Table of Contents

[Introduction 2](#_Toc191563590)

[Obsessive-Compulsive Disorder 2](#_Toc191563591)

[Problem statement 2](#_Toc191563592)

[Informant Discrepancies 4](#_Toc191563593)

[Modelling informant discrepancies 6](#_Toc191563594)

[Parent Rating 8](#_Toc191563595)

[Agreement as a test of validity for multiple informant 8](#_Toc191563596)

[MRI 9](#_Toc191563597)

[Voxel based morphometry 9](#_Toc191563598)

[Machine Learning 10](#_Toc191563599)

[The present study 11](#_Toc191563600)

[Hypothesis 11](#_Toc191563601)

[Methods 12](#_Toc191563602)

[Sample 12](#_Toc191563603)

[Clinical characteristics: 12](#_Toc191563604)

[Demographic characteristics: 13](#_Toc191563605)

[Itemwise CBCL OCS characterization: 13](#_Toc191563606)

[Comparison parent vs child reports (CBCL): 14](#_Toc191563607)

[Study Design 15](#_Toc191563608)

[Data acquisition 16](#_Toc191563609)

[Statistical analyses/Preliminary analyses(?) 19](#_Toc191563610)

[Modelling approach(?) 19](#_Toc191563611)

Introduction

Obsessive-Compulsive Disorder

OCD is recognized as a prevalent and persistent neuropsychiatric condition, impacting an estimated 2% to 3% of individuals worldwide (de Mathis et al., 2013). The disorder commonly arises in early life and is characterized by the presence of compulsions – ritualized behavioral or mental acts, and obsessions – intrusive and unwanted thoughts and worries (Karno et al., 1988). OCD is unique among mental illnesses in that it exhibits both externalizing and internalizing symptoms (Guzick et al., 2019). Externalizing features, like compulsivity and repetitive actions, are often outwardly disruptive and align with disorders such as attention-deficit/hyperactivity disorder (ADHD) and disruptive behavior disorders. In contrast, internalizing aspects, including anxiety, concerns, and obsessions, cause internal distress and align with conditions like depressive and anxiety disorders, often leading to avoidance or withdrawal (Achenbach, 2001). Understanding OCD within this dual framework enhances our grasp of its complexity and informs more effective therapeutic strategies. These frameworks are not only therapeutically beneficial but are also supported by empirical research (Kessler et al., 2011; Slade & Watson, 2006).

Liam (12 years) had been struggling with severe Obsessive-Compulsive Disorder (OCD) for several years. After starting therapy, he was showing signs of improvement. According to Liam, he felt he was making excellent progress. He managed to reduce his handwashing rituals from every hour to three times a day and had started joining some family meals. He was also beginning to meet his friends for short walks around the neighborhood. However, his parents observed a different reality. While Liam had made some progress, he often became trapped in lengthy rituals that caused him significant distress. He had yet to return to school full time, attending only partial days if he went at all. Though he started venturing out with friends, it was only to familiar, controlled environments. The family had rearranged the house to facilitate his progress—stocking rooms with hand sanitizers and scheduling regular therapist meetings to help him during challenging times. His parents continued to monitor his progress closely, supporting him in his journey while remaining aware of the continuous obstacles that his OCD presented.

This vignette demonstrates the importance of recognizing that the child's perspective is distinct, but equally valid. Traditionally, clinicians have depended on parents to supply comprehensive information about how an illness and its treatment affect their children. This reliance stems from the perception that children may not possess the cognitive and linguistic skills required to accurately understand and respond to surveys. However, the insights provided by Liam can differ significantly from those of his mother, highlighting the potential discrepancies in information regardless of whether the goal is clinical assessment or research. From this evaluation, we must make clear how parent and child assessments of mental health relate to one another.

Additionally, mental health problems can vary across different contexts (Bauducco et al., 2024; Beesdo et al., 2009). Children and adolescents may exhibit mental health concerns in certain environments, such as at home or school, while appearing unaffected in others, like during peer interactions. These contextual variations are evident across various domains, including conduct problems, attention, hyperactivity, and anxiety (Beesdo et al., 2009). Consequently, the source of information—whether from self, parents, other family members, healthcare professionals, or teachers—can lead to differing perceptions and understandings of the child's condition.

In parallel, advances in neuroimaging, particularly structural MRI, have elucidated the brain's role in OCD, pointing to abnormalities within the cortico-striato-thalamo-cortical circuit and other key regions. Such findings suggest that neuroimaging biomarkers hold promise for enhancing diagnostic accuracy and understanding the neurobiological underpinnings of OCD. Despite these advancements, the integration of multi-informant assessments and neuroimaging data in pediatric OCD research remains limited.(GPT)

This divergence in perspectives introduces complexities into clinical practice, research, and theory regarding child psychiatry and psychopathology(Chen et al., 2017; Salbach