognitive interpretations such as thought-action fusion (e.g., Shafran et al., 1996) which would equate the importance of thoughts/feelings to that of action/consequence.

Limitations

This study's findings should be interpreted in light of several limitations. Despite best efforts to capture a widely representative sample, the retained sample was limited in some demographic features. Even with an international recruitment approach, participants were overwhelmingly well-educated, Caucasian, and female. Moreover, reliance on social media forums may have introduced bias toward participants who are uniquely motivated and proactive about relying on the internet and community for information about OCD. Given the prolific use of the informal subtypes online as mentioned previously, this might mean that participants were more likely to have encountered or formed pre-existing expectations about how their OCD is structured, which may have influenced their approaches to answering specific questions (most specifically, a potential confound for the DOCS). Furthermore, participants were necessarily English speakers, and affluent enough to have access to internet services. Demographic information for the various social media groups were not available; however, the ratio of those who were confirmed to have seen the advertisements compared to the final numbers of volunteers, suggests a high potential for various volunteer-related biases. This is especially likely given the protracted period for recruitment; the low engagement supports an assumption that those who did choose to participate may be uniquely similar in some unknown variable(s) which facilitated involvement. A notable counterpoint is that those who participated in this study did so without any promises of compensation. With previous published literature, clinical participants are often on wait lists for treatment, which may prompt them toward extreme responding to demonstrate their dire need for support. Non-clinical samples are often

students who participate in research for course credit. Future studies would benefit from more proactive sampling across various demographically diverse groups, and groups with differing potential for perceived benefits or motivation to participate.

As mentioned prior, participants were also required to have existing OCD diagnoses; however, reliance on honest self-disclosure and the unknown history or quality of diagnoses meant that this sample may not homogeneously represent current clinical patients. Along with an unverified current clinical status, the lack of a diagnostic screening also meant that comorbidity rates within this sample was unknown. OCD is known for having a high co-occurrence with other psychological disorders (e.g., Sharma et al., 2021). Comparisons using validated metrics and previous literature norms suggested this group was comparable to other clinically validated OCD cohorts, but without the inclusion of an explicit diagnostic intake process, findings should be interpreted with a mindfulness to potential limitations of generalizability. Future studies may wish to include a diagnostic intake procedure or replicate within a known clinical cohort.

Findings of significance notwithstanding, a-priori power analyses suggested a target *N* of 92 based on typical effect size and power parameters ($f^2 = .15$, $p = .05$, *power* = .80). This means that the significant effects may indeed be attributable to true associations that were strong enough to be evident regardless of the relative lack of power, or it may reflect biases in the volunteer group toward certain tested parameters. Regardless, future studies would benefit from more participants to ensure that findings are more likely to be generalizable.

A final point worth noting relates to mindful interpretation of the results. While HA and INC are presented through the lens of a theoretical model which places them as lower-level (and by extension, possibly causal) constructs underlying overt symptoms, the cross-sectional design of this study cannot support inferences of directionality or causation. As candidate endophenotypes, it is logical to assume that trait HA and INC predispose individuals to developing specific overt symptom profiles; however, the primacy assumption inherent to that conceptualization would need to be explored through further empirical work. Relatedly, the approach of positioning HA and INC as separate 'predictor' variables may be an artificial abstraction. While it is certainly possible for someone to rate unitarily high on either variable (e.g., Summerfeldt et al., 2004, 2014), it is most common that individuals will experience some level of both (e.g., Taylor et al., 2014; Summerfeldt et al., 2014). This reality was reflected in the study sample through their moderate but significant correlation.

Conclusion

The results of this study suggest that in a multinational sample of 90 participants with a previous diagnosis of OCD, trait-level measures of the core dimensions HA and INC both significantly predicted overall symptoms of OCD and are further uniquely and differentially associated with dimensions of overt symptoms. Using the Dimensional Obsessive-Compulsive Scale—a truly dimensional measure of overt symptoms—and the Obsessive-Compulsive Core-Dimensions Questionnaire—a trait-level assessment of the core dimensions, trait HA uniquely predicted symptoms relating to harm/injury/bad luck, and symptoms related to unacceptable thoughts. Trait INC uniquely predicted symptoms relating to symmetry/completeness/'just right' feelings, and was unexpectedly also the

unique predictor of symptoms relating to germs/contamination. When considered together, HA and INC explained the greatest variance in harm-related overt symptoms, and the least for symptoms relating to unacceptable thoughts. When looking at overall symptom severity, both HA and INC significantly predicted total DOCS scores; however, INC was comparatively the stronger predictor overall.

It was unclear in this sample why INC was the only predictor of germ/contamination-related overt symptoms. However, this may have differed from previous studies due to the DOCS allowing for subjective allocation of symptoms within the given dimensions, as opposed to prior approaches which relied on pre-organized item-level symptoms. It was also unclear why the R^2 for the regression model exploring the core dimensions as predictors for unacceptable thoughts was so comparatively low. It may be that the overt dimension of unacceptable thoughts is truly unique in its score variance, or it may have been due to the limited number of study participants.

Regarding the current consideration of HA and INC as underlying predictors for overt symptoms in OCD, findings from this study support a continued focus on the core dimensions; especially for the dimension of INC which seems to be a strong predictor of overall OC-symptom severity. These findings support the assumption that the core dimensions cut across the dimensions of overt symptoms, and—taken with extant literature suggesting their potential for stability and their demonstrated presence across both clinical and non-clinical cohorts—continue to uphold the criteria for consideration as endophenotypes for OCD. If it is such that a causal or directional relationship exists between the core dimensions and overt symptoms, perhaps early assessments of the core dimensions could allow for more proactive or curated treatment approaches to intervene

before clinical symptoms emerge. The OC-CDQ-T is a quick and efficient measure that may demonstrate clinical utility as a screening tool. More broadly, these findings suggest that the core dimensions can be usefully incorporated alongside the overt symptoms model, and that further development and utilization of such structural models for OCD may indeed facilitate their use as supportive diagnostic resources and/or clinical specifiers.

Considering the established association between these two models, the endophenotype potential from the core dimensions, and the known cross-sectional associations found between each model and other important variables (e.g., etiology, symptom trajectory, and treatment), an important next step involves determining the long-term stability of symptoms as categorized by these models. Without relative stability, their use as potential diagnostic specifiers, and the importance of their established cross-sectional findings become increasingly less viable. The next study in this paper attempts to address these questions relating to stability.

Chapter 3: Study 2 – Organizing Symptoms of OCD: Assessing Stability in the Overt Symptom and Core Dimension Models

Due to high observed symptom heterogeneity in obsessive-compulsive disorder (OCD), two dimensional models have emerged which organize symptom variance in different ways. The first is a dimensional model that organizes symptoms based on their overt observable features (see Bloch et al., 2008). This overt symptom model groups symptoms across the four dimensions of: (1) germs and contamination; (2) harm, injury, and bad luck; (3) unacceptable thoughts; and (4) symmetry, balance, and 'just right' feelings (Abramowitz et al., 2010). The second is a model that describes two core underlying dimensions of symptom motivation: (1) harm avoidance (HA)—which describes motives relating to staving off harm, future-based fears, or other perceived aversive consequences, and (2) incompleteness (INC)—which describes motivations relating to achieving or maintaining symmetry, balance, or subjective feelings of things being (not) "just right" (Summerfeldt et al., 2014). These two core dimensions are further considered candidate endophenotypes for OCD (Bey et al., 2020; Kloosterman et al., 2013), meaning they are theorized to exist between genetic predisposing features for the disorder and overt symptoms themselves and are thus likely to play a causal role in overt symptoms (Taylor, 2012). Despite these theoretical associations, studies had yet to explore the relationship between the overt symptom and core dimension models.

The first study in this paper (Study 1) demonstrated that both trait HA and INC predicted overall OCD symptom severity. Furthermore, HA uniquely predicted overt symptom domains relating to harm/injury and unacceptable thoughts, whereas INC uniquely predicted symptoms relating to symmetry/completeness and

germs/contamination. These findings supported the theory that the core dimensions underlie overt symptoms, that they may be differentially responsible for expression at the overt symptoms, and that the two models are not mutually exclusive in their potential for research and/or clinical utility. Beyond these findings, important cross-sectional associations for the dimensions from each model have been established in prior literature. Important unique and shared correlates have been demonstrated between these model dimensions and: (1) genetic/heritability potential (Mahjani et al., 2021; Strom et al., 2021); (2) unique biological/neural correlates (Mataix-Cols et al., 2004; Van de Heuvel et al., 2009); (3) disorder etiology/progression (Kichuk et al., 2013; Lee & Wu, 2019); and (4) treatment efficacy (Cervin & Perrin, 2021; Sookman et al., 2005; Williams et al., 2013). Such findings make a strong case for these models' use in a clinical/diagnostic capacity; however, the studies' cross-sectional nature limits assumptions of longitudinal stability, which would be a meaningful requirement for practical use in a diagnostic context. Furthermore, specific to the core dimensions, stability is also a requisite criterion for their consideration as endophenotypes (e.g., Gottesman & Gould, 2003; for a review of endophenotypes in OCD, see Taylor, 2012).

Stability in OC-symptoms has been explored in prior literature; however, the number of such studies have been limited. Across the four existing publications that observed stability in the overt symptoms model, the authors generally concluded that the models represented stable dimensions—that each dimension over time was uniquely predicted by prior scores on the same dimension (i.e., that participants were not likely to 'switch' from one dimension to another over time)—but that within-dimension change was more common, especially among symptom groupings relating to

contamination/cleaning (de la Cruz et al., 2013; Delorme et al., 2006; Mataix-Cols, 2002; Rufer et al., 2005); symmetry/ordering; and aggressive/sexual symptoms (Mataix-Cols, 2002; Rufer et al., 2005). Despite these findings, several common methodological limitations have been observed: (1) the data reduction approaches used to group symptoms into categories and/or dimensions limited sensitivity to recording stability or change at other levels of organization (e.g., at the item-level symptom itself); (2) these same reduction approaches limited the studies' abilities to generate true dimensional representations of symptom severity at the overt symptoms level (for a summary of these reduction limitations, see Summerfeldt et al., 1999, pp. 298–299); (3) the analyses utilized by previous studies to determine if models were stable relied upon dimension-level (i.e., across participants) comparisons, and not within participant (i.e., across dimension; ipsative) stability; and (4) while not a specific methodological limitation—thus far, only unpublished data exist exploring stability of the core dimensions model (Till et al., 2024), and no known data exist for stability of the core dimensions when measured at the state (i.e., symptom) level.

Current Study

The current study aims to replicate the above findings for stability in the overt symptoms model while including and extending the same methods to explore stability from the core dimensions model. The study also aims to improve upon limitations of existing research by using truly continuous scores for overt symptom severity as afforded through the Dimensional Obsessive-Compulsive Scale's multifaceted approach to scoring symptom dimensions (DOCS; Abramowitz et a., 2010) as opposed to relying on the approaches from previous studies. More importantly, this study aims to incorporate

broader conceptualizations of stability through analyses not used in prior publications. Alongside the previous literature studies' exploration of mean-level change at the level of each dimension (i.e., across participants; absolute stability), this study will incorporate measures of qualitative (i.e., item-level) stability, and will be the first to include dimensional rank-order stability (i.e., across participants; relative stability) and within-person symptom profile stability (across dimensions; ipsative stability). The same analyses will be utilized for both overt symptom and core dimensions, marking the first known exploration of stability for the state-level core dimensions. Use of these additional analyses will provide a richer understanding of how and where symptom stability may exist at different levels of OCD symptom structure and reflects the growing utilization of these approaches to measuring stability in other dimensional or state/trait-like profiles as is seen in other areas of research such as personality psychology (e.g., Atherton et al., 2022; Poier, 2022; Terracciano et al., 2010).

Using a subset of participants from Study 1 ($N = 38$), longitudinal scores will be inferred via retrospective questioning in an interview format. Similar to Study 1, the core dimensions will be assessed; however, this time using a modified retrospective version of the DOCS. In light of the limitations of previous research, the DOCS is an advantageous measure in this capacity because it assesses symptom severity across multiple criteria (e.g., time spent engaging symptoms, difficulty not engaging with symptoms); it allows for a truly dimensional scoring of each symptom domain; it is quick to administer when compared to more in-depth symptom checklists; and it allows for participants to attribute symptoms that are subjectively relevant to them within the given symptom domains, allowing for less-common symptoms that may fail to be part of an established checklist

to still be recorded. A psychometric study by Kuckertz et al. (2021) found that within-person change reliability was good for the DOCS, and that the measure was "capable of capturing meaningful changes that exist within persons over time, and between-person change reliability was excellent" (pp.756, 761–762), further supporting its suitability for a study designed to make longitudinal inferences of stability.

State-level (i.e., symptom specific) measures of the core dimensions will be assessed using an interview-based measure developed by Summerfeldt et al. (2014) called the Obsessive-Compulsive Core Dimensions Interview (OC-CDI), which is ideal for this study because: (1) state-level stability of the core dimensions has yet to be assessed in the literature; (2) state-level scores of the core dimensions allows for a more interpretable comparison of stability when compared to the DOCS (i.e., as assessed using the same specific overt symptoms evoked for each DOCS dimension, rather than as an overall trait measure); and (3) the interview format allows for clarification and follow-up from both the participant and interviewer as needed, which is beneficial given the novel retrospective use of the measure. Stability will be assessed by comparing retrospective scores to participants' present-day scores; some of which will be carried over from the data collection from Study 1.

As a final step, the two emotion regulation scores assessed in Study 1: the Toronto Alexithymia Scale (TAS-20) and the Difficulty in Emotion Regulation Scale (DERS-16)—along with demographic variables, a measure of stressful life events, and a measure of treatment history—will be utilized in exploratory analyses to observe whether they are associated with outcomes of stability. The TAS-20 was included to highlight whether alexithymia or its subscales are associated with OCD symptom stability, as

alexithymia has been implicated elsewhere to have meaningful associations with OCD (e.g., Bagheri et al., 2020; Roh et al., 2011). Similarly, the DERS-16 was included due to established associations between emotion regulation difficulties, OCD, and psychopathology more generally (e.g., Yap et al., 2018). Demographic data was included due to known unique correlates with variables such as age of onset (e.g., Kichuk et al., 2013), and stressful live events due to the established waxing and waning symptom trajectory of OCD (e.g., Adams et al., 2018; American Psychiatric Association, 2013; Naftalovich et al., 2021).

Based on the reviewed research, it is hypothesized that (1) all participants will report some level of qualitative change over time; (2) that mean levels of the overt symptom model dimensions of (a) germs/contamination, (b) symmetry/completeness/ 'just right', and (c) overt symptom total scores will lower over time, whereas (d) harm/injury/bad luck, and (e) unacceptable thoughts will remain equivalent; (3) that mean levels of each core dimensions' scores will remain unchanged; (4) that all dimensions across both models will demonstrate moderate-to-high rank-order stability via test–retest correlations; (5) that both the overt symptom and core dimension models will demonstrate comparable ipsative stability over time; and (6) if/when one model demonstrates higher stability over the other, even if not statistically different, the core dimensions model will be the more stable.

Exploratory analyses will test whether retrospective dimension scores are unique predictors of the same dimension scores over time (an attempt to replicate previous literature suggesting participants do not 'change' dimensions; e.g., Mataix-Cols., 2002), and will further explore whether measures such as demographic features, emotional

competencies, life stress, and/or treatment history, confer predictive potential for symptom (in)stability.

Methods

Participants

Participants in this study represent a subset from Study 1 who chose to continue to the interview stage. Participants were internationally recruited online using social media forums and the Orchard OCD online registry (Orchard, 2024). Inclusion criteria required participants to be at least 18 years of age and to have an existing formal diagnosis for OCD. Seventy-nine participants indicated interest and requested to participate in Study 2. Of these, *N = 38* followed through and completed interviews. The final sample was primarily female (*n = 27*, 76%), and 89% of participants self-identified as White/Caucasian. Participants were an average of 35 years old (SD = 11.36) at the time of the interview (TP2) and the average age of onset (TP1) was 15 years (*SD* = 6.32; see Table B1 for sample demographics). All participants reported having used some form of treatment support, with slightly higher reported first-line pharmacotherapy use than from psychotherapies. Average total years utilizing support was 13 (*SD* = 11.01; Table B2). Based on this sample's reported age of onset, the longitudinal window of this study (TP1–TP2) represented an average of 20 years (*SD* = 12.35, *range* = 3–48). Despite a lower *N*, demographic and descriptive means from this sample were comparable to those from Study 1.

Procedure

After finishing Study 1, participants were given the opportunity to opt in to a virtual interview at a future date of their choosing. During the interviews, a brief period

was set aside for introductions, explanations, and for an opportunity to build rapport. Each online interview followed a structured format across three broad sections, facilitated by a shared-screen PowerPoint presentation with accompanying visual representations and the text for each measure. The interview format allowed for interaction and flexibility to clarify questions as needed. The first questions in the interview revisited answers from Study 1 by asking participants about the state-specific core dimensions (i.e., HA and INC via the OC-CDI) for their current symptoms (per dimension, as recorded by the Study 1 DOCS). In the second section, participants were asked questions about how their symptoms were experienced at the onset of their OCD (TP1), including the core dimension questions for that time (i.e., DOCS & OC-CDI). In the final section, participants were asked broad questions about stressful life events that may have happened between TP1 and current day (Life Events Questionnaire), along with questions regarding their treatment history (Treatment History Questionnaire). I personally conducted and coded all interviews.

As with Study 1, care was given to maximizing comfort and autonomy for participants given the study focus and interview format. Participants were not required to appear on camera for the interview, none of the included questions required detailed explanations of symptom or experiential content, and questions could be skipped, elaborated, or modified as needed to facilitate the interview process. After the interviews, participants were offered an opportunity to ask questions, and then were provided a formal debriefing form. Debriefing included an overview of the study purpose, recommendations for prior literature, contact information for the research lab and ethics

board, and a link to the International OCD Foundation website (International OCD Foundation, 2024) for further access to trusted OCD-specific resources.

Measures

Dimensional Obsessive-Compulsive Scale (Retrospective). The Dimensional Obsessive-Compulsive Scale (DOCS; Abramowitz et al., 2010) was utilized again in this follow up from Study 1. During the interviews, participants were verbally and visually guided through the DOCS a second time, but this time with the preface that they were to respond "…as if [they] were back in time to when [their] symptoms first started causing problems with daily functioning" (i.e., age of onset; TP1). Verbal and visual progression through the DOCS was provided to participants via shared screen, with flexibility to ask questions or clarify the retrospective focus as needed.

Slight modifications were applied to the original DOCS, appropriately reframing the tense of the questions from present to past; however, it should be noted that using the DOCS in a retrospective capacity has not been validated. Despite the novel use, internal consistency remained excellent in this sample, with high Cronbach's alphas for the dimensions of germs/contamination (α = .98); harm/injury/bad luck (α = .91); unacceptable thoughts (α = .97); symmetry/order/'just right' (α = .94); and total measure (α = .91).

Obsessive-Compulsive Core-Dimensions Interview (State). The Obsessive-Compulsive Core-Dimensions Interview (Summerfeldt et al., 2014) is a measure related to the OC-CDQ utilized in Study 1 that assesses the same core dimensions of INC and HA. However, instead of utilizing the 20-question self-report format, the OC-CDI is designed as an interview-based assessment. When measuring core dimensions at the state

(i.e., symptom) level, this measure was designed to accompany the Y-BOCS. For each symptom endorsed in the Y-BOCS, the participant is asked two follow-up questions: (a) "To what extent do you associate this with the fear that something harmful/bad might happen, often linked with a feeling of anxiety or apprehension?"; and (b) "To what extent do you associate this with the need to have things 'just so' or 'just right', otherwise you feel dissatisfied, tense, or very uncomfortable?" (Summerfeldt et al., 2014). The participant rates their answer on a five-point Likert scale ranging from "Not at all" (0) to "Extremely" (4), and then the scores for each dimension are separated by obsession or compulsion, leading to four raw scores: INC obsessions, INC compulsions, HA obsessions, HA compulsions. Raw scores are then divided by the number of respective endorsed symptoms in each of the four categories to generate mean total scores for each. The OC-CDI has been tested using clinical samples and was found to support the structural validity of the two-factor core dimension model. Intercorrelations between the OC-CDI and OC-CDQ ranged from $r = .69$ (INC) to $r = .79$ (HA; Summerfeldt et al., 2014). Cervin & Perrin (2019) further demonstrated the construct validity of the OC-CDI when compared to measures of disgust and the OC-CDQ.

Due to this current study not utilizing the Y-BOCS, and thus, not generating a comprehensive symptom checklist, the OC-CDI was applied instead to each of the four DOCS dimensions. After participants were prompted to think of symptoms from a dimension (and after answering the corresponding DOCS questions), participants were asked the two OC-CDI questions with reference to the symptoms they had identified. Separate total scores for HA and INC were then calculated for that time point by averaging across scores given for each of the four DOCS dimensions. The OC-CDI was

utilized in this manner twice: once as a follow-up to Study 1 for participants' current symptoms (TP2) and then again following the retrospective version of the DOCS (TP1). It should be noted that this use of the OC-CDI—as inclusive of all symptoms within each DOCS dimension and then as a retrospective questionnaire—was an adaptation for the present study and marks the first time the OC-CDI has been used in this manner.

Qualitative Symptom Change. After completing each symptom dimension of the retrospective version of the DOCS, participants were asked if "…the content/details of [their] symptoms relating to [the dimension] had changed or evolved between then (TP1) until now (TP2)". The differences between qualitative change (i.e., content-based factors), and quantitative change (i.e., various markers of severity) were thoroughly explained, and answers were recorded as yes/no. In cases where participants indicated that full DOCS categories of symptoms either no longer existed at TP2, or that new symptom dimensions appeared, change was appropriately recorded for that dimension.

Toronto Alexithymia Scale. The twenty-item Toronto Alexithymia Scale (TAS-20; Bagby, Parker & Taylor, 1994; Bagby, Taylor & Parker, 1994) is a revised version of the original 26-item Toronto Alexithymia Scale, which was designed to measure problems related to identifying, describing, and processing emotions. This self-report measure presents items as self-statements (e.g., "I am often confused about what emotion I am feeling"), uses five-point Likert scales for each item ranging from "strongly disagree" (1) to "strongly agree" (5), and is organized into three subscales: difficulty describing feelings, difficulty identifying feelings, and externally oriented thinking. Subscale scores are summed to generate a total score (20–100) with suggested cut-off scores as follows: non-alexithymia (< 51); possible alexithymia (52–60); alexithymia (>

61; Bagby, Parker, & Taylor, 1994). Assessment of the 20-item TAS reported good internal consistency (α = .81–.84), and generally good test–retest reliability (r =.39–.77, p < .01; Bagby, Taylor & Parker, 1994; Joukamaa et al., 2001; Thorberg et al., 2010) and when compared to established contemporaneous measures, was found to demonstrate convergent, concurrent, and discriminant validity (Bagby, Parker & Taylor, 1994; Thorberg et al., 2010). The TAS measures have demonstrated validity across both clinical and non-clinical cohorts and have been used in populations with varying health complications (e.g., Cleland et al., 2005; Thorberg et al., 2010). They have been translated across at least 17 languages and have become the literature benchmark for alexithymia research (e.g., Bagby et al., 2020; Leising et al., 2009). Internal consistency within this study was consistent with previous findings in the literature. Subscale alpha scores for this sample were as follows: identifying feelings (α = .88); describing feelings (α = .81); externally oriented thinking (α = .57)*;* and total measure (α = .89).

Difficulty of Emotion Regulation Scale. The 16-item Difficulty of Emotion Regulation Scale (DERS-16; Bjureberg et al., 2016;) is a shortened version of the original 36-item DERS (Gratz & Roemer, 2004) designed to assess difficulties in emotional regulation nonspecific to any individual disorder(s). The authors operationalize difficulty in emotion regulation as any deficits in awareness, understanding, or modification of emotions that are involved in accomplishing ones' subjective goals (Bjureberg et al., 2016, p.g. 284). Measure inclusion for the DERS-16 from the original DERS was empirically informed and built around five subscales: (1) lack of emotional clarity, (2) difficulty in engaging in goal-directed behaviour, (3) impulse control difficulties, (4) limited access to emotion regulation strategies, and (5)

nonacceptance of emotional responses. Bjureberg et al. (2016) concluded that the final 16 items demonstrated excellent internal consistency (α = .92) and were strongly correlated with the 36-item DERS from its inceptive sample (r = .80; p <.001). Construct validity and reliability was demonstrated across three diverse independent clinical and community samples, demonstrating test–retest validity (r = .85; p <.001) along with good convergent and discriminant validity when compared to other established measures (Bjureberg et al., 2016). Internal consistency with this study sample was good-to-excellent across all five subscales, with Cronbach's alpha scores presenting as follows: nonacceptance of emotional responses (α = .78); difficulty in engaging in goal-directed behaviour (α = .81); impulse control difficulties (α = .82); limited access to emotion regulation strategies (α = .86); lack of emotional clarity (α = .93); and total measure (α = .92).

Stressful Life Events Checklist. The Stressful Life Events Checklist (Appendix E) is a proprietary measure with items adapted from the National Longitudinal Survey of Children and Youth (Statistics Canada, 2010). The questionnaire includes eight items that broadly describe known distressing life events, including: (1) personal breakups, (2) school/work problems, (3) death of a close companion, (4) illness of close companion, (5) divorce/separation of parents, (6) financial issues, (7) personal health problems, and (8) other/misc. The measure frames the item questions to be inclusive of the time elapsed between when symptoms "…were first problematic enough to cause problems with daily functioning" (TP1) and current day (TP2). Answers were recorded using a 5-point Likert scale for each item based on how 'affective', 'stressful', or 'traumatic' the event(s) were, with possible scores ranging from "not applicable" (0) to "extremely affective" (4).

Participants in this study were asked to only include the event if it happened within the given time range (TP1–TP2), and if multiple events within a category were applicable (e.g., multiple breakups) they were asked to rate the 'worst' example. Data collected from this measure were summed to create a "stressful life events" total score—loosely reflecting breadth and severity of known stressful life events. This measure was utilized exclusively in exploratory analyses and was not required nor expected to be wholly comprehensive or discriminant as a measure of experienced life stress.

Treatment History Checklist. No known measures exist that assess the quality of participant therapies specific to OCD. The Treatment History Checklist (Appendix E) is a measure I created for this study that assesses: (1) details about treatments used—both pharmaceutical and therapeutic, and (2) overall time spent utilizing said treatment. The questions relating to treatment details are qualitative and specific to the elapsed time between age-of-onset and current day (TP1–TP2). Answers were coded based on contemporary use and empirical support of each reported therapy. For medication history, the participant was asked if they had used any pharmaceutical means to treat their OCD symptoms from age of onset to current day (TP1–TP2). If so, they were asked to identify which types. Answers were coded on a five-point scale based on the support for the identified drug class in the literature. For example: no use of exogenous compounds was coded as 0; non-psychiatric medications (e.g., recreational substances or nutritional supplements) were coded as a 1; psychiatric medications that are not considered first-line treatments (e.g., stimulants, beta blockers, benzodiazepines) were coded as 2; first-line psychiatric medications (e.g., SSRI, SNRI) were coded as 3; and multiple first-line treatments were coded as 4. Questions about psychotherapies followed

the same format, with: no support as 0; unregulated supports (e.g., talking to a friend, online social group, teacher) as 1; structured, but non-specialized supports (e.g., non-specialist therapies, life coaches, religious mentors) as 2; structured, empirically supported OCD-specific therapies (e.g., exposure and response prevention [ERP], acceptance and commitment therapy [ACT], inference-based cognitive-behavioural therapt [I-CBT]) as 3; and multiple empirically-supported OCD therapies as 4.

Each category score deferred to the participant's highest score. For example, if they reported having used THC and benzodiazepines, and had received ERP and used peer support, their scores for medication and psychotherapy would be 2 and 3 respectively. The scores for these categories were used in exploratory analyses both as separate scores and as an overall treatment quality score. To assess overall time utilizing therapy, participants were simply asked to estimate the total number of years they had spent utilizing the resources above specific to their OCD.

Analyses

Prior to running the main analyses, two-sample t-test comparisons of descriptive sample means were utilized to justify parity between this subset sample and those from Study 1. For the main analyses, measures of stability were conceptualized and analysed in four different ways. The first observed individuals' direct reports of within-dimension (i.e., qualitative) symptom change, which related to any qualitative differences or evolutions in symptom details, despite remaining present within the same overt-symptom dimension. Second, dimensional stability—or the sample's mean-level (absolute) stability over time for each dimension—was analyzed through Wilcoxon signed-rank tests, and then by observing change scores (i.e., raw and standardized mean differences

from TP1 to TP2) for all DOCS dimensions, DOCS total scores, and HA and INC dimension scores. Third, dimensional rank-order (relative) stability was assessed using Spearman test–retest correlations for all DOCS dimensions and total scores, as well as for HA and INC. Finally, ipsative profile stability was examined for the DOCS dimensions, and then separately for the core dimensions, using modified double-entry intraclass correlation analyses (ICC_{DE}; see Appendix C for a description and examples of the ICC_{DE} procedure, and for an explanation of the modifications utilized in this study).

Exploratory Analyses. As a last step, two exploratory analyses were performed. Despite the low sample *N*, multiple regression analyses were used to see if TP1 dimension scores uniquely predicted the same dimensions at TP2. This was to explore the potential of replicating findings from extant research showing the core dimensions to be stable in this regard (e.g., Mataix-Cols, 2002). Furthermore, the resulting stability coefficients and difference scores were compared against the other collected measures from Study 1 using zero-order correlations to loosely inform possible directions for future study.

Correlation, regression, and Wilcoxon signed-rank analyses were conducted using the JASP statistical software (Intel version. 0.18; JASP Team, 2023). Double-entry intraclass correlations were conducted using Microsoft Excel (Mac version 16.84).

Results

Data Preparation

In the subset of data retained for Study 2, only a single participant ($n = 1$) had missing item-level data from their Study 1 answers. Due to the limited total *N*, the participant was retained, utilizing the same case-median imputed data from Study 1.

Study 1 and 2 Sample Demographic and Descriptives Comparison

Representation from participants in this sample were equivalent to those from Study 1. With exception to a slightly higher percentage of male representation (21% vs. 16%), no significant differences were found between the total score means on the DERS-16, TAS-20, the DOCS, or other important participant demographics (see Table B3).

Intracategory Qualitative Change

Qualitative change was calculated by tallying the answers given by participants following each of the DOCS dimensional retrospective categories (Table B2). In this sample, 36 participants (95%) reported experiencing qualitative changes in symptoms within one or more of the DOCS dimensions over time.

Dimensional Mean-Level (Absolute) Stability

Mean-level scores of each dimension, for both the overt symptoms model and the core dimensions model, were compared from TP1 and TP2. Shapiro-Wilk tests, used in light of the small *N*, revealed non-normally distributed dimension scores at both TP1 and TP2 that were supported by visual appraisal of their respective histogram and QQ plots. Thus, Wilcoxon signed-rank tests were used to compare mean dimension scores (Table B4). None of the dimensions or total scores from the overt-symptoms model were found to be significantly different from TP1 to TP2. Both core dimensions were the same as well, showing no significant mean-level change.

Even with non-significant mean-level changes, mean difference scores were calculated to show trends in the dimensions over time (Table B5). For the overt symptoms, all dimensions and the DOCS total score demonstrated a trending increase in mean-level severity from TP1 to TP2, with exception of the harm/injury/bad luck

dimension, which decreased slightly over time. The core dimension of harm avoidance similarly decreased over time, whereas incompleteness remained essentially unchanged.

Rank-Order (Relative) Stability

Stability in the rank order for each of the dimensions and the DOCS total scores were assessed using test–retest correlations (Table B6). Spearman's *rho* was utilized due to the violations of normality observed in previous analyses. With exception to the overt symptom dimension relating to harm/injury/bad luck, each dimension from both the overt symptom and core dimension models, as well as overall overt-symptoms scores, demonstrated statistically significant rank-order correlations. Of all the measures, the state-level core dimension HA showed the greatest rank-order stability; however, all significant correlations were moderate-to-high, ranging from r_s = .41 to r_s = .54 (Cohen, 1988; Gignac & Szodorai, 2016). The overt symptom dimension of harm/injury/bad luck was the only non-significantly correlating measure, with a significantly lower test–retest correlation than the others (r_s = .28, p = .09).

Individual Profile (Ipsative) Stability

Based on the averaged correlation coefficients across participants, both the overt symptoms model and the core dimensions model demonstrated moderate-to-high ipsative stability from TP1 to TP2, with the core dimensions demonstrating slightly higher stability (r_{ICCDE} = .40) than overt symptoms (r_{ICCDE} = .36). The associated Pearson correlation for the overt symptoms model is higher as it only factors in the changes of profile shape from TP1 to TP2, whereas the ICC_{DE}—after controlling for elevation changes—includes the effect of changes in both profile shape and scatter (Table B7; for more on the difference between Pearson and the ICC_{DE}, see Appendix C; Furr 2010). As

the resulting model scores were averaged from the stability of each participant, the coefficients did not include an associated *p*-value. Post-hoc *p*-values based on the *N* and the averaged ICC_{DE} coefficients for the overt symptom and core dimension models suggested both to be significant at $p = .026$ and $p = .013$ respectively.

Exploratory Analyses

Regression assessment of Cohort-Level Interdimensional Change. Dimensions specific to each model at TP1 were entered as independent variables, and separate multiple regression analyses were run using each respective models' TP2 dimensions as the dependant variable. In almost all cases—both at the overt symptoms level and the core dimension level—the TP1 dimension scores uniquely predicted the same dimension at TP2 after controlling for the other dimensions within the model. Resulting *p*-values ranged from <.001 to .02. The exception to this was seen in the DOCS overt dimension of harm/injury/bad luck–related symptoms. For that regression model, none of the TP1overt-symptom dimension scores significantly predicted harm symptoms at TP2; however, this may have been a result of insufficient power.

Zero-Order Correlates of Stability. For the purpose of guiding future research, a correlation matrix was constructed to examine potential relationships between stability and other variables assessed (e.g., demographic characteristics, emotional competencies, treatment history, and life stress; see Table D8). Despite being restrictively underpowered, the most obvious patterns of significant correlations emerged between the subscales of the DERS-16 and (1) change scores from overt symptoms relating to harm/injury/bad luck ($r_s = .40$ to $r_s = .47$); and (2) change scores in state INC ($r_s = .40$ to $r_s = .47$). Changes in mean state HA and INC seemed to be associated with unique

correlates. Furthermore, at the ipsative level, only stability in the overt symptom model seemed to be associated with the other measured variables.

Discussion

The goal of the present study was to replicate findings in the current literature of longitudinal symptom stability in OCD, using a new measure for OC-symptom dimensions (DOCS) and a flexible retrospective test–retest window anchored by developmental milestones of the disorder. The study also sought to extend these findings of stability to encompass the core dimension model state measures (OC-CDI) of harm avoidance and incompleteness. Finally, the study broadened previous tests of stability for OCD by including measures of dimensional rank-order (i.e., relative) stability and a novel measure of ipsative stability (ICC_{DE}); the last of which allowed for the exploration of symptom profiles at the level of the individual (i.e., how stable each dimension was in relation to the other over time) and to explore if other personally-recorded measures were associated with that stability. The test–retest window was anchored by participants' recalled age at clinical symptom onset (TP1) and their current age (TP2). In this cohort ($N = 38$), the average span for this period was 20 years ($SD = 12.35$, *range* = 3–48), with an average TP1 age of 15 years ($SD = 6.32$, *range* = 5–32) and an average TP2 age of 35 years ($SD = 11.36$, *range* = 21–61).

Hypothesis 1

The first hypothesis was that all participants would report some level of qualitative, within-symptom changes. This position was based on the wide-spread anecdotal clinical understanding that OCD demonstrates a largely varying and evolving symptom course, and from prior literature that either empirically or anecdotally reported

within-dimension change at the item-level of overt symptoms (Besiroglu et al., 2007; de la Cruz et al., 2013; Rettew et al., 1992; Skoog & Skoog, 1999). While the hypothesis was not technically supported, it was close, with 36 of the 38 participants (95%) reporting some level of within-dimension changes from one or more of the DOCS categories. While it is certainly possible that symptoms might manifest unchanging across the course of the disorder, it is also very possible that the concept of the question was not properly or adequately communicated, or that preconceptions about symptom organization left room for differing interpretation.[6] Neither participant were obvious outliers in any meaningful demographics (e.g., notably short elapsed times between TP1 to TP2). Despite the technically unsupported hypothesis, the findings are still consistent with previous literature identifying high within-category change (e.g., Rettew et al., 1992; Skoog & Skoog, 1999).

Hypothesis 2

The second hypothesis posited that the DOCS dimensions of germs/contamination, symmetry/completeness/'just right', and the DOCS total symptom scores would significantly drop over time, whereas harm/injury/bad luck and unacceptable thoughts would stay the same. These assumptions were made based on the observed mean changes from previous studies (de la Cruz et al., 2013; Delorme et al., 2006; Mataix-Cols, 2002; Rufer et al., 2005), with best guestimates as to how those changes may have translated to the four DOCS dimensions. This hypothesis was partially

[6] One of the two who did not report any symptom changes claimed very specifically in their interview that they had "pure OCD" (i.e., one of the informal labels gaining popularity online), and despite the organization presented through the DOCS and OC-CDI, was steadfast in that specific conceptualization for their symptoms. This might suggest that pre-existing conceptualizations of symptom organization could influence participant answers.

supported, as none of the DOCS scores significantly changed from TP1 to TP2. This finding should be interpreted mindfully, as it is based on mean-level changes at the averaged sample level which still allows for changes at the level of the individual so long as it is not enough to significantly move the entire group mean or is otherwise counteracted by other participant changes within the group. Despite the lack of significant mean-level change, further examination of mean changes through the identification of difference scores revealed that the DOCS dimension of harm was trending toward a decrease over time. This stood in contrast to the other overt symptom dimensions of unacceptable thoughts, symmetry/completeness/'just right' feelings, and DOCS total scores, which all showed slight trends of increase over time. There are various possible explanations for these outcomes. It may be that there was indeed no change, and that this cohort was especially stable in their overt symptoms across time. It may also be that there were significant changes in mean scores, but the lower *N* for this study was unable to capture them due to limitations in power. The increase trend in symptoms may have been an artifact of recall biases (e.g., the assumption through retrospect that symptoms would have developed to be as bad as they currently were), or it may have been due to the TP1 anchor being placed at the start of clinically relevant symptom development, meaning that symptoms would likely legitimately get worse before they got better as they progressed. The previous longitudinal study designs would be more robust against retrospection; however, participants in those studies would all have had diagnoses prior to their study windows, meaning they may have been more likely to show symptom decreases than increases as the study progressed. This is a likely contributing factor, as when de la Cruz et al. (2013) controlled for participants whose

symptoms went into remittance due to therapy in their test–retest window, the statistically significant drops in mean scores for germ-related symptoms disappeared.

Hypothesis 3

The third hypothesis was that state harm avoidance and incompleteness mean scores would remain significantly comparable from TP1 to TP2. This was based on the findings by Besiroglu et al. (2007) that showed greater stability when a model represented similar symptom variability across fewer categories or dimensions. As the Y-BOCS category means were seen to be mostly stable in prior studies (e.g., de la Cruz, 2013; Delorme et al., 2006), it was presumed that a model averaging across the same symptoms to form fewer underlying dimensions would show greater stability. Furthermore, as the state-level core dimensions are thought to represent the same underlying constructs as the trait-level core dimensions (Summerfeldt et al., 2014), their representation of candidate endophenotypes would suggest a position of greater stability than the downstream phenotype dimensions, as they theoretically sit closer to the level of the genotype. This hypothesis was supported, as neither the means for HA nor INC changed significantly over time. The change scores for HA; however, did show a slight negative trend, whereas INC stayed essentially the same. All possibilities raised for the mean-level analyses of the overt symptoms would be relevant in these analyses as well. The limited *N* and the study design warrants mindfulness when considering the generalizability of these findings.

Hypothesis 4

The fourth hypothesis was that rank-order stability would display moderate-to-high correlations across all dimensions in both models. Large deviations in rank order

suggest that the rate, degree, or direction to which people change in their symptoms within that dimension are dissimilar compared to those of the other participants. Conversely, high positive correlation coefficients would suggest that the way in which those symptoms progress are more universal within the group. It was difficult to know what degree of effect size should be expected for the test–retest coefficients for each dimension, as no known studies had utilized this approach for these constructs. For the core dimensions, test–retest semi-partial correlations over a 15-year test window have been explored in unpublished data, and found to be significant for both HA (r = .23) and INC (r = .38; Till et al., 2024). While these finding were informative, the semi-partial correlations were part of regression analyses and controlled for the other variable, and the measures were taken at the trait level. Relatedly, Josefsson et al. (2013) found very high rank-order stability for HA (r = .70–.82, p <.01) across 10 years; however, this was also observing HA as a trait measure as opposed to the symptom/state level and had used a different measure to assess the construct. Specific to overt OCD symptoms, one study by van Grootheest et al. (2009) exploring the contribution of genetics to stability, assessed time 1 to time 2 symptom severity scores (i.e., across sample, rank-order stability) as recorded via selected items from the Young Adult Self-Report Scale for Obsessive-Compulsive Symptoms (Archenbach, 1997) and a modified version of the Padua Inventory (originally sourced from Sanavio, 1988). They found rank-order severity stability to be r = .60 across two years, r = .40 across four-to-five years, and r = .20 across 11 years. However, the study switched measures over their longest retest window. A second study looking to replicate these results similarly utilized test–retest (i.e. sample rank-order stability) scores on a differently modified version of the Padua Inventory,

which reported an overall symptom severity correlation of $r = .63$ using non-clinical participants over a six-year test window (Zilhão et al., 2014). Findings such as these were potentially informative; however, the use of different measures, and the lack of dimension-level specificity limits relevance to the current study. Studies exploring rank-order stability of personality are mixed, but typically find comparatively higher coefficients, especially as participants enter adulthood. An often-cited meta-analytic review in the personality literature by Roberts and DelVecchio (2000) assessing over 150 studies found that rank-order stability in personality traits increased from $r = .31$ in early childhood, to $r = .74$ between the ages of 50 and 70, with typical adult traits ranging from $r = .46$ for neuroticism to $r = .55$ for extraversion. Duration of test–retest window and age at the start of testing also seem to be factors in stability. Finn (1986) demonstrated an average personality rank-order stability of $r = .38$ over 30 years for people who started the study window in their teens, and a median of $r = .56$ for those who began in adulthood. Conversely, Terracciano et al. (2006) demonstrated rank order for the Big Five personality dimensions to range from $r = .76$ to $r = .84$ for those between 30 and 50 years old; while Hopwood et al. (2013) found the Big Five ranged from $r = .63$ for neuroticism and conscientiousness, to $r = .75$ for openness over ten years in a clinical sample. Such variance is a likely result of utilizing differing measures, test–retest windows, participant ages, and sample demographics. Studies looking at related psychopathology found rank-order test–retest coefficients comparatively lower (e.g., $r = .30$ for obsessive-compulsive personality disorder over 10 years; Hopwood et al., 2013).

For this current study, the assumption for parity across sample rank-order correlations was further informed by the underlying theoretical structure belief that

symptom phenotypes are at least partially a manifestation of unique genotype. It was assumed that participants ranking higher or lower within a dimension may be so due to etiological bases that are stable in their differences from participant to participant. If anything, it could have been expected that the core dimensions would demonstrate higher rank-order stability, as they are describing constructs theoretically closer to genotype than overt symptoms.

When looking at the test–retest Spearman coefficients, this hypothesis was not fully supported, as the DOCS dimension relating to harm/injury/bad luck was statistically insignificant (r_s = .28, p = .09) and was noticeably lower than the other dimensions and total scores. It was interesting to see that harm/injury/bad luck was the same dimension observed from the mean-difference analyses to deviate uniquely from the other DOCS dimensions in direction trend change. Even if it could be argued that the lower study N made it possible that significance was missed due to lower power, the correlation coefficient is low compared to the others. This lower correlation means that of all the groups, something about the overt symptoms relating to harm/injury/bad luck made them more variable at the level of the participant than was seen in other dimensions. For example, this may mean that harm/injury/bad luck–related symptoms are more amenable to treatments, sensitive to stressful personal life events, or simply may wax and wane more than the others; regardless, this dimension was less universal in the way it progressed for those who had symptoms within the group. When looking at the other groups, all resulting Spearman coefficients were significant, and all were medium-to-high, with coefficients ranging from r_s = .41 to r_s = .54. Interestingly, between just the dimension scores (i.e., excluding the DOCS total score) symptoms relating to

harm/injury/bad luck and the HA core dimension had both the lowest and highest coefficients respectively. This was not expected, but the finding demonstrates how core motives can cut across the overt symptom domains (i.e., that they can remain stable across overt symptoms, even while symptom severity in the dimension they most directly influence are less stable) and further highlights the different role the core dimension construct describes compared to overt symptoms. In this case, underlying harm-avoidant motivations across all present symptoms seem to remain comparatively stable between individuals at the level of the group, whereas severity in overt symptoms relating to harm/injury/bad luck are comparatively fragile in their relative order among participants over time. This is the first known direct examination of rank-order stability for OCD symptoms when organized by the overt symptom and state-based core dimension models.

Hypothesis 5

The fifth hypothesis was that both the overt symptoms model and the core dimension model would show comparable ipsative stability. The method for assessing ipsative stability was modified to control for overall changes in severity over time to specifically assess how the structure of each profile (i.e., the relationships between each dimension score) progressed for each person (see Appendix C). The overt symptoms (DOCS) profile (r_{ICCDE} = .36) and the core dimensions (OC-CDI) profile (r_{ICCDE} = .40) were both calculated to be significant through post-hoc testing using the coefficients and study *N* at p = .026 and p = .013 respectively. As these are both within the range for medium effect sizes using general rules of thumb and not statistically different, this hypothesis was interpreted as being tentatively supported. While these statistics can be compared within this study, there are several reasons why they are difficult to interpret

against findings from extant literature, which is why an expected effect size was not explicitly included in the hypothesis. The other somewhat comparable area of literature using this analysis is personality psychology; most specifically in examining the progression of the well-established and replicated Big Five personality traits. At first glance, the ICC_{DE} averages from this study seem lower or comparable to those of replicated personality measures. For example, Atherton et al. (2022), found an average ICC_{DE} in the Big Five to be .48 across a 12-year study window in adults, and Poier (2022) found 4-year Big Five ICC_{DE} scores ranging from .36 to .41 in adolescents, .38 to .46 in older adults, and .30 to .38 in old age. Relatedly, Terracciano et al. (2010) found the average ICC_{DE} to be .63 using the Guilford-Zimmerman temperament survey (Guilford, 1990) across their whole sample. Despite these apparent similarities, there are some notable differences that make outside comparisons difficult. First, the choice to control for elevation in this study means that the resulting correlation coefficients will almost always be higher than those that do not make this adjustment. In personality psychology, there is little need to assume large universal changes in 'personality severity' across time, so there would be a much lower potential for degradation of the final omnibus coefficient with observed changes in dimension means within the profiles. A second point is that when assessing psychopathology symptoms, it is much more likely that the measure may encounter floor effects where the participants report no symptoms within a certain dimension, but then later develop symptoms that could easily take precedence over others; or vice versa. In other words, large changes in within-person dimension rank order would be more normative in OC-symptoms than personality traits. A third related point directs attention toward what Furr (2008) refers to as the

'normativeness' problem. One premise of normativeness is the assumption that certain constraints or influences (e.g., the demands and expectations of a functional society) may shape specific norms to appear within profiles. If such shaping is occurring for constructs like personality, then it may support the maintenance or regression of dimension scores toward the mean of their 'normative' range over time, which would have the effect of bolstering measures of stability. While anecdotal, it may be that because OC-symptoms are already maladaptive, the existence of a 'normative' dimension profile for symptoms are a less likely phenomenon. Furthermore, time-related factors may impart a different outcome for traits than it does for symptoms. Whereas personality traits are known to stabilize in adulthood and stay as such into older age (e.g., Bleidorn et al., 2022), the waxing and waning of symptoms in OCD may require a different precedent for what is considered 'stable' within a given time lag or starting age. Finally, personality traits are a mix of adaptive and maladaptive traits. Literature typically demonstrates that less adaptive traits have lower longitudinal stability. As OC-symptoms and the state core dimensions are more definitionally maladaptive (especially as their levels increase), it may be expected that their normative stability should closer approximate other more maladaptive traits such as neuroticism. Neuroticism coefficients in the literature vary and are often represented by measures of rank-order stability, but they typically show comparatively lower metrics of stability than those of more adaptive trait measures (e.g., Hopwood et al., 2013). Considering the above points, it should not be expected that the ICC_{DE} coefficients from this study are directly comparable to those of the replicated Big Five personality traits. For these reasons, it is hard to know what a meaningful deviation in effect size is for these analyses, or what a typical effect size for this cohort should be

to begin with. Given these considerations, further exploration and replication will be required to better contextualize their interpretation. This was the first exploration of ipsative stability found in any measure of OCD and related symptomatology.

Hypothesis 6

The sixth hypothesis was that if, or when, there were differences in stability, even if not statistically significant, the core dimensions would be the more stable model. The prediction was included to try and capture evidence for the structural hierarchy of the models these analyses are purporting to elucidate. While certain methodological issues (e.g., cross-sectional retrospection) might confound differentiations between the models, if the core dimensions cut across and inform overt symptoms, and exist at a level closer to genotype, then there is reason to expect that their stability, relative to the overt symptoms, should be greater. When assessing stability through the lens of sample dimension means, mean-level (absolute) stabilities were comparable between models; all were similarly insignificant. When observing stability through the lens of rank-order (relative) change, the correlations were also very similar, with exception to the noted deviation in DOCS harm/injury/bad luck–related symptoms; meaning that the core dimensions model provided a more stable framework for organizing symptoms over time. When observing stability through the lens of within-person (ipsative) profile change, the core dimensions model demonstrated a higher coefficient; however, as was established, this difference lacks the context to understand how significant or meaningful the effect might be. Based on these observations, this hypothesis was considered tentatively supported, with an understanding that further replication would be needed to provide greater confidence in the core dimension model's superior stability.

Exploratory Analyses

Time 1 Dimensions as Predictors for Time 2. With the small sample in this study, it was understood a-priori that regression analyses would be restrictively underpowered; thus, it was decided not to include any regression models in the core analyses. However, for the sake of exploration, it was of interest to know whether the findings from previous literature showing TP1 scores as unique predictors of TP2 scores after controlling for the other dimensions in the model would replicate. Despite concerns of power, with exception again to the DOCS harm/injury/bad luck dimension, these findings replicated previous literature (e.g., de la Cruz et al., 2013; Delorme et al., 2006; Fullana et al., 2007; Mataix-Cols et al., 2002; Rufer et al., 2005), demonstrating a lack of cohort-level patterns of change from one dimension to the other over time. These findings were similarly observed for both core dimensions. The exception in these findings, both in relation to the other dimensions in this study and to findings from previous literature, was seen in the overt-symptom dimension of harm/injury/bad luck, which at TP2 was not significantly predicted by any of the TP1 overt symptom dimensions. This analysis again highlights a unique deviance in the overt-symptom scores relating to harm/injury/bad luck in this sample compared to the other dimensions. It would be informative to observe if these same findings replicated with more participants.

Zero-Order Correlations. Finally, a broad exploratory look at associations with stability highlighted the potential for emotion regulation difficulties to be implicated in the mean-level change scores of harm/injury/bad luck–related overt symptoms and the state core dimension of incompleteness. Given the repeated observations of comparative

instability within the dimension of overt symptoms relating to harm/injury/bad luck, this connection may be worthwhile exploring in future studies. The main issue with this analysis was that many of these variables share variance, and due to the participant number, it is very likely that either potentially significant correlations were not identified due to issues with power, and/or that the cohort itself might not validly represent the target population. While these associations are interesting as a purely explorative approach, they would need significantly more data before they could be assumed to justifiably inform directions for future research.

Limitations

Findings from this study should be interpreted in light of certain limitations. First and foremost, this study shared the same basic demographic limitations as was seen in Study 1. Despite a slightly larger representation of males in this sample, the rest of the demographic means were almost identical, and thus represented a bias toward well-educated Caucasian female participants. Relatedly, but perhaps more detrimentally, the limited participant count meant that certain analyses were either not feasible, as was the case with the exploratory regression analyses, or were otherwise bordering on low power. A-priori power analyses for the Wilcoxon signed-rank test suggested a minimum of 35 participants to ensure adequate power for a medium effect size (dz = .50, p = .05, *power* = .80); whereas analyses for the Spearman correlations suggested a minimum of 84 people for a medium effect size (r = .30, p = .05, *power* = .80) and 29 for a large effect size (r= .50, p = .05, *power* = .80). This lower power meant that many analyses might have failed in identifying significant relationships, and the lower N in general might have meant that the participants were more likely to be biased in specific areas, making the

findings from this study potentially limited in their generalizability. As mentioned in Study 1, given the limited number of participants who followed through with this project over the year-long recruitment period, it is also possible that these participants were unique in some variable that made them especially willing or driven to participate when the vast majority chose not to. While every effort was made to allay fears or discomfort during the lead up to the interviews, it was understood that the process of speaking to a stranger online about potentially embarrassing, upsetting, or otherwise uncomfortable symptoms, might be a limiting factor or deterrent. While acknowledging that a benefit of this sample was their own motivation to participate (i.e., no compensation was provided), given the low engagement, future studies would be advised to consider possible compensation or incentives for the time spent, to utilize a longer overall recruitment windows, and/or to utilize established clinical resources (e.g., hospitals or clinics) to ensure greater participant numbers, and better target population representation.

Another important limitation was that many of the measures in Study 2 were employed outside of their intended and/or validated use. The Life Events Checklist and Treatment History Checklist were adapted or designed for this study, and while their use was relegated to informing exploratory analyses and largely only required to be face-valid, their empirical validity was untested beyond the scope of the immediate study. Second, while the psychometric properties of the DOCS have been well validated, its use as a retrospective measure has not been tested in any found literature. While many of the questions on the DOCS are likely to be reasonably robust to retrospection error (i.e., most of the dimension questions are context-based, such as how hard it was to avoid compulsions in that dimension, allowing answers to be deduced as a comparison to the

other dimensions), others would be less robust (such as those asking how many hours a day were spent engaged with symptoms). Finally, the use of the OC-CDI deviated rather significantly from its intended and validated use. As designed, the OC-CDI would follow up with every endorsed symptom after the participant had completed the Y-BOCS symptom checklist, whereas in this study, the OC-CDI followed every endorsed DOCS category. This required the participant to average across the symptoms they had called to mind and then to rate their experience of each core dimension based on that averaged representation of the overt-symptom dimension. Mechanistically, the process is similar, as the resulting scores for the validated approach are averaged across the symptoms eventually. However, by putting the onus on the participant to average across symptoms, the process might be more likely to bias specific symptoms over others and may not be as accurate or articulate as when using it as designed. While admittedly not ideal to deviate in this way, the state core dimensions are thought to reflect the same underlying core dimensions as the OC-CDQ-T, meaning that even with some of the nuance lost through the process of in-vivo averaging, the underlying core dimension motives should still be proximately represented. Like with the DOCS, this was also the first time the OC-CDI was used as a retrospective measure. Future studies could attempt to replicate these findings using the Y-BOCS and the original factor-analyzed symptom organization. While this would not address the issues of retrospection, it would allow for a more accurate use of the OC-CDI and would provide the opportunity to replicate or refute these results.

The problem of retrospection highlights another limitation of the study's design. While cross-sectional retrospection beneficially allowed for the test–retest window to be

variable and predicated on important developmental anchors of symptom trajectory, it may have been susceptible to certain recall biases (e.g., Talari & Goyal, 2020). The design of the study was set up in such a way as to try and maximize separation between TP1 and TP2 answers; whereby the participants filled out their current symptoms in Study 1 and answered their retrospective symptoms at a later date in Study 2. Despite these limitations, retrospective recall has shown to be more robust when participants are reporting on instances of high frequency, which would be the case in clinical-level OC-symptoms (Hser et al., 1992). Additionally, other longitudinal biases such as response shifts can be minimized when answering cross-sectionally (see Lam & Bengo, 2003). To the defence of the current design, the findings from this study mostly replicated those of longitudinal prospective designs, even in instances where analyses had low statistical power. Furthermore, many well-validated measures in psychology regularly utilize retrospection; including the gold standard Y-BOCS symptom checklist, which includes questions about past symptom prevalence. Beyond retrospection specifically, the variable test window allowed for flexible timeframes to try and capture meaningful stages of change. However, this also meant that the tested span for stability was highly variable (3–48 years) which may limit direct comparisons to other longitudinal studies where participants all repeated the tests within similar time lags. This is especially important considering the replicated observation that other trait measures (i.e., personality) are often more or less variable across differing timeframes (e.g., van Grootheest et al., 2009). As a pilot study for this process, these findings are still important and informative so long as they are interpreted mindfully.

As a final notable limitation, many of the hypotheses lacked comparable literature from which to contextualize the results. The processes in this study for exploring stability were borrowed largely from the growing literature in personality psychology. However, despite the appropriateness and breadth of the analytic strategies for this research question, the stability expected from a profile of traits with varying levels of benefit and adaptive potential is likely to differ from a profile of symptoms representing a notably debilitating psychological disorder. Thus, understandings and norms for meaningful stability will be somewhat contextually bound, and until the results of this study are replicated, support for the hypotheses should be interpreted within the context of their novelty. Findings of stability thus far in the literature have mostly been segregated to that of mean-level (absolute) change, and time-one-to-time-two regression analyses. Future studies should continue to use a breadth of analytical tools to explore various facets of stability in OCD and other psychopathology.

Conclusion

This study explored the stability of OCD symptoms in 38 multinational participants across an average symptom progression of 20 years. Outcomes from this study generally replicated those from previous literature and extended to find similar findings from the core dimensions model. Use of additional novel measures also provided groundwork for integrating broader conceptualizations of stability. To summarize: qualitative experiences of symptoms typically changed for people over time; however, after organizing symptoms into higher-order dimensions—either by grouping them with similarly-occurring overt symptoms, or by measuring their underlying motives—those dimensions of symptoms typically remained stable over time. More

specifically, (a) mean dimension severity scores did not get meaningfully better or worse over time, (b) current symptoms within a dimension seemed to be best predicted by previous symptoms within the same dimension, and (c) people tended to maintain their rank order within those dimensions, at least as much as some well-established personality traits. Finally, (d) individual symptom profiles—or the relationship among dimensions within each of the two models for each person—tended to remain moderately stable; possibly comparable to established trait profiles like the Big Five.

The one notable exception to these findings was seen in the overt dimension of harm/injury/bad luck. While this dimension was similar to the others in that it did not show a significant change in overall severity, it demonstrated markedly lower rank-order stability than the other dimensions and was not significantly predicted by any prior symptom dimension. This means that severity of symptoms within that dimension tended to change in ways that were more individually driven, perhaps meaning they were more susceptible to personally specific environmental or situational influence. Given the findings from the exploratory analyses, those influences may have been related to difficulties with emotional regulation.

Another interesting point was that despite these comparatively lower metrics of stability in the overt symptoms of harm, the core dimension HA (i.e., harm-related motives) demonstrated the highest rank-order stability. This serves to highlight functional differences between the two models and supports the theory that the core dimensions cut across dimensions of overt symptoms. To expand on this difference, symptom-specific motives of HA, as averaged across all dimensions, displayed high rank-order stability over time—meaning changes within the sample tended to reflect more universal trends

(i.e., they moved together as a group). Conversely, the personal scores of overt-symptom severities relating to harm/injury/bad luck (i.e., the overt symptom domain most associated with HA) were following more individual and variable trajectories within the sample. Motives across symptoms stayed comparatively stable, whereas the overt symptoms in that dimension did not. These differences—the prevalence across overt symptoms and the higher relative stability in HA—together support the conceptualization of the core dimensions as endophenotypes for OCD.

While previous studies indicated decreases in various mean-level dimension severities over time, this was possibly due to prospective study windows beginning after symptoms had become well-established. The design of this study sought to track symptom progression from the initial development of clinical symptoms through to present day. Despite the instances of changes in past literature, previous authors generally found that symptoms organized at the dimension level were stable, whereas item-level symptom change (i.e., qualitative change) was normal. Taken together, I consider the findings of stability from this study to largely replicate and expand on what was understood from prior research.

Overall, it seems that symptom organization through the overt symptom and core dimension models remains relatively stable over time, with possible comparisons in some cases to that of highly established personality traits—especially when considering the ages of the participants, the test–retest window this study represents, and the maladaptive nature of OCD symptoms. Between the two models, the core dimensions model was tentatively considered the more stable; however, these models should not be considered mutually exclusive as they appear to describe different levels of a larger structural

conceptualization of obsessive-compulsive disorder. When considering the previous literature demonstrating that these symptom domains carry predictive power and share meaningful associations with topics like etiological factors, symptom trajectory, and/or clinical outcomes, such findings of stability might serve to validate their clinical use as further specifiers during diagnosis or treatment evaluation.

Chapter 4: Summary of Findings, Common Limitations, and Future Directions

The two studies from this paper explored the relationship between two leading conceptual models of symptom variance in obsessive-compulsive disorder, and then how stable dimensions from each model and the models themselves would remain over time. Findings from Study 1 demonstrated that the two trait-level core dimensions of harm avoidance (HA) and incompleteness (INC) both predicted overall symptom severity, and that after controlling for the opposite dimension, trait INC demonstrated a greater effect size of association with overall symptoms than HA. Furthermore, the two core dimensions each uniquely predicted scores in opposing and non-overlapping overt symptoms dimensions, with HA predicting scores for overt symptoms relating to harm/injury/bad luck, and those relating to unacceptable thoughts; and INC predicting symmetry/completeness/'just right' symptoms, and symptoms relating to germs/contamination.

Study 2 explored the concept of longitudinal stability in symptoms when organized by the DOCS overt symptom dimensions and the state-level core dimensions (OC-CDI). It was found that despite a high level of within-dimension change over time—where qualitative changes in symptom experience were common—symptom dimensions, when organized by both models, seemed relatively stable. The dimensions from both models showed comparable moderate-to-high dimensional mean level (absolute), rank-order (relative), and full-profile (ipsative) stability, with exception to overt symptoms relating to harm/injury/bad luck, which showed comparatively lower non-significant rank-order stability. When differences in stability were observed between the models,

such as was seen in the ipsative model stability analyses, the core dimensions model was the more stable.

Overt symptoms relating to harm/injury/bad luck were a focus at various points in these two studies. The dimension was notably low in most measures of stability; this was interesting as it stood in contrast to the relatively high stability seen in the related core dimension of HA. The regression model from Study 1 also indicated that overt symptoms relating to harm/injury/bad luck was the dimension with variance most explained by trait HA and INC. It is unclear what these findings mean, as they seem at first glance to be contradictory. However, the core dimensions from both studies were measured using different approaches (i.e., trait vs. state), and contexts (i.e., cross-sectional vs. longitudinal). Given the progressive build in stability in other traits (e.g., Roberts and DelVecchio 2000), it may be that trait core dimensions are likely to predict the eventual outcome of overt symptoms relating to harm/injury/bad luck once they have had time to stabilize, but perhaps overt harm/injury/bad luck–related symptoms are especially variable in the early stages of the disorder. Further exploration would be needed to explore these assumptions.

Some notable limitations were present across both studies. Because both studies used the same base sample, the lack of a clinical intake interview meant that participants' current clinical status was not confirmed. However, using suggested cutoffs from the symptom assessments, and comparison to other validated OCD and clinical cohorts, the sample was considered as clinically representative. Similarly, the number of participants for both experiments were borderline low for the statistical analyses used. While both studies still found results of statistical significance, it is possible that other findings were

missed due to low-powered analyses, or that the sample was unique in specific measures that would otherwise limit generalizability. Participants were also overwhelmingly Caucasian, female, well-educated English speakers. However, a notable counterpoint is that participants were truly multinational, from many differing countries and cultures, and all participants volunteered without promises of compensation or overt personal gain. This contrasts with many study designs that leverage clinical support or grade requirements to bolster participation, and that sample from single areas. Future replications of these findings will be needed with more participants from more diverse groups.

Taken together, these findings build on prior research exploring symptom dimensions and their stability while extending to include the core dimensions model. The findings of core-dimension stability further adds to the literature by supporting their conceptualization as endophenotypes for OCD, as stability is an established requisite for endophenotype candidacy (Gottesman & Gould, 2003; Taylor et al., 2012). Furthermore, considering the growing exploration of associations with meaningful risk factors, symptom trajectories, and possible clinical outcomes, the relationships between and stability within these models further supports their clinical and research potential. As a practical example: because the core dimensions are known to extend across both clinical and non-clinical populations (e.g., Summerfeldt et al., 2014), the benefit of screening those at risk of developing OCD using the OC-CDQ-T becomes increasingly apparent, as assessments like these might inform specific risk factors, disorder precipitants, or unique trends in future overt symptom development. Additionally, further diagnostic specificity using these models could eventually facilitate highly curated treatment approaches based

on patients' unique symptom profiles. Even now, if used as broader diagnostic specifiers—much like that of high or low insight—there may be various tangible and actionable clinical benefits; especially in cases where patients are specific-dimension dominant and/or for those who deviate from typical symptom presentations (e.g., those with symptoms notably dominated by motives of incompleteness; Cervin & Perrin, 2021; Mathes et al., 2019; McKay, 2020; Schwartz, 2018; Summerfeldt, 2004; Summerfeldt, 2007). Continued exploration will be required to examine whether similar results would replicate in prospective longitudinal designs and with more representative samples. However, these findings of model associations and their stability address current gaps in the literature, offer compelling support for the models' hypothesized relative co-relationships, mechanisms, and meta-structures, and provide further support for their ongoing consideration for clinical use and research.

References

Abramowitz, J. S., Deacon, B. J., Olatunji, B. O., Wheaton, M. G., Berman, N. C., Losardo, D., Timpano, K. R., McGrath, P. B., Riemann, B. C., Adams, T., Björgvinsson, T., Storch, E. A., & Hale, L. R. (2010). Assessment of obsessive-compulsive symptom dimensions: Development and evaluation of the Dimensional Obsessive-Compulsive Scale. *Psychological Assessment*, *22*(1), 180–198. https://doi.org/10.1037/a0018260

Abramowitz, J. S., McKay, D., & Taylor, S. (Eds.). (2008). *Clinical handbook of obsessive-compulsive disorder and related problems*. John Hopkins University Press.

Adams, T. G., Kelmendi, B., Brake, C. A., Gruner, P., Badour, C. L., & Pittenger, C. (2018). The role of stress in the pathogenesis and maintenance of obsessive-compulsive disorder. *Chronic Stress*, *2*, 2470547018758043–2470547018758054. https://doi.org/10.1177/2470547018758043

Alemany-Navarro, M., Cruz, R., Real, E., Segalàs, C., Bertolín, S., Rabionet, R., Carracedo, Á., Menchón, J. M., & Alonso, P. (2020). Looking into the genetic bases of OCD dimensions: A pilot genome-wide association study. *Translational Psychiatry*, *10*(1), 151–167. https://doi.org/10.1038/s41398-020-0804-z

Alonso, P., Menchon, J. M., Pifarre, J., Mataix-Cols, D., Torres, L., Salgado, P., & Vallejo, J. (2001). Long-term follow-up and predictors of clinical outcome in obsessive-compulsive patients treated with serotonin reuptake inhibitors and behavioral therapy. *Journal of Clinical Psychiatry, 62*(7), 535–544. http://doi.org/10.4088/jcp.v62n07a06

Alsobrook II, J. P., Leckman, J. F., Goodman, W. K., Rasmussen, S. A., & Pauls, D. L. (1999). Segregation analysis of obsessive-compulsive disorder using symptom-based factor scores. *American Journal of Medical Genetics*, *88*(6), 669–675. https://doi.org/10.1002/(SICI)1096-8628(19991215)88:6<669::AID-AJMG17>3.0.CO;2-N

Aluja, A., Malas, O., Urieta, P., Worner, F., & Balada, F. (2020). Biological correlates of the Toronto Alexithymia Scale (TAS-20) in cardiovascular disease and healthy community subjects. *Physiology & Behavior*, *227*, 113151–113159. https://doi.org/10.1016/j.physbeh.2020.113151

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Achenbach, T. M. (1997). *Manual for the young adult self-report and young adult behavior checklist*. University of Vermont, Department of Psychiatry.

Atherton, O. E., Sutin, A. R., Terracciano, A., & Robins, R. W. (2022). Stability and change in the Big Five personality traits: Findings from a longitudinal study of Mexican-origin adults. *Journal of Personality and Social Psychology*, *122*(2), 337–350. https://doi.org/10.1037/pspp0000385

Baer, L. (1994). Factor analysis of symptom subtypes of obsessive-compulsive disorder and their relation to personality and tic disorders. *The Journal of clinical psychiatry*, *55*, 18–23. PMID: 8077163

Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale—I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research, 38*(1), 23–32. http://doi.org/10.1016/0022-3999(94)90005-1

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (2020). Twenty-five years with the 20-item Toronto Alexithymia Scale. *Journal of Psychosomatic Research*, *131*, 109940–109940. https://doi.org/10.1016/j.jpsychores.2020.109940

Bagby, R. M., Taylor, G. J., & Parker, J. D. (1994). The twenty-item Toronto Alexithymia Scale—II: Convergent, discriminant, and concurrent validity. *Journal of Psychosomatic Research, 38*(1), 33–40. https://doi.org/10.1016/0022-3999(94)90005-1

Bagheri, M., Nematollah Zadeh Mahani, K., & Pour Amrollahi, M. (2020). The mediating role of alexithymia in the relationship of coping strategies and personality traits with obsessive-compulsive disorder. *Internal Medicine Today*, *27*(1), 62–81. http://doi.org/10.32598/hms.27.1.3355.1

Belloch, A., Fornés, G., Carrasco, A., López-Solá, C., Alonso, P., & Menchón, J. M. (2016). Incompleteness and not just right experiences in the explanation of obsessive–compulsive disorder. *Psychiatry Research*, *236*, 1–8. https://doi.org/10.1016/j.psychres.2016.01.012

Besiroglu, L., Uguz, F., Ozbebit, O., Guler, O., Cilli, A. S., & Askin, R. (2007). Longitudinal assessment of symptom and subtype categories in obsessive-compulsive disorder. *Depression and Anxiety*, *24*(7), 461–466. https://doi.org/10.1002/da.20240

Bey, K., Weinhold, L., Grützmann, R., Heinzel, S., Kaufmann, C., Klawohn, J., Riesel, A., Lennertz, L., Schmid, M., Ramirez, A., Kathmann, N., & Wagner, M. (2020). The polygenic risk for obsessive-compulsive disorder is associated with the personality trait harm avoidance. *Acta Psychiatrica Scandinavica*, *142*(4), 326–336. https://doi.org/10.1111/acps.13226

Bishara, A. J., & Hittner, J. B. (2015). Reducing bias and error in the correlation coefficient due to nonnormality. *Educational and Psychological Measurement*, *75*(5), 785–804. https://doi.org/10.1177/0013164414557639

Bjureberg, J., Ljótsson, B., Tull, M. T., Hedman, E., Sahlin, H., Lundh, L.-G., Bjärehed, J., DiLillo, D., Messman-Moore, T., Gumpert, C. H., & Gratz, K. L. (2016). Development and validation of a brief version of the Difficulties in Emotion Regulation Scale: The DERS-16. *Journal of Psychopathology and Behavioral Assessment*, *38*(2), 284–296. https://doi.org/10.1007/s10862-015-9514-x

Bleidorn, W., Schwaba, T., Zheng, A., Hopwood, C. J., Sosa, S. S., Roberts, B. W., & Briley, D. A. (2022). Personality stability and change: A meta-analysis of longitudinal studies. *Psychological Bulletin*, *148*(7–8), 588–619. https://doi.org/10.1037/bul0000365

Bloch, M. H., Green, C., Kichuk, S. A., Dombrowski, P. A., Wasylink, S., Billingslea, E., Landeros-Weisenberger, A., Kelmendi, B., Goodman, W. K., Leckman, J. F., Coric, V., & Pittenger, C. (2013). Long-term outcome in adults with obsessive-compulsive disorder: Research article: Long-term outcome in adult OCD. *Depression and Anxiety*, *30*(8), 716–722. https://doi.org/10.1002/da.22103

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. *The American Journal of Psychiatry*, *165*(12), 1532–1542. https://doi.org/10.1176/appi.ajp.2008.08020320

Borghuis, J., Denissen, J. J. A., Oberski, D., Sijtsma, K., Meeus, W. H. J., Branje, S., Koot, H. M., & Bleidorn, W. (2017). Big Five personality stability, change, and codevelopment across adolescence and early adulthood. *Journal of Personality and Social Psychology*, *113*(4), 641–657. https://doi.org/10.1037/pspp0000138

Bragdon, L. B., & Coles, M. E. (2017). Examining heterogeneity of obsessive-compulsive disorder: Evidence for subgroups based on motivations. *Journal of Anxiety Disorders,* 45, 64–71. https://doi.org/10.1016/j.janxdis.2016.12.002

Burton, A. L., Brown, R., & Abbott, M. J. (2022). Overcoming difficulties in measuring emotional regulation: Assessing and comparing the psychometric properties of the DERS long and short forms. *Cogent Psychology*, *9*(1), 2060629–2060646. https://doi.org/10.1080/23311908.2022.2060629

Calkins, D. S. (1974). Some effects of non-normal distribution shape on the magnitude of the Pearson Moment Correlation Coefficient. *Revista Interamericana de Psicología/Interamerican Journal of Psychology*, *8*(3 & 4). 261–288. https://doi.org/10.30849/rip/ijp.v8i3%20&%204.708

Cameron, D. H., Rowa, K., McKinnon, M. C., Rector, N. A., & McCabe, R. E. (2020). Neuropsychological performance across symptom dimensions of obsessive-compulsive disorder: A comment on the state and critical review of the literature. *Expert Review of Neurotherapeutics*, *20*(5), 425–438. https://doi.org/10.1080/14737175.2020.1746644

Cameron, D. H., Streiner, D. L., Summerfeldt, L. J., Rowa, K., McKinnon, M. C., & McCabe, R. E. (2019). A comparison of cluster and factor analytic techniques for identifying symptom-based dimensions of obsessive-compulsive disorder. *Psychiatry Research*, *278*, 86–96. https://doi.org/10.1016/j.psychres.2019.05.040

Cameron, D. H., Summerfeldt, L. J., Rowa, K., McKinnon, M. C., Rector, N. A., Richter, M. A., Ornstein, T. J., & McCabe, R. E. (2019). Differences in neuropsychological performance between incompleteness- and harm avoidance-related core dimensions in obsessive-compulsive disorder. *Journal of Obsessive-Compulsive and Related Disorders*, *22*, 100448–100455. https://doi.org/10.1016/j.jocrd.2019.100448

Cervin, M., & Perrin, S. (2019). Measuring harm avoidance, incompleteness, and disgust in youth with obsessive-compulsive disorder and anxiety disorders. *Journal of Obsessive-Compulsive and Related Disorders*, *22*, 100442–100449. https://doi.org/10.1016/j.jocrd.2019.100442

Cervin, M., & Perrin, S. (2021). Incompleteness and disgust predict treatment outcome in pediatric obsessive-compulsive disorder. *Behavior Therapy*, *52*(1), 53–63. https://doi.org/10.1016/j.beth.2020.01.007

Chase, T., Wetterneck, C. T., Bartsch, R. A., Leonard, R. C., & Riemann, B. C. (2015). Investigating treatment outcomes across OCD symptom dimensions in a clinical sample of OCD patients. *Cognitive Behaviour Therapy*, *44*(5), 365–376. https://doi.org/10.1080/16506073.2015.1015162

Clark, D. A. (2005). Lumping versus splitting: A commentary on subtyping in OCD. *Behavior Therapy*, *36*(4), 401–404. https://doi.org/10.1016/S0005-7894(05)80122-4

Clark, D. A. (2006). *Cognitive-behavioral therapy for OCD.* Guilford Press.

Cleland, C., Magura, S., Foote, J., Rosenblum, A., & Kosanke, N. (2005). Psychometric properties of the Toronto Alexithymia Scale (TAS-20) for substance users. *Journal of Psychosomatic Research*, *58*(3), 299–306. https://doi.org/10.1016/j.jpsychores.2004.11.002

Cloninger, C. R. (1986). A unified biosocial theory of personality and its role in the development of anxiety states. *Psychiatric developments, 3*(2), 167–226.

Cloninger, C. R., & Svrakic, D. M. (1997). Integrative psychobiological approach to psychiatric assessment and treatment. *Psychiatry*, *60*(2), 120–141. https://doi.org/10.1080/00332747.1997.11024793

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*. Lawrence Erlbaum Associates

Coles, M. E., Heimberg, R. G., Frost, R. O., & Steketee, G. (2005). Not just right experiences and obsessive–compulsive features: Experimental and self-monitoring perspectives. *Behaviour Research and Therapy*, *43*(2), 153–167. https://doi.org/10.1016/j.brat.2004.01.002

Coles, M. E., Johnson, E. M., & Schubert, J. R. (2011). Retrospective reports of the development of obsessive-compulsive disorder: Extending knowledge of the protracted symptom phase. *Behavioural and Cognitive Psychotherapy*, *39*(5), 579–589. https://doi.org/10.1017/S135246581100004X

Compas, B. E., Jaser, S. S., Bettis, A. H., Watson, K. H., Gruhn, M. A., Dunbar, J. P., Williams, E., & Thigpen, J. C. (2017). Coping, emotion regulation, and psychopathology in childhood and adolescence: A meta-analysis and narrative review. *Psychological Bulletin*, *143*(9), 939–991. https://doi.org/10.1037/bul0000110

De Berardis, D., Campanella, D., Gambi, F., Sepede, G., Salini, G., Carano, A., La Rovere, R., Pelusi, L., Penna, L., Cicconetti, A., Cotellessa, C., Salerno, R. M., & Ferro, F. M. (2005). Insight and alexithymia in adult outpatients with obsessive-compulsive disorder. *European Archives of Psychiatry and Clinical Neuroscience*, *255*(5), 350–358. https://doi.org/10.1007/s00406-005-0573-y

de Fruyt, F., Bartels, M., Van Leeuwen, K. G., De Clercq, B., Decuyper, M., & Mervielde, I. (2006). Five types of personality continuity in childhood and adolescence. *Journal of personality and social psychology*, *91*(3), 538–552. https://doi.org/10.1037/0022-3514.91.3.538

de la Cruz, F. L., Micali, N., Roberts, S., Turner, C., Nakatani, E., Heyman, I., & Mataix-Cols, D. (2013). Are the symptoms of obsessive-compulsive disorder temporally stable in children/adolescents? A prospective naturalistic study. *Psychiatry Research*, *209*(2), 196–201. https://doi.org/10.1016/j.psychres.2012.11.033

Delorme, R., Bille, A., Betancur, C., Mathieu, F., Chabane, N., Mouren-Simeoni, M. C., & Leboyer, M. (2006). Exploratory analysis of obsessive-compulsive symptom dimensions in children and adolescents: A prospective follow-up study. *BMC Psychiatry*, *6*(1), 1–10. https://doi.org/10.1186/1471-244X-6-1

Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, *56*(293), 52–64. https://doi.org/10.1080/01621459.1961.10482090

Ecker, W., & Gönner, S. (2008). Incompleteness and harm avoidance in OCD symptom dimensions. *Behaviour Research and Therapy*, *46*(8), 895–904. https://doi.org/10.1016/j.brat.2008.04.002

Eisen, J. L., Goodman, W. K., Keller, M. B., Warshaw, M. G., DeMarco, L. M., Luce, D. D., & Rasmussen, S. A. (1999). Patterns of remission and relapse in obsessive-compulsive disorder: A 2-year prospective study. *The Journal of Clinical Psychiatry, 60*(5), 346–351. http://doi.org/10.4088/jcp.v60n0514

Eisen, J. L., Sibrava, N. J., Boisseau, C. L., Mancebo, M. C., Stout, R. L., Pinto, A., & Rasmussen, S. A. (2013). Five-year course of obsessive-compulsive disorder: Predictors of remission and relapse. *The Journal of Clinical Psychiatry, 74*(3), 233–239. http://doi.orh/10.4088/JCP.12m07657

Eilertsen, T., Hansen, B., Kvale, G., Abramowitz, J. S., Holm, S. E. H., & Solem, S. (2017). The Dimensional Obsessive-Compulsive Scale: Development and validation of a short form (DOCS-SF). *Frontiers in Psychology*, *8*, 1503–1510. https://doi.org/10.3389/fpsyg.2017.01503

Enander, J., Andersson, E., Kaldo, V., Lindefors, N., Andersson, G., & Rück, C. (2012). Internet administration of the Dimensional Obsessive-Compulsive Scale: A psychometric evaluation. *Journal of Obsessive-Compulsive and Related Disorders*, *1*(4), 325–330. https://doi.org/10.1016/j.jocrd.2012.07.008

Ferrão, Y. A., Shavitt, R. G., Bedin, N. R., De Mathis, M. E., Lopes, A. C., Fontenelle, L. F., Torres, A. R., & Miguel, E. C. (2006). Clinical features associated to refractory obsessive–compulsive disorder. *Journal of affective disorders*, *94*(1-3), 199–209. https://doi.org/10.1016/j.jad.2006.04.019

Fineberg, N. A., Dell'Osso, B., Albert, U., Maina, G., Geller, D., Carmi, L., Sireau, N., Walitza, S., Grassi, G., Pallanti, S., Hollander, E., Brakoulias, V., Menchon, J. M., Marazziti, D., Ioannidis, K., Apergis-Schoute, A., Stein, D. J., Cath, D. C., Veltman, D. J., … Zohar, J. (2019). Early intervention for obsessive compulsive disorder: An expert consensus statement. *European Neuropsychopharmacology*, *29*(4), 549–565. https://doi.org/10.1016/j.euroneuro.2019.02.002

Finn, S. E. (1986). Stability of personality self-ratings over 30 years: Evidence for an age/cohort interaction. *Journal of Personality and Social Psychology*, *50*(4), 813–818. https://doi.org/10.1037/0022-3514.50.4.813

Foa, E. B., Abramowitz, J. S., Franklin, M. E., & Kozak, M. J. (1999). Feared consequences, fixity of belief, and treatment outcome in patients with obsessive-compulsive disorder. *Behavior Therapy*, *30*(4), 717–724. https://doi.org/10.1016/S0005-7894(99)80035-5

Foa, E. B., Huppert, J. D., Leiberg, S., Langner, R., Kichic, R., Hajcak, G., & Salkovskis, P. M. (2002). The Obsessive-Compulsive Inventory: Development and validation of a short version. *Psychological Assessment*, *14*(4), 485–496. https://doi.org/10.1037/1040-3590.14.4.485

Foa, E. B., & Kozak, M. J. (1986). Emotional processing of fear: Exposure to corrective information. *Psychological Bulletin*, *99*(1), 20–35. https://doi.org/10.1037/0033-2909.99.1.20

Foa, E. B., Kozak, M. J., Goodman, W. K., Hollander, E., Jenike, M. A., & Rasmussen, S. A. (1995). DSM-IV field trial: Obsessive-compulsive disorder. *The American Journal of Psychiatry*, *152*(1), 90–96. https://doi.org/10.1176/ajp.152.1.90

Fullana, M. A., Tortella-Feliu, M., Caseras, X., Taberner, J., Torrubia, R., & Mataix-Cols, D. (2007). Temporal stability of obsessive-compulsive symptom dimensions in an undergraduate sample: A prospective 2-year follow-up study. *Behavior Modification, 31*(6), 815–824. https://doi.org/10.1177/0145445507301649

Furr, R. M. (2008). A framework for profile similarity: Integrating similarity, normativeness, and distinctiveness. *Journal of Personality*, *76*(5), 1267–1316. https://doi.org/10.1111/j.1467-6494.2008.00521.x

Furr, R. M. (2010). The double-entry intraclass correlation as an index of profile similarity: Meaning, limitations, and alternatives. *Journal of Personality Assessment*, *92*(1), 1–15. https://doi.org/10.1080/00223890903379134

Garyfallos, G., Katsigiannopoulos, K., Adamopoulou, A., Papazisis, G., Karastergiou, A., & Bozikas, V. P. (2010). Comorbidity of obsessive–compulsive disorder with obsessive–compulsive personality disorder: Does it imply a specific subtype of obsessive–compulsive disorder? *Psychiatry Research*, *177*(1), 156–160. https://doi.org/10.1016/j.psychres.2010.01.006

Ghisi, M., Chiri, L. R., Marchetti, I., Sanavio, E., & Sica, C. (2010). In search of specificity: "Not just right experiences" and obsessive–compulsive symptoms in non-clinical and clinical Italian individuals. *Journal of Anxiety Disorders*, *24*(8), 879–886. https://doi.org/10.1016/j.janxdis.2010.06.011

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences*, *102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., Heninger, G. R., & Charney, D. S. (1989). The Yale-Brown Obsessive-Compulsive Scale: I. Development, use, and reliability. *Archives of General Psychiatry, 46*(11), 1006–1011. http://doi.org/10.1001/archpsyc.1989.01810110048007

Gottesman, I. I., & Gould, T. D. (2003). The endophenotype concept in psychiatry: Etymology and strategic intentions. *American Journal of Psychiatry, 160*(4), 636–645. https://doi.org/10.1176/appi.ajp.160.4.636

Grados, M. A., Riddle, M. A., Samuels, J. F., Liang, K. Y., Hoehn-Saric, R., Bienvenu, O. J., Walkup, J. T., Song, D. H., & Nestadt, G. (2001). The familial phenotype of obsessive-compulsive disorder in relation to tic disorders: The Hopkins OCD Family Study. *Biological Psychiatry*, *50*(8), 559–565. https://doi.org/10.1016/S0006-3223(01)01074-5

Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of Psychopathology and Behavioral Assessment*, *26*(1), 41–54. https://doi.org/10.1023/B:JOBA.0000007455.08539.94

Grisham, J. R., Fullana, M. A., Mataix-Cols, D., Moffitt, T. E., Caspi, A., & Poulton, R. (2011). Risk factors prospectively associated with adult obsessive–compulsive symptom dimensions and obsessive–compulsive disorder. *Psychological Medicine*, *41*(12), 2495–2506. https://doi.org/10.1017/S0033291711000894

Guilford, J. P. (1990). Guilford-Zimmerman Temperament Survey. In C. A. Beere (Ed.), *Gender roles: A handbook of tests and measures,* (131). Greenwood Publishing.

Hansenne, M., & Ansseau, M. (1999). Harm avoidance and serotonin. *Biological psychology*, *51*(1), 77–81. https://doi.org/10.1016/S0301-0511(99)00018-6

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, J. O., Shugart, Y., Liang, K., Hoehn-Saric, R., … Murphy, D. L. (2007). Familiarity of factor analysis-derived YBOCS dimensions in OCD-affected sibling pairs from the OCD Collaborative Genetics Study. *Biological Psychiatry*, *61*(5), 617–625. https://doi.org/10.1016/j.biopsych.2006.05.040

Hezel, D. M., Rose, S. V., & Simpson, H. B. (2022). Delay to diagnosis in OCD. *Journal of Obsessive-Compulsive and Related Disorders*, *32*, 100709–100713. https://doi.org/10.1016/j.jocrd.2022.100709

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70.

Hopwood, C. J., Morey, L. C., Donnellan, M. B., Samuel, D. B., Grilo, C. M., McGlashan, T. H., Shea, M. T., Zanarini, M. C., Gunderson, J. G., & Skodol, A. E. (2013). Ten-year rank-order stability of personality traits and disorders in a clinical sample. *Journal of Personality*, *81*(3), 335–344. https://doi.org/10.1111/j.1467-6494.2012.00801.x

Horncastle, T., Ludlow, A. K., & Gutierrez, R. (2022). Not just right experiences and incompleteness as a predictor of obsessive compulsive symptoms in clinical and community samples: A meta-analysis. *Journal of Obsessive-Compulsive and Related Disorders*, *35*, 100762–100768. https://doi.org/10.1016/j.jocrd.2022.100762

Hser, Y., Anglin, M. D., & Chou, C. (1992). Reliability of retrospective self-report by narcotics addicts. *Psychological Assessment, 4*(2), 207-213. https://doi.org/10.1037/1040-3590.4.2.207

Hunt, C. (2020). Differences in OCD symptom presentations across age, culture, and gender: A quantitative review of studies using the Y-BOCS symptom checklist. *Journal of Obsessive-Compulsive and Related Disorders*, *26*, 100533–100544. https://doi.org/10.1016/j.jocrd.2020.100533

International OCD Foundation. (2024). *Organization home directory page.* International OCD Foundation. http://iocdf.org

Irwin, L. D., & Jones, M. K. (2017). The relationship between obsessive-compulsive symptoms, perfectionism, and anxiety sensitivity for not just right experiences. *Behaviour Change, 34*(3), 134–155. https://doi.org/10.1017/bec.2017.10

Janowitz, D., Grabe, H. J., Ruhrmann, S., Ettelt, S., Buhtz, F., Hochrein, A., Schulze-Rauschenback, S., Meyer, K., Kraft, S., Ferber, S., Pukrop, R., Freyberger, H., Klosterkötter, J., Falkai, P., John, U., Maier, W., & Wagner, M. (2009). Early onset of obsessive–compulsive disorder and associated comorbidity. *Depression and Anxiety*, *26*(11), 1012–1017. https://doi.org/10.1002/da.20597

JASP Team. (2023). *JASP* (Version 0.18). University of Amsterdam. https://jasp-stats.org/

Jaspers-Fayer, F., Han, S. H. J., Chan, E., McKenney, K., Simpson, A., Boyle, A., Ellwyn, R., & Stewart, S. E. (2017). Prevalence of acute-onset subtypes in pediatric obsessive-compulsive disorder. *Journal of Child and Adolescent Psychopharmacology*, *27*(4), 332–341. https://doi.org/10.1089/cap.2016.0031

Jenike, M. A., Baer, L., Minichiello, W. E., Rauch, S. L., & Buttolph, M. L. (1997). Placebo-controlled trial of fluoxetine and phenelzine for obsessive-compulsive disorder. *The American Journal of Psychiatry*, *154*(9), 1261–1264. https://doi.org/10.1176/ajp.154.9.1261

Josefsson, K., Jokela, M., Cloninger, C. R., Hintsanen, M., Salo, J., Hintsa, T., Pulkki-Råback, L., & Keltikangas-Järvinen, L. (2013). Maturity and change in personality: Developmental trends of temperament and character in adulthood. *Development and Psychopathology*, *25*(3), 713–727. https://doi.org/10.1017/S0954579413000126

Joukamaa, M., Miettunen, J., Kokkonen, P., Koskinen, M., Julkunen, J., Kauhanen, J., Jokelainen, J., Veijola, J., Läksy, K., & Järvelin, M. (2001). Psychometric properties of the Finnish 20-item Toronto Alexithymia Scale. *Nordic Journal of Psychiatry*, *55*(2), 123–127. https://doi.org/10.1080/08039480116694

Kichuk, S. A., Torres, A. R., Fontenelle, L. F., Rosário, M. C., Shavitt, R. G., Miguel, E. C., Pittenger, C., & Bloch, M. H. (2013). Symptom dimensions are associated with age of onset and clinical course of obsessive–compulsive disorder. *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, *44*, 233–239. https://doi.org/10.1016/j.pnpbp.2013.02.003

Kloosterman, P. H., Summerfeldt, L. J., Parker, J. D., & Holden, J. J. (2013). The obsessive-compulsive trait of incompleteness in parents of children with autism spectrum disorders. *Journal of Obsessive-Compulsive and Related Disorders*, *2*(2), 176–182. https://doi.org/10.1016/j.jocrd.2012.11.004

Kuckertz, J. M., Rhee, D. W., Schreck, M., Van Kirk, N., Dattolico, D., & Falkenstein, M. J. (2021). Reevaluating the psychometric properties of symptom measures within longitudinal contexts: The Yale–Brown Obsessive Compulsive Scale and Dimensional Obsessive–Compulsive Scale. *Psychological Assessment, 33*(8), 756–765. https://doi.org/10.1037/pas0001017

Lam, T. C. M., & Bengo, P. (2003). A comparison of three retrospective self-reporting methods of measuring change in instructional practice. *The American Journal of Evaluation*, *24*(1), 65–80. https://doi.org/10.1177/109821400302400106

Landeros-Weisenberger, A., Bloch, M. H., Kelmendi, B., Wegner, R., Nudel, J., Dombrowski, P., Pittenger, C., Krystal, J. H., Goodman, W. K., Leckman, J. F., & Coric, V. (2010). Dimensional predictors of response to SRI pharmacotherapy in obsessive–compulsive disorder. *Journal of Affective Disorders*, *121*(1), 175–179. https://doi.org/10.1016/j.jad.2009.06.010

Leckman, J. F., Bloch, M. H., & King, R. A. (2009). Symptom dimensions and subtypes of obsessive-compulsive disorder: A developmental perspective. *Dialogues in Clinical Neuroscience, 11*(1), 21–33. https://doi.org/10.31887/DCNS.2009.11.1/jfleckman

Leckman, J. F., Denys, D., Simpson, H. B., Mataix-Cols, D., Hollander, E., Saxena, S., Miguel, E. C., Rauch, S. L., Goodman, W. K., Phillips, K. A., & Stein, D. J. (2010). Obsessive-compulsive disorder: A review of the diagnostic criteria and possible subtypes and dimensional specifiers for DSM-V. *Depression and Anxiety*, *27*(6), 507–527. https://doi.org/10.1002/da.20669

Leckman, J. F., Pauls, D. L., Zhang, H., Rosario-Campos, M. C., Katsovich, L., Kidd, K. K., Pakstis, A. J., Alsobrook, J. P., Robertson, M. M., McMahon, W. M., Walkup, J. T., van de Wetering, B. J. M., King, R. A., & Cohen, D. J. (2003). Obsessive-compulsive symptom dimensions in affected sibling pairs diagnosed with Gilles de la Tourette syndrome. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*, *116B*(1), 60–68. https://doi.org/10.1002/ajmg.b.10001

Lee, H.-J., & Kwon, S.-M. (2003). Two different types of obsession: Autogenous obsessions and reactive obsessions. *Behaviour Research and Therapy*, *41*(1), 11–29. https://doi.org/10.1016/S0005-7967(01)00101-2

Lee, S. R., & Wu, K. D. (2019). Feelings of incompleteness explain symptoms of OCD and OCPD beyond harm avoidance. *Journal of Obsessive-Compulsive and Related Disorders, 21*, 151–157. https://doi.org/10.1016/j.jocrd.2019.04.002

Leising, D., Grande, T., & Faber, R. (2009). The Toronto Alexithymia Scale (TAS-20): A measure of general psychological distress. *Journal of Research in Personality*, *43*(4), 707–710. https://doi.org/10.1016/j.jrp.2009.03.009

Lincoln, T. M., Schulze, L., & Renneberg, B. (2022). The role of emotion regulation in the characterization, development and treatment of psychopathology. *Nature Reviews Psychology, 1*(5), 272–286. https://doi.org/10.1038/s44159-022-00040-4

López-Solà, C., Gutiérrez, F., Alonso, P., Rosado, S., Taberner, J., Segalàs, C., Real, E., Menchón, J. M., & Fullana, M. A. (2014). Spanish version of the Dimensional Obsessive–Compulsive Scale (DOCS): Psychometric properties and relation to obsessive beliefs. *Comprehensive Psychiatry*, *55*(1), 206–214. https://doi.org/10.1016/j.comppsych.2013.08.015

Mahjani, B., Bey, K., Boberg, J., & Burton, C. (2021). Genetics of obsessive-compulsive disorder. *Psychological Medicine*, *51*(13), 2247–2259. https://doi.org/10.1017/S0033291721001744

Mantovani, A., Shubeck, J., Gowatsky, J., Simpson, H. B., & Greenberg, B. J. C. N. (2013). Incompleteness and harm avoidance in obsessive compulsive disorder: Different response to transcranial magnetic stimulation. *Clinical Neurophysiology, 124*(10), e182–e183. https://doi.org/10.1016/j.clinph.2013.04.323

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. *American Journal of Psychiatry, 159*(2), 263–268. https://doi.org/10.1176/appi.ajp.159.2.263

Mataix-Cols, D., Rosario-Campos, M. C., & Leckman, J. F. (2005). A multidimensional model of obsessive-compulsive disorder. *American Journal of Psychiatry, 162*(2), 228–238. https://doi.org/10.1176/appi.ajp.162.2.228

Mataix-Cols, D., Nakatani, E., Micali, N., & Heyman, I. (2008). Structure of obsessive-compulsive symptoms in pediatric OCD. *Journal of the American Academy of Child & Adolescent Psychiatry, 47*(7), 773–778. https://doi.org/10.1097/CHI.0b013e31816b73c0

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. *Archives of general psychiatry*, *61*(6), 564–576. http://doi.org/10.1001/archpsyc.61.6.564

Mathes, B. M., Kennedy, G. A., Wilver, N. L., Carlton, C. N., & Cougle, J. R. (2019). A multi-method analysis of incompleteness in behavioral treatment of contamination-based OCD. *Behaviour Research and Therapy*, *114*, 1–6. https://doi.org/10.1016/j.brat.2018.12.008

Matsunaga, H., Hayashida, K., Kiriike, N., Maebayashi, K., & Stein, D. J. (2010). The clinical utility of symptom dimensions in obsessive–compulsive disorder. *Psychiatry Research*, *180*(1), 25–29. https://doi.org/10.1016/j.psychres.2009.09.005

Matsunaga, H., Maebayashi, K., Hayashida, K., Okino, K., Matsui, T., Iketani, T., Kiriike, N., & Stein, D. J. (2008). Symptom structure in Japanese patients with obsessive-compulsive disorder. *The American Journal of Psychiatry*, *165*(2), 251–253. https://doi.org/10.1176/appi.ajp.2007.07020340

McCrae, R. R. (2008). A note on some measures of profile agreement. *Journal of Personality Assessment*, *90*(2), 105–109. https://doi.org/10.1080/00223890701845104

McGrath, P. (2024). *What is OCD?* NOCD. https://www.treatmyocd.com/what-is-ocd

McGuire, J. F., Crawford, E. A., Park, J. M., Storch, E. A., Murphy, T. K., Larson, M. J., & Lewin, A. B. (2014). Neuropsychological performance across symptom dimensions in pediatric obsessive-compulsive disorder: Research article: Neuropsychological functioning and OCD symptoms. *Depression and Anxiety*, *31*(12), 988–996. https://doi.org/10.1002/da.22241

McKay, D. (2020). Treatment of incompleteness in obsessive-compulsive disorder. In E. A. Storch, D. McKay, & J. S. Abramowitz (Eds.), *Advanced casebook of obsessive-compulsive and related disorders* (pp. 57–69). Academic Press. https://doi.org/10.1016/B978-0-12-816563-8.00004-8

McKay, D., Abramowitz, J. S., Calamari, J. E., Kyrios, M., Radomsky, A., Sookman, D., Taylor, S., & Wilhelm, S. (2004). A critical evaluation of obsessive–compulsive disorder subtypes: Symptoms versus mechanisms. *Clinical Psychology Review*, *24*(3), 283–313. https://doi.org/10.1016/j.cpr.2004.04.003

McKay, D., & Neziroglu, F. (2009). Methodological issues in the obsessive–compulsive spectrum. *Psychiatry Research*, *170*(1), 61–65. https://doi.org/10.1016/j.psychres.2009.01.004

Melke, J., Westberg, L., Nilsson, S., Landén, M., Soderstrom, H., Baghaei, F., Rosmond, R., Holm, G., Björntorp, P., Nilsson, L. G., Adolfsson, R., & Eriksson, E. (2003). A polymorphism in the serotonin receptor 3A (HTR3A) gene and its association with harm avoidance in women. *Archives of General Psychiatry*, *60*(10), 1017–1023. http://doi.org/10.1001/archpsyc.60.10.1017

Melli, G., Chiorri, C., Bulli, F., Carraresi, C., Stopani, E., & Abramowitz, J. (2015). Factor congruence and psychometric properties of the Italian version of the Dimensional Obsessive-Compulsive Scale (DOCS) across non-clinical and clinical samples. *Journal of Psychopathology and Behavioral Assessment, 37*(2), 329–339. https://doi.org/10.1007/s10862-014-9450-1

Melli, G., Moulding, R., Puccetti, C., Pinto, A., Caccico, L., Drabik, M. J., & Sica, C. (2020). Assessing beliefs about the consequences of not just right experiences: Psychometric properties of the Not Just Right Experience-Sensitivity Scale (NJRE-SS). *Clinical Psychology & Psychotherapy*, *27*(6), 847–857. https://doi.org/10.1002/cpp.2468

Miguel, E. C., Leckman, J. F., Rauch, S., do Rosario-Campos, M. C., Hounie, A. G., Mercadante, M. T., Chacon, P., & Pauls, D. L. (2005). Obsessive-compulsive disorder phenotypes: Implications for genetic studies. *Molecular Psychiatry, 10*(3), 258–275. https://doi.org/10.1038/sj.mp.4001617

Naftalovich, H., Anholt, G. E., Keren, R., Ben Arush, O., & Kalanthroff, E. (2021). Waxing and waning: The roles of chronotype and time of day in predicting symptom fluctuations in obsessive-compulsive disorder using a daily-monitoring design. *Journal of Psychiatric Research*, *143*, 91–97. https://doi.org/10.1016/j.jpsychires.2021.08.032

Ojalehto, H. J., Abramowitz, J. S., Hellberg, S. N., Buchholz, J. L., & Twohig, M. P. (2020). Adherence to exposure and response prevention as a predictor of improvement in obsessive-compulsive symptom dimensions. *Journal of Anxiety Disorders*, *72*, 102210–102218. https://doi.org/10.1016/j.janxdis.2020.102210

Ojalehto, H. J., Pinciotti, C. M., Hellberg, S. N., Myers, N. S., DuBois, C. M., Rodriguez, C. S., & Abramowitz, J. S. (2023). Obsessive-compulsive symptom profiles vary by index trauma type in a trauma-exposed community sample. *Journal of Obsessive-Compulsive and Related Disorders*, *38*, 100827–100835. https://doi.org/10.1016/j.jocrd.2023.100827

Ólafsson, R. P., Emmelkamp, P. M., Ólason, D. Þ., & Kristjánsson, Á. (2020). Disgust and contamination concerns: The mediating role of harm avoidance and incompleteness. *International Journal of Cognitive Therapy, 13*(3), 251–270. https://doi.org/10.1007/s41811-020-00076-5

Olatunji, B. O., Davis, M. L., Powers, M. B., & Smits, J. A. J. (2013). Cognitive-behavioral therapy for obsessive-compulsive disorder: A meta-analysis of treatment outcome and moderators. *Journal of Psychiatric Research*, *47*(1), 33–41. https://doi.org/10.1016/j.jpsychires.2012.08.020

Orchard. (2024). *Welcome to the Orchard OCD registry*. OCD Registry. https://orchardocdregistry.org

Pallanti, S., Hollander, E., Bienstock, C., Koran, L., Leckman, J., Marazziti, D., Pato, M., Stein, D., & Zohar, J. (2002). Treatment non-response in OCD: Methodological issues and operational definitions. *The International Journal of Neuropsychopharmacology*, *5*(2), 181–191. https://doi.org/10.1017/S1461145702002900

Pelissolo, A., & Corruble, E. (2002). Personality factors in depressive disorders: Contribution of the psychobiologic model developed by Cloninger. *L'encephale, 28*(4), 363–373. PMID: 12232546

Pietrefesa, A. S., & Coles, M. E. (2008). Moving beyond an exclusive focus on harm avoidance in obsessive compulsive disorder: Considering the role of incompleteness. *Behavior Therapy, 39*(3), 224–231. https://doi.org/10.1016/j.beth.2007.08.004

Pietrefesa, A. S., & Coles, M. E. (2009). Moving beyond an exclusive focus on harm avoidance in obsessive-compulsive disorder: Behavioral validation for the separability of harm avoidance and incompleteness. *Behavior Therapy, 40*(3), 251–259. https://doi.org/10.1016/j.beth.2008.06.003

Pinto, A., Greenberg, B. D., Grados, M. A., Bienvenu, O. J., Samuels, J. F., Murphy, D. L., Hasler, G., Stout, R. L., Rauch, S. L., Shugart, Y. Y., Pauls, D. L., Knowles, J. A., Fyer, A. J., McCracken, J. T., Piacentini, J., Wang, Y., Willour, V. L., Cullen, B., Liang, K.-Y., … Nestadt, G. (2008). Further development of YBOCS dimensions in the OCD Collaborative Genetics Study: Symptoms vs. categories. *Psychiatry Research*, *160*(1), 83–93. https://doi.org/10.1016/j.psychres.2007.07.010

Pinto, A., Mancebo, M. C., Eisen, J. L., Pagano, M. E., & Rasmussen, S. A. (2006). The Brown Longitudinal Obsessive Compulsive Study: Clinical features and symptoms of the sample at intake. *Journal of Clinical Psychiatry, 67*(5), 703–711. http://doi.org/10.4088/jcp.v67n0503

Poier, S. (2022). How stable is your customer? Individual and ipsative consistency of consumers' Big Five personality traits. *Contemporary Economics*, *16*(3), 297–316. https://doi.org/10.5709/ce.1897-9254.483

Radomsky, A. S., & Taylor, S. (2005). Subtyping OCD: Prospects and problems. *Behavior Therapy, 36*(4), 371–379. https://doi.org/10.1016/S0005-7894(05)80119-4

Rasmussen, S. A., & Eisen, J. L. (1992). The epidemiology and clinical features of obsessive-compulsive disorder. *Psychiatric Clinics*, *15*(4), 743–758. https://doi.org/10.1016/S0193-953X(18)30205-3

Raymond, M. R. (1986). Missing data in evaluation research. *Evaluation & the Health Professions*, *9*(4), 395–420. https://doi.org/10.1177/016327878600900401

Rettew, D. C., Swedo, S. E., Leonard, H. L., Lenane, M. C., & Rapoport, J. L. (1992). Obsessions and compulsions across time in 79 children and adolescents with obsessive-compulsive disorder. *Journal of the American Academy of Child & Adolescent Psychiatry, 31*(6), 1050–1056. https://doi.org/10.1097/00004583-199211000-00009

Roberts, B. W., & DelVecchio, W. F. (2000). The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychological bulletin*, *126*(1), 3–25. https://doi.org/10.1037/0033-2909.126.1.3

Roh, D., Kim, W. J., & Kim, C. H. (2011). Alexithymia in obsessive-compulsive disorder: Clinical correlates and symptom dimensions. *The Journal of Nervous and Mental Disease*, *199*(9), 690–695. https://doi.org/10.1097/NMD.0b013e318229d209

Rowsell, M., & Francis, S. E. (2015). OCD subtypes: Which, if any, are valid? *Clinical Psychology*, *22*(4), 414–435. https://doi.org/10.1111/cpsp.12130

Rufer, M., Grothusen, A., Maß, R., Peter, H., & Hand, I. (2005). Temporal stability of symptom dimensions in adult patients with obsessive–compulsive disorder. *Journal of Affective Disorders*, *88*(1), 99–102. https://doi.org/10.1016/j.jad.2005.06.003

Ruscio, A. M., Stein, D. J., Chiu, W. T., & Kessler, R. C. (2010). The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. *Molecular Psychiatry*, *15*(1), 53–63. https://doi.org/10.1038/mp.2008.94

Salkovskis, P. M. (1985). Obsessional-compulsive problems: A cognitive-behavioural analysis. *Behaviour Research and Therapy*, *23*(5), 571–583. https://doi.org/10.1016/0005-7967(85)90105-6

Sanavio, E. (1988). Obsessions and compulsions: The Padua Inventory. *Behaviour Research and Therapy, 26*(2), 169–177. https://doi.org/10.1016/0005-7967(88)90116-7

Schwartz, R. A. (2018). Treating incompleteness in obsessive-compulsive disorder: A meta-analytic review. *Journal of Obsessive-Compulsive and Related Disorders*, *19*, 50–60. https://doi.org/10.1016/j.jocrd.2018.08.001

See, C. C. H., Tan, J. M., Tan, V. S. Y., & Sündermann, O. (2022). A systematic review on the links between emotion regulation difficulties and obsessive-compulsive disorder. *Journal of Psychiatric Research*, *154*, 341–353. https://doi.org/10.1016/j.jpsychires.2022.07.023

Shafran, R., Thordarson, D. S., & Rachman, S. (1996). Thought-action fusion in obsessive compulsive disorder. *Journal of Anxiety Disorders*, *10*(5), 379–391. https://doi.org/10.1016/0887-6185(96)00018-7

Shahabi, M., Hasani, J., & Bjureberg, J. (2020). Psychometric properties of the brief Persian version of the Difficulties in Emotion Regulation Scale (The DERS-16). *Assessment for Effective Intervention*, *45*(2), 135–143. https://doi.org/10.1177/1534508418800210

Sharma, E., Sharma, L. P., Balachander, S., Lin, B., Manohar, H., Khanna, P., Lu, C., Garg, K., Thomas, T. L., Au, A. C. L., Selles, R. R., Højgaard, D. R. M. A., Skarphedinsson, G., & Stewart, S. E. (2021). Comorbidities in obsessive-compulsive disorder across the lifespan: A systematic review and meta-analysis. *Frontiers in Psychiatry, 12*, 703701–703720. https://doi.org/10.3389/fpsyt.2021.703701

Sheppes, G., Suri, G., & Gross, J. J. (2015). Emotion regulation and psychopathology. *Annual Review of Clinical Psychology, 11*(1), 379–405. https://doi.org/10.1146/annurev-clinpsy-032814-112739

Sica, C., Bottesi, G., Caudek, C., Orsucci, A., & Ghisi, M. (2016). "Not Just Right Experiences" as a psychological endophenotype for obsessive-compulsive disorder: Evidence from an Italian family study. *Psychiatry Research*, *245*, 27–35. https://doi.org/10.1016/j.psychres.2016.08.005

Sica, C., Caudek, C., Chiri, L. R., Ghisi, M., & Marchetti, I. (2012). "Not just right experiences" predict obsessive–compulsive symptoms in non-clinical Italian individuals: A one-year longitudinal study. *Journal of Obsessive-Compulsive and Related Disorders, 1*(3), 159–167. https://doi.org/10.1016/j.jocrd.2012.03.006

Sibrava, N. J., Boisseau, C. L., Eisen, J. L., Mancebo, M. C., & Rasmussen, S. A. (2016). An empirical investigation of incompleteness in a large clinical sample of obsessive compulsive disorder. *Journal of Anxiety Disorders, 42*, 45–51. https://doi.org/10.1016/j.janxdis.2016.05.005

Skoog, G., & Skoog, I. (1999). A 40-year follow-up of patients with obsessive-compulsive disorder. *Archives of General Psychiatry*, *56*(2), 121–127. https://doi.org/doi:10.1001/archpsyc.56.2.121

Smit, D. J. A., Cath, D., Zilhão, N. R., Ip, H. F., Denys, D., Braber, A., Geus, E. J. C., Verweij, K. J. H., Hottenga, J., & Boomsma, D. I. (2020). Genetic meta-analysis of obsessive–compulsive disorder and self-report compulsive symptoms. *American Journal of Medical Genetics. Part B, Neuropsychiatric Genetics*, *183*(4), 208–216. https://doi.org/10.1002/ajmg.b.32777

Sookman, D., Abramowitz, J. S., Calamari, J. E., Wilhelm, S., & McKay, D. (2005). Subtypes of obsessive-compulsive disorder: Implications for specialized cognitive behavior therapy. *Behavior Therapy*, *36*(4), 393–400. https://doi.org/10.1016/S0005-7894(05)80121-2

Starcevic, V., Berle, D., Brakoulias, V., Sammut, P., Moses, K., Milicevic, D., & Hannan, A. (2011). Functions of compulsions in obsessive–compulsive disorder. *Australian & New Zealand Journal of Psychiatry, 45*(6), 449–457. https://doi.org/10.3109/00048674.2011.567243

Statistics Canada. (2010). *National Longitudinal Survey of Children and Youth, Cycle 8: Survey instruments*. Ottawa: Special Surveys Division.

Stein, D. J., Andersen, E. W., & Overo, K. F. (2007). Response of symptom dimensions in obsessive-compulsive disorder to treatment with Citalopram or placebo. *Revista Brasileira de Psiquiatria*, *29*(4), 303–307. https://doi.org/10.1590/s1516-44462007000400003

Stein, D. J., Carey, P. D., Lochner, C., Seedat, S., Fineberg, N., & Andersen, E. W. (2008). Escitalopram in obsessive-compulsive disorder: Response of symptom dimensions to pharmacotherapy. *CNS Spectrums, 13*(6), 492–498. https://doi.org/10.1017/S1092852900016722

Stewart, S. E., Rosario, M. C., Baer, L., Carter, A. S., Brown, T. A., Scharf, J. M., Illmann, C., Leckman, J. F., Sukhodolsky, D., Katsovich, L., Rasmussen, S., Goodman, W., Delorme, R., Leboyer, M., Chabane, N., Jenike, M. A., Geller, D. A., & Pauls, D. L. (2008). Four-factor structure of obsessive-compulsive disorder symptoms in children, adolescents, and adults. *Journal of the American Academy of Child and Adolescent Psychiatry*, *47*(7), 763–772. https://doi.org/10.1097/CHI.0b013e318172ef1e

Stivaleti Colombarolli, M., Carolina Zuanazzi, A., Koich Miguel, F., & Giromini, L. (2019). Psychometric properties of the Toronto Alexithymia Scale (TAS-20) in Brazil. *Transcultural Psychiatry*, *56*(5), 992–1010. https://doi.org/10.1177/1363461519847312

Storch, E. A., McNamara, J., Jordan, C., Marien, W. E., Jacob, M. L., Murphy, T. K., Goodman, W. K., & Geffken, G. R. (2008). Associations between miscellaneous symptoms and symptom dimensions in adults with obsessive-compulsive disorder. *Anxiety, Stress, and Coping, 21*(2), 199–212. https://doi.org/10.1080/10615800701885369

Strom, N. I., Soda, T., Mathews, C. A., & Davis, L. K. (2021). A dimensional perspective on the genetics of obsessive-compulsive disorder. *Translational Psychiatry*, *11*(1), 401–401. https://doi.org/10.1038/s41398-021-01519-z

Summerfeldt, L. J. (2004). Understanding and treating incompleteness in obsessive-compulsive disorder. *Journal of Clinical Psychology*, *60*(11), 1155–1168. https://doi.org/10.1002/jclp.20080

Summerfeldt, L. J. (2007). Treating Incompleteness, Ordering, and Arranging Concerns. In M. M. Antony, C. Purdon, & L. J. Summerfeldt (Eds.), *Psychological treatment of obsessive-compulsive disorder: Fundamentals and beyond* (pp. 187–207). American Psychological Association. https://doi.org/10.1037/11543-008

Summerfeldt, L. J. (2014). *Establishing norms for presentation of the core dimensions in clinical and non-clinical samples* [Unpublished manuscript]. Department of Psychology, Trent University.

Summerfeldt, L.J., Antony, M. M., & Swinson, R. P. (2002, March). Age of onset and symptom profile in obsessive-compulsive disorder. In J. Abramowitz & C. Carmin (Chairs). *Treatment of psychopathology of obsessive-compulsive disorder across the lifespan* [Symposium]. National Conference of the Anxiety Disorders Association of America, Austin, TX, USA.

Summerfeldt, L. J., Gilbert, S. J., & Reynolds, M. (2015). Incompleteness, aesthetic sensitivity, and the obsessive-compulsive need for symmetry. *Journal of Behavior Therapy and Experimental Psychiatry*, *49*(Part B), 141–149. https://doi.org/10.1016/j.jbtep.2015.03.006

Summerfeldt, L. J., Richter, M. A., Antony, M. M., & Swinson, R. P. (1999). Symptom structure in obsessive-compulsive disorder: A confirmatory factor-analytic study. *Behaviour Research and Therapy, 37*(4), 297–311. https://doi.org/10.1016/S0005-7967(98)00134-X

Summerfeldt, L. J., Kloosterman, P. H., Antony, M. M., Richter, M. A., & Swinson, R. P. (2004). The relationship between miscellaneous symptoms and major symptom factors in obsessive-compulsive disorder. *Behaviour Research and Therapy, 42*(12), 1453–1467. https://doi.org/10.1016/j.brat.2003.09.006

Summerfeldt, L. J., Kloosterman, P. H., Antony, M. M., & Swinson, R. P. (2014). Examining an obsessive-compulsive core dimensions model: Structural validity of harm avoidance and incompleteness. *Journal of Obsessive-Compulsive and Related Disorders, 3*(2), 83-94. https://doi.org/10.1016/j.jocrd.2014.01.003

Summers, B. J., Wilver, N. L., Garratt, G. H., & Cougle, J. R. (2020). A multimethod analysis of incompleteness and visual "not just right" experiences in body dysmorphic disorder. *Behavior Therapy, 51*(5), 764–773. https://doi.org/10.1016/j.beth.2019.11.001

Talari, K., & Goyal, M. (2020). Retrospective studies–utility and caveats. *Journal of the Royal College of Physicians of Edinburgh*, *50*(4), 398–402. https://doi.org/10.4997/JRCPE.2020.409

Taylor, P. J. (2010). *An introduction to intraclass correlation that resolves some common confusions.* [Unpublished manuscript]. University of Massachusetts. https://faculty.umb.edu/peter_taylor/09b.pdf

Taylor, S. (2011). Early versus late onset obsessive–compulsive disorder: Evidence for distinct subtypes. *Clinical Psychology Review*, *31*(7), 1083–1100. https://doi.org/10.1016/j.cpr.2011.06.007

Taylor, S. (2012). Endophenotypes of obsessive–compulsive disorder: Current status and future directions. *Journal of Obsessive-Compulsive and Related Disorders, 1*(4), 258–262. https://doi.org/10.1016/j.jocrd.2012.06.004

Taylor, S., Abramowitz, J. S., & McKay, D. (2007). Cognitive-behavioral models of obsessive-compulsive disorder. In M. M. Antony, C. Purdon, & L. J. Summerfeldt (Eds.), *Psychological treatment of obsessive-compulsive disorder: Fundamentals and beyond* (pp. 9–29). American Psychological Association. https://doi.org/10.1037/11543-001

Taylor, G. J., Bagby, R. M., & Parker, J. D. A. (2003). The 20-Item Toronto Alexithymia Scale: IV. Reliability and factorial validity in different languages and cultures. *Journal of Psychosomatic Research*, *55*(3), 277–283. https://doi.org/10.1016/S0022-3999(02)00601-3

Taylor, S., McKay, D., Crowe, K. B., Abramowitz, J. S., Conelea, C. A., Calamari, J. E., & Sica, C. (2014). The sense of incompleteness as a motivator of obsessive-compulsive symptoms: An empirical analysis of concepts and correlates. *Behavior Therapy, 45*(2), 254–262. https://doi.org/10.1016/j.beth.2013.11.004

Terracciano, A., Costa Jr, P. T., & McCrae, R. R. (2006). Personality plasticity after age 30. *Personality and Social Psychology Bulletin, 32*(8), 999–1009. https://doi.org/10.1177/0146167206288599

Terracciano, A., McCrae, R. R., & Costa, P. T. (2010). Intra-individual change in personality stability and age. *Journal of Research in Personality*, *44*(1), 31–37. https://doi.org/10.1016/j.jrp.2009.09.006

Thorberg, F. A., Young, R. M., Sullivan, K. A., Lyvers, M., Connor, J. P., & Feeney, G. F. (2010). A psychometric comparison of the Toronto Alexithymia Scale (TAS-20) and the Observer Alexithymia Scale (OAS) in an alcohol-dependent sample. *Personality and Individual Differences*, *49*(2), 119–123. https://doi.org/10.1016/j.paid.2010.03.018

Till, J. (2023). *Trent obsessive-compulsive disorder (OCD) research study* [Video]. YouTube. https://www.youtube.com/watch?v=4GO2jpBlxXA

Till. J., Parker, J. D. A. & Summerfeldt, L. J. (2024, June 22). *Stability of obsessive-compulsive-related core dimensions: Results from a 15-year longitudinal study* [Poster presentation]. Canadian Psychological Association National Convention, Ottawa, On, Canada.

Tolin, D. F. (2009). Alphabet Soup: ERP, CT, and ACT for OCD. *Cognitive and Behavioral Practice*, *16*(1), 40–48. https://doi.org/10.1016/j.cbpra.2008.07.001

Torresan, R. C., Ramos-Cerqueira, A. T. A., Shavitt, R. G., do Rosário, M. C., de Mathis, M. A., Miguel, E. C., & Torres, A. R. (2013). Symptom dimensions, clinical course and comorbidity in men and women with obsessive-compulsive disorder. *Psychiatry Research*, *209*(2), 186–195. https://doi.org/10.1016/j.psychres.2012.12.006

Tsikriktsis, N. (2005). A review of techniques for treating missing data in OM survey research. *Journal of Operations Management, 24*(1), 53–62. https://doi.org/10.1016/j.jom.2005.03.001

van den Heuvel, O. A., Remijnse, P. L., Mataix-Cols, D., Vrenken, H., Groenewegen, H. J., Uylings, H. B. M., van Balkom, A. J. L. M., & Veltman, D. J. (2009). The major symptom dimensions of obsessive-compulsive disorder are mediated by partially distinct neural systems. *Brain*, *132*(4), 853–868. https://doi.org/10.1093/brain/awn267

van Grootheest, D. S., Cath, D., Hottenga, J. J., Beekman, A. T., & Boomsma, D. I. (2009). Genetic factors underlie stability of obsessive–compulsive symptoms. *Twin Research and Human Genetics*, *12*(5), 411–419. https://doi.org/10.1375/twin.12.5.411

Walker, R., Moraine, A. A., Black, J. K., (2021). *Running and interpreting multiple linear regressions in JASP*. Exploring diversity with statistics. https://utc.pressbooks.pub/step-by-step-JASP-guides/chapter/running-and-interpreting-multiple-regression-in-jasp/

Weyers, P., Krebs, H., & Janke, W. (1995). Reliability and construct validity of the German version of Cloninger's Tridimensional Personality Questionnaire. *Personality and Individual Differences*, *19*(6), 853–861. https://doi.org/10.1016/S0191-8869(95)00128-X

Wiborg, I. M., Falkum, E., Dahl, A. A., & Gullberg, C. (2005). Is harm avoidance an essential feature of patients with panic disorder? *Comprehensive Psychiatry, 46*(4), 311–314. https://doi.org/10.1016/j.comppsych.2004.08.009

Williams, M. T., Farris, S. G., Turkheimer, E., Pinto, A., Ozanick, K., Franklin, M. E., Liebowitz, M., Simpson, H. B., & Foa, E. B. (2011). Myth of the pure obsessional type in obsessive-compulsive disorder. *Depression and Anxiety*, *28*(6), 495–500. https://doi.org/10.1002/da.20820

Williams, M. T., Mugno, B., Franklin, M., & Faber, S. (2013). Symptom dimensions in obsessive-compulsive disorder: Phenomenology and treatment outcomes with exposure and ritual prevention. *Psychopathology*, *46*(6), 365–376. https://doi.org/10.1159/000348582

World Health Organization. (2008). *The global burden of disease: 2004 update.* World Health Organization. https://iris.who.int/bitstream/handle/10665/43942/9789241563710_eng.pdf?sequence=1

Yap, K., Mogan, C., Moriarty, A., Dowling, N., Blair-West, S., Gelgec, C., & Moulding, R. (2018). Emotion regulation difficulties in obsessive-compulsive disorder. *Journal of Clinical Psychology*, *74*(4), 695–709. https://doi.org/10.1002/jclp.22553

Yiğit, İ., & Guzey Yiğit, M. (2019). Psychometric properties of Turkish version of Difficulties in Emotion Regulation Scale-Brief Form (DERS-16). *Current Psychology, 38*, 1503–1511. https://doi.org/10.1007/s12144-017-9712-7

Ziegler, S., Bednasch, K., Baldofski, S., & Rummel-Kluge, C. (2021). Long durations from symptom onset to diagnosis and from diagnosis to treatment in obsessive-compulsive disorder: A retrospective self-report study. *PloS One*, *16*(12), e0261169–e0261184. https://doi.org/10.1371/journal.pone.0261169

Zilhão, N. R., Smit, D. J., den Braber, A., Dolan, C. V., Willemsen, G., Boomsma, D. I., & Cath, D. C. (2015). Genetic and environmental contributions to stability in adult obsessive compulsive behavior. *Twin Research and Human Genetics*, *18*(1), 52–60. https://doi.org/10.1017/thg.2014.77

Appendix A
Study 1 Tables and Figures

Table A1

Study 1 Sample Demographics

Demographic Measures	Study 1 (*N=90*)		
	Mean	***SD***	**Min – Max**
Age	–	–	–
Current Age (TP2)	33.32	11.47	18 – 67
Age First Symptoms	9.99	5.69	1 – 33
Age of Onset (TP1)	16.1	6.48	5 – 35
Age First Helped	21.16	7.544	3 – 40
Age First Diagnosis	24.38	8.84	3 – 50
Elapsed Times	–	–	–
First Symptoms – Current Age (TP2)	23.33	10.87	3 – 53
Age of Onset (TP1) – Current age (TP2)	17.22	11.63	0 – 47
First Symptoms – First Diagnosis	14.39	8.54	0 – 39
Age of Onset (TP1) – First Diagnosis	8.27	8.39	-11 – 33
First Diagnosis – Current Age (TP2)	9.94	9.816	0 – 36
	n (%)		
Biological Sex	–		
Male	14 (16)		
Female	74 (82)		
Other/Prefer Not to Say	2 (2)		
Gender	–		
Man	15 (17)		
Woman	67 (74)		
Other/Prefer Not to Say	8 (9)		
Ethnicity	–		
African/Black/Caribbean	4 (4)		
Asian (South, Central, East, West)	4 (4)		
Caucasian/White	79 (88)		
Hispanic/Latino	3 (3)		
Other	2 (2)		
Prefer Not to say	2 (2)		
Education	–		
Elementary School	1 (1)		
High School	15 (17)		
Post Secondary (University/College)	48 (53)		
Graduate Studies (MA, PhD)	24 (27)		
Other	1 (1)		
Prefer Not to Say	1 (1)		

Note. TP1 = time-point 1; TP2 = time-point 2.

Table A2

Study 1 Measure Descriptives

Measures	Study 1 (*N* = *90*)					
	Mean	***SD***	**Min – Max**	***α***	**Possible Range**	***n (%)***
Trait Core Dimensions (OC-CDQ-T)						
Trait Harm Avoidance	28.3	8.5	4 – 40	.914	0 – 40	–
Trait Incompleteness	24.97	8.25	0 – 40	.894	0 – 40	–
Overt Symptom Dimensions (DOCS)						
Germs/Contamination	6.77	6.37	0 – 20	.963	0 – 20	–
Harm/Injury/Bad Luck	9.41	5.94	0 – 20	.949	0 – 20	–
Unacceptable Thoughts	9.21	6.53	0 – 20	.957	0 – 20	–
Symmetry/Completeness/'Just Right'	7.31	5.91	0 – 20	.950	0 – 20	–
DOCS Total Score	32.7	14.02	6 – 72	.879	0 – 80	–
Cutoff Scores:	–	–	–	–	–	–
Clinical Cutoff (18+; Suggested Diagnostic Threshold)	–	–	–	–	–	78 (87)
Difficulty in Emotion Regulation Scale (DERS - 16)						
Subscale 1 (Clarity)	5.93	2.62	2 – 10	.927	2 – 10	–
Subscale 2 (Goals)	12.37	2.71	5 – 15	.812	3 – 15	–
Subscale 3 (Impulse)	8.56	3.54	3 – 15	.822	3 – 15	–
Subscale 4 (Strategies)	18.88	4.89	5 – 25	.862	5 – 25	–
Subscale 5 (Nonacceptance)	10.81	3.17	4 – 15	.777	3 – 15	–
DERS-16 Total Score	56.55	13.52	22 – 80	.921	16 – 80	–
Toronto Alexithymia Scale (TAS-20)						
Subscale 1 (Identifying Feelings)	21.59	7.47	7 – 35	.883	7 – 35	–
Subscale 2 (Describing Feelings)	14.33	5.4	5 – 25	.811	5 – 25	–
Subscale 3 (Externally Oriented Thinking)	17.69	4.83	8 – 28	.572	8 – 40	–
TAS-20 Total Score	53.61	14.97	20 – 86	.886	20 – 100	–
Cutoff Scores:	–	–	–	–	–	–
Non-Alexithymia (< 51)	–	–	–	–	–	36 (40)
Possible Alexithymia (52-60)	–	–	–	–	–	22 (24)
Alexithymia (61+)	–	–	–	–	–	32 (36)

Note. OC-CDQ-T = Obsessive-Compulsive Core-Dimension Questionnaire (trait version); DOCS = Dimensional Obsessive-Compulsive Scale; DERS = Difficulty with Emotion Regulation Scale; TAS = Toronto Alexithymia Scale.

Table A3

Study 1 Measure Zero-Order Correlations

		Study 1 (*N* = 90)						
	Variable	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	Trait Harm Avoidance	—						
2	Trait Incompleteness	.42***	—					
3	DOCS Germs/Contamination	.18	.41***	—				
4	DOCS Harm Injury/Bad Luck	.55***	.38***	.16	—			
5	DOCS Unacceptable Thoughts	.20	.01	-.27**	.25*	—		
6	DOCS Symmetry/Balance/'Just Right'	.12	.49***	.18	.39***	-.08	—	
7	DOCS Total Score	.47***	.57***	.45***	.80***	.38***	.62***	—
8	DERS 1 (Clarity)	.25*	.32**	-.02	.14	.32**	.16	.26*
9	DERS 2 (Goals)	.40***	.19	.12	.18	.22*	.00	.25*
10	DERS 3 (Impulse)	.17	.23*	.01	-.04	.17	.21*	.16
11	DERS (Strategies)	.42***	.23*	-.03	.19	.33**	.06	.27*
12	DERS (Nonacceptance)	.49***	.31**	.12	.26*	.28**	.21*	.35***
13	DERS Total Score	.42***	.30**	.04	.15	.35***	.16	.31**
14	TAS 1 (Identifying)	.25*	.26*	-.07	.19	.28**	.16	.28**
15	TAS 2 (Describing)	.13	.19	-.01	.15	.15	.19	.26*
16	TAS 3 (External)	-.071	-.05	-.20	-.06	.01	-.05	-.10
17	TAS Total Score	.17	.21*	-.08	.13	.21	.13	.21*
18	Current Age	-.29**	-.06	-.20	-.19	-.04	-.02	-.19
19	First Symptoms	-.01	-.21*	-.16	-.01	.10	-.21*	-.10
20	Age of Onset	-.08	-.15	.01	-.14	-.12	-.14	-.18
21	Age First Helped	-.10	-.11	.00	-.19	-.14	-.10	-.21*
22	Age First Diagnosis	-.14	-.11	-.16	-.03	-.05	.11	-.05
23	First Symptoms – Current Age	-.25*	-.01	-.10	-.22*	-.07	.08	-.15
24	Age of Onset – Current Age	-.18	.02	-.16	-.08	.01	.09	-.05
25	First Symptoms – First Diagnosis	-.16	-.07	-.06	-.10	-.12	.19	-.04
26	Age of Onset – First Diagnosis	-.07	-.04	-.19	.05	-.02	.21	.04
27	First Diagnosis – Current Age	-.12	.02	.04	-.12	-.01	-.11	-.10

Note. DOCS = Dimensional Obsessive-Compulsive Scale; DERS = Difficulty with Emotion Regulation Scale; TAS = Toronto Alexithymia Scale; TP1 = time point 1; TP2 = time point 2. *p <.05; **p <.01; ***p <.001.

Table A4

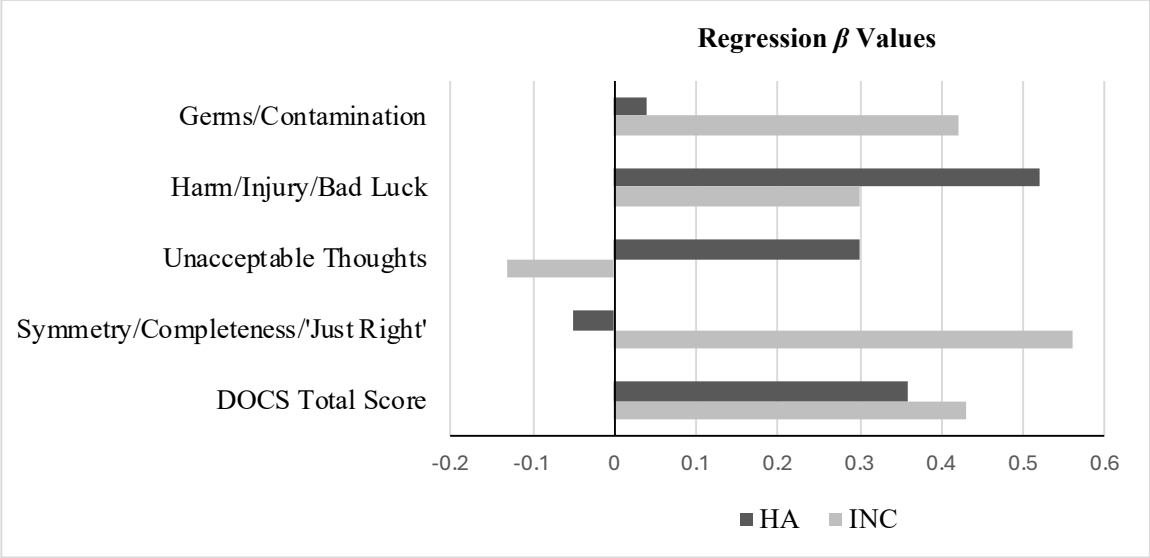
Regression Models: Core Dimensions as Predictors for Overt Symptom Dimensions

		Regression Statistics				
Variable Models	***R^2***	***B***	***SE B***	***β***	***t***	***p***
DOCS Germs/Contamination	.192	–	–	–	–	–
Trait HA		0.030	0.079	.04	0.382	.704
Trait INC		0.324	0.081	.42	3.983	< .001
DOCS Harm/Injury/Bad Luck	.349	–	–	–	–	–
Trait HA		0.360	0.066	.52	5.445	< .001
Trait INC		0.106	0.068	.15	1.56	.122
DOCS Unacceptable Thoughts	.077	–	–	–	–	–
Trait HA		0.234	0.087	.30	2.701	.008
Trait INC		-0.101	0.089	-.13	-1.129	.262
DOCS Symmetry/'Just Right'	.294	–	–	–	–	–
Trait HA		-0.037	0.069	-.05	-0.541	.590
Trait INC		0.403	0.071	.56	5.701	< .001
DOCS Total Score	.437	–	–	–	–	–
Trait HA		0.587	0.145	.36	4.044	< .001
Trait INC		0.733	0.15	.43	4.897	< .001

Note. DOCS = Dimensional Obsessive-Compulsive Scale; HA = core dimension harm avoidance; INC = core dimension incompleteness.

Figure A1

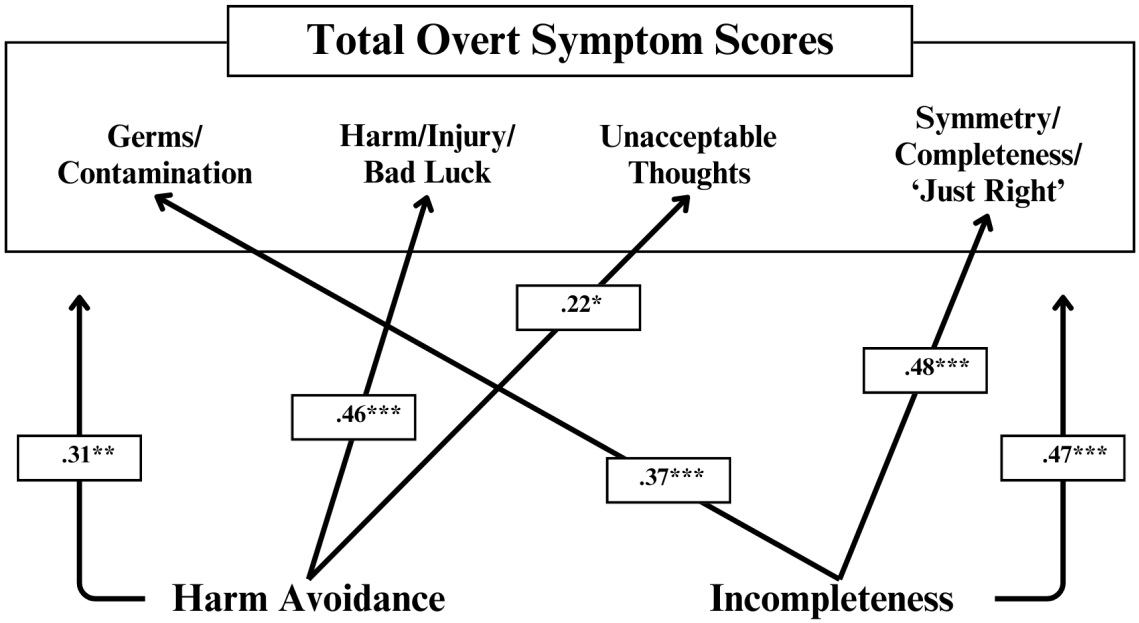
Study 1 Regression Model β Comparisons



Note. Y axis = DOCS dimensions and total score as regression model dependant variables; x-axis = regression beta coefficients; DOCS = Dimensional Obsessive-Compulsive Scale; HA = core dimension harm avoidance; INC = core dimension incompleteness.

Figure A2

Study 1 Partial Spearman Correlations



Note. Spearman partial correlations controlling for the opposite core dimension.
*p <.05; **p <.01; ***p <.001.

Appendix B

Study 2 Tables

Table B1

Study 2 Sample Demographics

Demographic Measures	Study 2 (*N=38*)		
	Mean	***SD***	**Min – Max**
Age	–	–	–
Current Age (TP2)	35.18	11.36	21 – 61
Age First Symptoms	10.08	6.2	1 – 33
Age of Onset (TP1)	15.29	6.32	5 – 34
Age First Helped	20.92	7.58	7 – 40
Age First Diagnosis	24.76	9.03	7 – 47
Elapsed Times	–	–	–
First Symptoms – Current Age (TP2)	25.37	12.1	8 – 53
Age of Onset (TP1) – Current age (TP2)	19.89	12.35	3 – 48
First Symptoms – First Diagnosis	14.95	9.35	0 – 39
Age of Onset (TP1) – First Diagnosis	9.47	9.3	0 – 33
First Diagnosis – Current Age (TP2)	10.42	9.91	0 – 34
	n (%)		
Biological Sex	–		
Male	8 (21)		
Female	29 (76)		
Other/Prefer Not to Say	1 (3)		
Gender	–		
Man	9 (24)		
Woman	27 (71)		
Other/Prefer Not to Say	2 (5)		
Ethnicity	–		
African/Black/Caribbean	1 (3)		
Asian (South, Central, East, West)	1 (3)		
Caucasian/White	34 (89)		
Hispanic/Latino	3 (8)		
Other	1 (3)		
Prefer Not to say	1 (3)		
Education	–		
High School	9 (24)		
Post Secondary (University/College)	20 (53)		
Graduate Studies (MA, PhD)	9 (24)		

Note. TP1 = time-point 1; TP2 = time-point 2.

Table B2

Study 2 Measure Descriptives

Measures	Study 2 (*N = 38; Subset of Study I*)					
	Mean	*SD*	Min – Max	*α* =	Possible Range	*n (%)*
Current (TP2) Overt Symptom Dimensions (DOCS)						
Germs/Contamination	6.11	6.23	0 – 19	.962	0 – 20	–
Harm/Injury/Bad Luck	10	6.41	0 – 20	.960	0 – 20	–
Unacceptable Thoughts	9.9	6.43	0 – 19	.954	0 – 20	–
Symmetry/Completeness/'Just Right'	8.26	5.83	0 – 20	.951	0 – 20	–
DOCS Total Score	34.26	16.67	9 – 72	.930	0 – 80	–
Cutoff Scores:	–	–	–	–	–	–
Clinical Cutoff (18+; Suggested Diagnostic Threshold)	–	–	–	–	–	31 (81)
Current (TP2) State Core Dimensions (OC-CDI)						
Current State Harm Avoidance	2.22	1.16	0 – 4	–	0 – 4	–
Current State Incompleteness	2.35	1	0.25 – 4	–	0 – 4	–
Retrospective (TP1) Overt Symptom Dimensions (DOCS)						
Germs/Contamination	5.74	7.37	0 – 20	.982	0 – 20	–
Harm/Injury/Bad Luck	10.95	6.34	0 – 19	.908	0 – 20	–
Unacceptable Thoughts	7.87	7.8	0 – 20	.971	0 – 20	–
Symmetry/Completeness/'Just Right'	6.61	5.59	0 – 17	.943	0 – 20	–
DOCS Total Score	31.15	14.8	6 – 61	.910	0 – 80	–
Cutoff Scores:	–	–	–	–	–	–
Clinical Cutoff (18+; Suggested Diagnostic Threshold)	–	–	–	–	–	31 (81)
Retrospective (TP1) State Core Dimensions (OC-CDI)						
Past State Harm Avoidance	2.47	1.34	0 – 4	–	0 – 4	–
Past State Incompleteness	2.31	1.14	0 – 4	–	0 – 4	–
Within-Dimension Qualitative Changes (TP1 – TP2)						
Yes	–	–	–	–	–	36 (95)
No	–	–	–	–	–	2 (5)
Difficulty in Emotion Regulation Scale (DERS - 16)						
Subscale 1 (Clarity)	5.87	2.67	2 – 10	.910	2 – 10	–
Subscale 2 (Goals)	11.763	3.09	5 – 15	.824	3 – 15	–
Subscale 3 (Impulse)	8.105	3.623	3 – 15	.838	3 – 15	–
Subscale 4 (Strategies)	17.4	5.33	5 – 25	.886	5 – 25	–
Subscale 5 (Nonacceptance)	10.11	3.59	4 –15	.863	3 – 15	–
DERS - 16 Total Score	53.24	15.78	22 – 80	.948	16 – 80	–
Toronto Alexithymia Scale (TAS-20)						
Subscale 1 (Identifying Feelings)	20.5	7.93	7 – 35	.915	7 – 35	–
Subscale 2 (Describing Feelings)	13.763	5.69	5 – 25	.851	5 – 25	–
Subscale 3 (Externally Oriented Thinking)	18.03	5.26	8 – 28	.690	8 – 40	–
TAS-20 Total Score	52.29	15.76	20 – 86	.905	20 – 100	–
Cutoff Scores:	–	–	–	–	–	–
Non-Alexithymia (< 51)	–	–	–	–	–	17 (45)
Possible Alexithymia (52-60)	–	–	–	–	–	9 (24)
Alexithymia (61+)	–	–	–	–	–	12 (31)
Stressful Life Events Questionnaire						
Total Score	16.32	7.16	4 – 30	.739	0 – 32	–
Treatment History Questionnaire						
Medication/Pharmacological	3.605	0.718	1 – 4	–	0 – 4	–
Therapeutic/Psychological	2.6	0.775	1 – 4	–	0 – 4	–
Treatment History Total Score	6.289	1.113	3 – 8	–	0 – 8	–
Estimated Total Years	12.72	11.01	1 – 35	–	–	–

Note. OC-CDI = Obsessive-Compulsive Core-Dimension Interview; DOCS = Dimensional Obsessive-Compulsive Scale; DERS = Difficulty with Emotion Regulation Scale; TAS = Toronto Alexithymia Scale.

Table B3

Study 1 and 2 Sample Mean Comparison

Measure	Study 1 (*N* = 90)		Study 2 (*N* = 38)			
	M	***SD***	***M***	***SD***	***t***	***p***
Age Demographics						
Current Age	33.32	11.47	35.18	11.36	0.84	.402
Age First Symptoms	9.99	5.69	10.08	6.20	0.08	.937
Age of Onset	16.10	6.48	15.29	6.32	0.65	.516
Age First Helped	21.16	7.54	20.92	7.58	0.16	.870
Age First Diagnosis	24.38	8.84	24.76	9.03	0.22	.826
Emotion Regulation	–	–	–	–	–	–
DERS – 16 Total Score	56.55	13.52	53.24	15.78	1.20	.231
Alexithymia	–	–	–	–	–	–
TAS – 20 Total Score	53.61	14.97	52.29	15.76	0.45	.654
DOCS Overt Symptoms	–	–	–	–	–	–
DOCS Total Score	32.7	14.02	34.26	16.67	0.54	.588

Note. DOCS = Dimensional Obsessive-Compulsive Scale. Two-sample t-tests calculated based on sample mean, *N*, and standard deviation.

Table B4

Dimension Mean Stability: Significant Mean Change from Time 1 to Time 2

Measure	Mean Comparisons (TP1–TP2)		
	W	***z***	***p***
DOCS Overt Symptom Dimensions	–	–	–
Germs/Contamination	178.0	-0.264	.800
Harm/Injury/Bad Luck	359.5	1.412	.160
Unacceptable Thoughts	177.0	-1.627	.105
Symmetry/Completeness/'Just Right'	190.0	-1.384	.168
DOCS Total Score	346.0	-0.355	.728
OC-CDI State Core Dimensions	–	–	–
State Harm Avoidance	412.0	1.589	.113
State Incompleteness	291.0	-0.111	.918

Note. DOCS = Dimensional Obsessive-Compulsive Scale; OC-CDI = Obsessive-Compulsive Core Dimensions Interview; TP1 = time point 1; TP2 = time point 2. Mean comparisons made using Wilcoxon signed-rank tests.

Table B5

Dimension Mean Stability: Mean Change Scores

Measure	Raw Mean		Standardized Mean (Z-Score)
	Difference (*SD*)	Min-Max	Difference (*SD*)
DOCS Overt Symptom Dimensions	–	–	–
Germs/Contamination	0.37 (6.61)	-15 – 17	0.061 (1.10)
Harm/Injury/Bad Luck	-0.95 (7.57)	-14 – 20	-0.157 (1.26)
Unacceptable Thoughts	2.03 (6.95)	-14 – 17	0.336 (1.15)
Symmetry/Completeness/'Just Right'	1.66 (6.30)	-11 – 19	0.275 (1.05)
DOCS Total Score	3.11 (17.756)	-29 – 44	–
OC-CDI State Core Dimensions	–	–	–
State Harm Avoidance	-0.24 (1.21)	-2.75 – 3	-0.273 (1.37)
State Incompleteness	0.04 (1.12)	-2 – 2.67	0.045 (1.26)

Note. DOCS = Dimensional Obsessive-Compulsive Scale; OC-CDI = Obsessive-Compulsive Core Dimensions Interview. Positive scores indicate an increase in dimension severity over time.

Table B6

Dimension Rank-Order Stability

Measure	Spearman Test–Retest	
	r_s	*p*
DOCS Overt Symptom Dimensions	–	–
Germs/Contamination	.51	<.001
Harm/Injury/Bad Luck	.28	.09
Unacceptable Thoughts	.53	<.001
Symmetry/Completeness/'Just right'	.42	.009
DOCS Total Score	.41	.012
OC-CDI State Core Dimensions	–	–
State Harm Avoidance	.54	<.001
State Incompleteness	.46	.003

Note. DOCS = Dimensional Obsessive-Compulsive Scale; OC-CDI = Obsessive-Compulsive Core Dimensions Interview.

Table B7

Ipsative Stability in the Overt Symptom and Core Dimension Models

Dimension Profile	Double-Entry Intraclass			Pearson		
	ICC_{DE}	*SD*	Min – Max	*r*	*SD*	Min - Max
Overt Symptom (DOCS)	.36	.44	-.48 – .99	0.39	.52	-.78 –.99
Core Dimension (State OC-CDI)	.40	.69	-.98 – 1	–	–	–

Note. DOCS = Dimensional Obsessive-Compulsive Scale; OC-CDI = Obsessive-Compulsive Core Dimensions Questionnaire. Ipsative profiles have been normed to their respective means, meaning ICC_{DE} values are controlling for changes in profile elevation. Regular Pearson correlations not calculated for the core dimensions, as they represent < 3 variables per time point.

Table B8

Significant Exploratory Zero-Order Spearman Correlations

	Profile Stability		Mean Difference Scores					
Correlate Measures	**Overt**	**Core**	**DOCS 1**	**DOCS 2**	**DOCS 3**	**DOCS 4**	**State HA**	**State INC**
DERS Clarity	–	–		.40*	–	–	–	–
DERS Goals	–	–	.40*	.47**	–	–	–	.42**
DERS Impulse	–	–	–	–	–	–	–	.47**
DERS Strategies	–	–	–	.44**	.33*	–	.41*	.40*
DERS Nonacceptance	–	–	–	.47**	–	–	–	–
TAS Identifying	–	–	–	.42**	–	–	.47**	–
TAS Describing	–	–	–	–	–	–	.35*	–
Medication Quality	.39*	–	–	–	–	–	–	–
Therapy Quality	–	–	–	–	–	–	–	-.34*
Diagnosis-Current Age (TP2)	.34*	–	–	–		–	–	–
Age of Onset (TP1) - First Diagnosis	–	–	–	–	.33*	–	–	–
Current Age	.35*	–	–	–	–	–	–	–

Note. Overt = Overt symptom model dimensions; Core = core dimension model dimensions; DOCS = Dimensional Obsessive-Compulsive Scale; DOCS 1 = germs/contamination; DOCS 2 = harm/injury/bad luck; DOCS 3 = unacceptable thoughts; DOCS 4 = symmetry/completeness/'just right'; HA = core dimension harm avoidance; INC = core dimension incompleteness; DERS = Difficulty with Emotion Regulation Scale; TAS = Toronto Alexithymia Scale; TP1 = time-point 1; TP2 = time-point 2. Higher correlations in difference scores suggest lower stability. Table only shows significant correlations. Correlations calculated using Spearman's *rho*. *p <.05; **p <.01; ***p <.001.

Appendix C

The Double-Entry Intraclass Correlation

The double-entry intraclass correlation is a novel approach unique in its ability to measure the stability of a profile of dimensions at the level of an individual (see Furr 2010; McCrae, 2008; Taylor, 2010). The ICC_{DE} has typically been used in personality psychology to assess person-level stability across multidimensional profiles such as the "Big Five" personality inventory (e.g., Atherton et al., 2022; Poier, 2022; Terracciano et al., 2010). In this analysis, within-person (i.e., across-dimension) profile scores are correlated from TP1 to TP2 much like a typical Pearson's test–retest correlation. However, in the ICC_{DE}, prior to the correlation, the TP2 scores are appended to TP1 and the TP1 scores to TP2 (Furr, 2010). This analysis returns a Pearson 'stability coefficient' (-1–1, with higher values indicating greater stability) that stands as an omnibus measure of stability in the profile's shape (i.e., the rank order of the dimensions), scatter (i.e., the variance of each dimension from the mean), and elevation (i.e., the mean across all dimension scores; for a more detailed description of each facet, see Furr, 2010).

For this Study 2, individual ICC_{DE} correlation coefficients were generated for each participant for both the overt symptoms model (i.e., across four dimensions), and the core dimensions model (across two dimensions). Based on the way the measures were utilized—where a valid score for the core dimensions relied on symptoms to be present within the respective DOCS category—any instance where no symptoms were recorded for the DOCS meant that the score on the OC-CDI for that dimension was changed to N/A instead of a zero. The core dimension scores for TP1 and TP2 thus represented an average score for each time point after removing scores for the DOCS

dimensions with no reported symptoms. That is, if the participant only reported symptoms in two of the DOCS dimensions at TP1, the OC-CDI scores for that time point would be based on an average of two values for HA and two for INC, instead of the typical four. Failure to remove the OC-CDI scores of zero where no DOCS categories were endorsed, artificially inflated the profile stability of the core dimensions. Prior to the ICC_{DE} analyses, raw scores for each dimension were also standardized to z-scores based on their respective dimension means and standard deviations (as recommended by McRae, 2008, p. 106) which acted to guard against inflated stability coefficients based on normative trends across dimension means. The resulting correlation coefficients were then averaged across participants for each model, allowing for a comparison of the models' relative stability. Individuals' personal ICC_{DE} stability coefficients were further used for the subsequent exploratory analyses.

Due to the propensity of overall OCD symptom severity to wax and wane with normal environmental changes or life stress (e.g., Adams et al., 2018; American Psychiatric Association, 2013; Naftalovich et al., 2021), I decided to modify the approach described by Furr (2010) to control for changes in elevation by centering all profile scores for the ICC_{DE} by their respective timepoint means. This decision was reinforced by the observed aggressive effect changes in elevation had on the degradation of the correlation coefficient (see Furr 2010, figure 3b) and because the research question is more accurately addressed by observing relational stability among profile factors (i.e., how stable their relationships are with each other) than by overall changes in symptom severity. To be more concise, I decided it more important to measure relational changes among dimensions—as they are theorized to be uniquely anchored by stable etiological

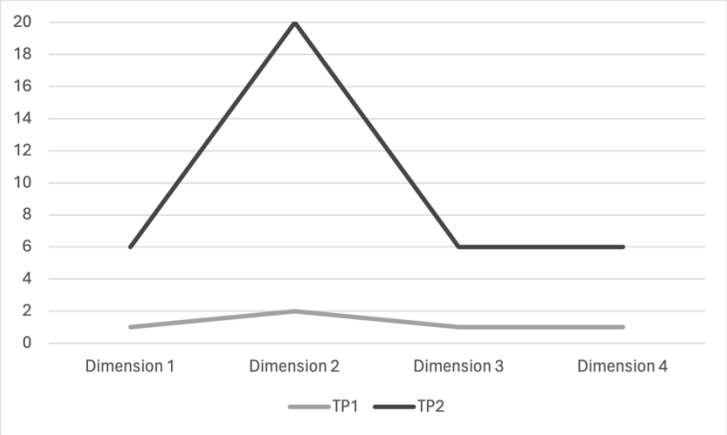
factors—than to have measures of stability affected by universal changes in symptom severity, as symptoms are known to wax-and-wane normally over time and in response to environmental stressors. Thus, for this analysis, I chose to control for universal symptom severity change. While I believe that that the ICC_{DE} is the most appropriate tool for this analysis and controlling for elevation is a required step given this application of the analysis, the modification mentioned limits direct comparability of findings of stability to those elsewhere in the literature. These limitations are discussed in Study 2.

Case Examples, and Modifications

The following examples highlight and justify the modifications made to the ICC_{DE} analyses for this study. Furr (2010) generally advocated the use of a regular Pearson intraclass correlation between the profile points of TP1 and TP2 as opposed to the double-entry method. The argument he made was that typically, the most important variable for determining stability is that of profile shape over time. This might be true for many research questions or measures. However, the broad theory underlying my research questions is that specific OCD symptom domains are influenced and/or anchored by other stable etiological factors (i.e., genetics, neural structure) and therefore, observing the relationships between all dimension scores within the larger profile models themselves should also be informative. Figure C1 demonstrates the issues inherent with considering only profile shape.

Figure C1

The Limitations of Only Considering Profile Shape



Note. Y-axis = dimension severity scores; x-axis = symptom dimensions; TP1 = time point 1; TP2 = time point 2.

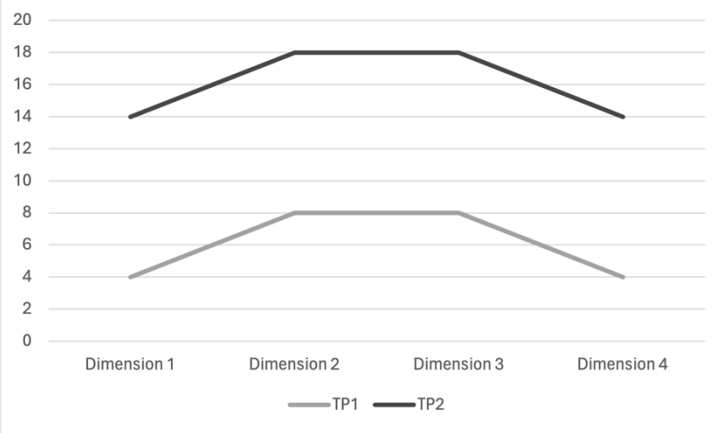
In this example, when only considering shape by using a regular Pearson correlation, TP1 (1,2,1,1) and TP2 (6,20,6,6) correlate perfectly ($r = 1$). Looking at these two profiles, there is clearly unique and informative change happening in dimension 2 compared to the other three dimensions that goes unrecorded. Relatedly, the overall rise in symptoms altogether go unrecorded when only considering shape. When using a basic Pearson correlation, despite these changes in the relative relationships between the dimensions themselves, the correlation is still considered 'perfect' because the shape stays the same for each profile. For this study, I did not consider this outcome to be optimal, as the resulting coefficients would have been less sensitive to these important facets of change and would have overestimated and overrepresented the interpretation of ipsative stability. Furthermore, because the double-entry process requires the opposing time-point data to be appended to each profile, the ICC_{DE} allows for the same statistical approach to be used with the two-factor core dimensions profiles, which only have two

variables at each time point; one for HA and one for INC. Running a basic Pearson correlation is not feasible with less than three data points in each profile.

The ICC_{DE} is unique in that it considers changes in shape, scatter, and elevation, which means that it would account for the changes in the relationships of the variables from the example above. What I found; however, was that change in overall elevation had an aggressive effect on the resulting correlation coefficients (See Furr, 2010, p. 5 for more on the individual contribution of each profile characteristic). Figure C2 illustrates the effect of elevation changes.

Figure C2

The Effect of Changes in Elevation



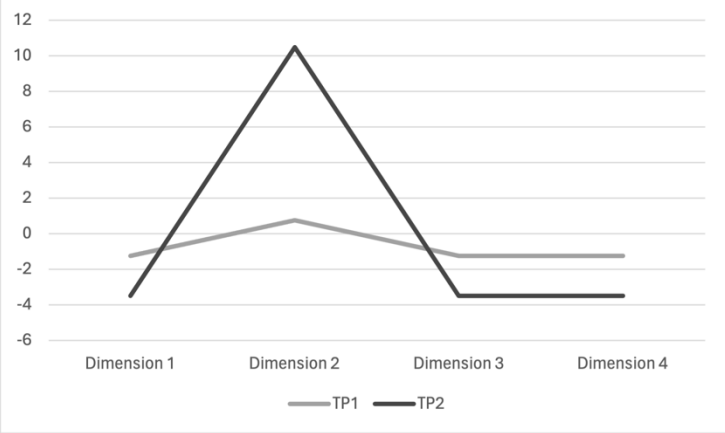
Note. Y-axis = dimension severity scores; x-axis = symptom dimensions; TP1 = time point 1; TP2 = time point 2.

In this example, the shape and scatter are the same for both profiles. TP1(4,8,8,4) and TP2 (14,18,18,14) only differ in their profile means (i.e., their elevations). Using the ICC_{DE}, the coefficient returned for this example is r = -.72, which is difficult to interpret given the purpose of the research question. While the relationship between the variables—or the profile's overall structure—remained identical, the symptoms in this

example show a clear uniform increase over time. The influence that changes in elevation have over the final stability coefficient in this analysis was overpowering in this example to the point of driving the resulting correlation statistic far into the negative. With the known and expected severity fluctuations in symptom course for OCD, I decided that it would be best to control for changes in elevation by centering each set of profile scores around their respective means. Not doing so would have strongly underestimated and underrepresented interpretations of stability. The final example, Figure C3, demonstrates the modification I made to the original ICC_{DE} approach reported in Furr (2010).

Figure C3

Centering Profiles by their Mean



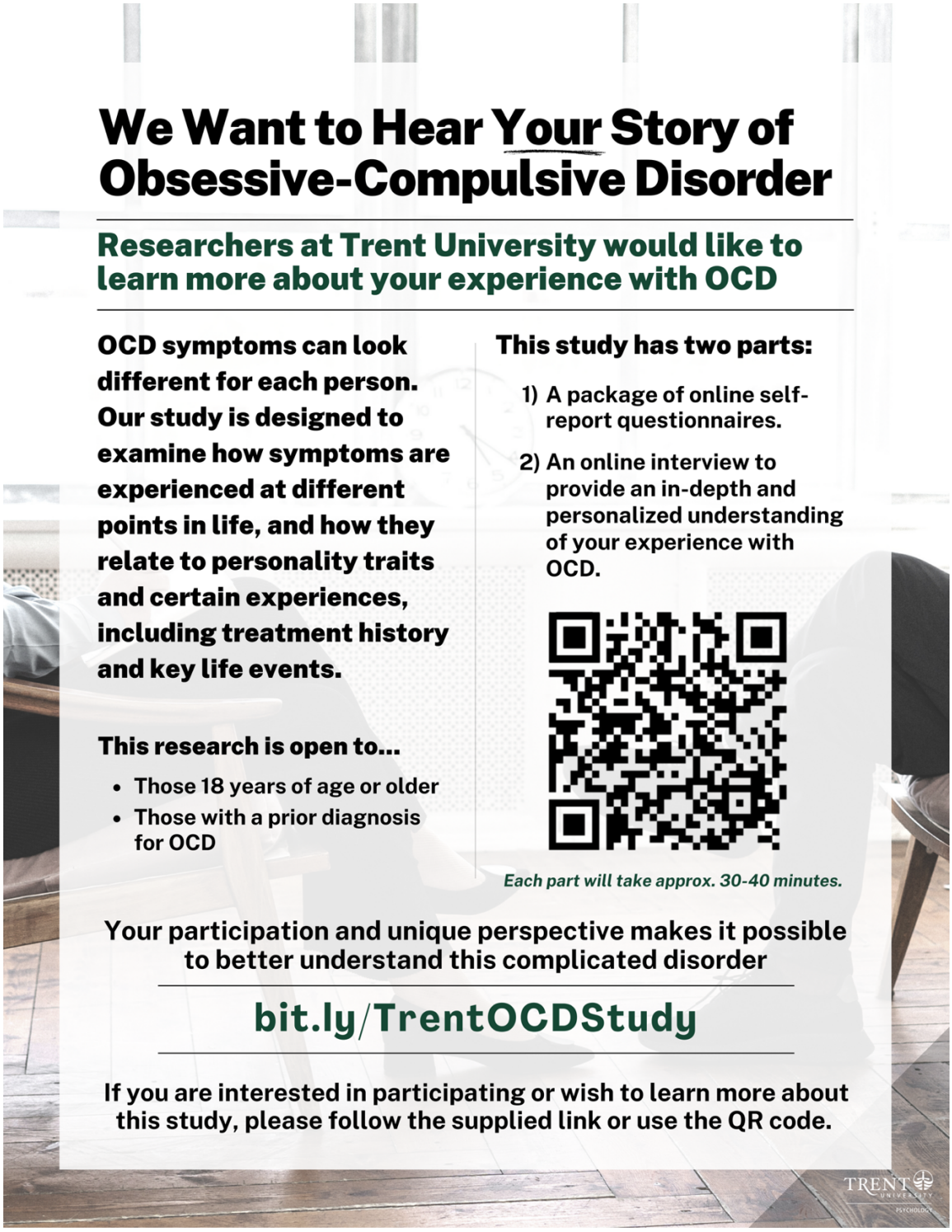
Note. Y-axis = dimension severity scores; x-axis = symptom dimensions; TP1 = time point 1; TP2 = time point 2.

This final example recreates the graph from Figure C1 and centers each time point by their respective mean. By doing this prior to calculating the ICC_{DE}, the correlation coefficient is created through consideration of the profiles' shape and scatter, while controlling for changes in elevation. In the first example, the Pearson *r* was 1; the ICC_{DE} without controlling for elevation would have been r_{ICCDE} = -.41; and the ICC_{DE}

after controlling for elevation is r_{ICCDE} = .27. After running numerous models of possible combinations of TP1 to TP2 scores, I considered this approach the most appropriate for maintaining meaningful and interpretable correlations, while still maintaining sensitivity to between-dimension relationship change within the profile over time. Because overall changes in symptom severity were not important for answering the specific research question, the ICC_{DE} analyses were conducted using this final method.

Appendix D

Study Advertisement



Appendix E

Created and Modified Measures

Life Events Questionnaire

Measure Prompt

"Thinking back to the time period between when your obsessive-compulsive symptoms first became problematic enough to interfere with your normal functioning until now, please identify and rate the following experiences, if applicable, based on how negative (e.g., stressful, painful, unpleasant, overwhelming) the experience was for you."

Stressful Life Events

1. A painful break-up with a boyfriend/girlfriend, partner, or spouse
2. A serious problem at school or at work
3. The death of someone close to you
4. The illness of someone close to you
5. The divorce or separation of your parents
6. A serious money problem
7. A serious health problem (i.e., personal)
8. Other

Scoring Metric

- Not applicable – 0
- Not very negative –1
- Negative but manageable – 2
- Negative to the point of being disruptive – 3
- Extremely negative – 4

Scores are only applicable if the event happened during the targeted time frame (i.e., age of onset [TP1]–current day [TP2]). If participants experienced many examples within the categories above (e.g., multiple problems at school or work), they were instructed to speak to the one that was the most affective. Scores from each item were summed to create an overall stressful life events score (0–32) which represents a rough estimate of the breadth and subsequent severity of numerous experienced stressful live events.

Treatment History Questionnaire

Measure Prompt

"The following three questions will ask about your treatment history. You do not need to provide any specifics beyond what you are comfortable with. Specifics, such as names of certain medications, are acceptable answers but not required. If you cannot remember specific names or are more comfortable speaking to the class of medication (e.g., SSRI), you may choose to do so."

Treatment History Questions

1. **Pharmacological Supports.** "Within this time period [TP1–TP2], have you regularly taken (i.e., more than once) any medications, recreational drugs, or supplements – formally prescribed or otherwise – to purposefully attend to, control, lessen, or support your OCD or OCD-related symptoms? If so, what were they?"
2. **Psychological Supports.** "Within this time period [TP1–TP2], have you regularly utilized (i.e., more than once) any therapy, social support (group or singular), or other therapeutic modalities – formally prescribed or otherwise – to purposefully attend to, control, lessen, or support your OCD or OCD-related symptoms? If so, what were they?"
3. **Overall Time Spent Utilizing Supports.** "Overall, combining the methods mentioned above: how many years, in total, have you actively utilized treatment to purposefully attend to, or control, your OCD or OCD-related symptoms?"

Scoring Metric

Scoring of the qualitative answers from questions 1 and 2 utilized a coding system based on a 4-point Likert scale:

- No resources used – 0
- Non-psychiatric/psychological resources used (e.g., recreational drugs/supplements; religious/social support) – 1
- Psychiatric/psychological resources used but not considered first-line approaches for OCD (e.g., benzodiazepines; emotion-focused therapies) – 2
- Psychiatric/psychological resources used that represent first-line approaches for OCD (e.g., selective-serotonin reuptake inhibitors; exposure therapies with response prevention) – 3
- Multiple psychiatric/psychological resources used that represent first-line approaches for OCD – 4

Scores from these questions were considered separately (i.e., Pharmacotherapy and Psychotherapy; 0-4) and combined as a total treatment quality score (0-8). Question 3 was recorded directly as participants answered it.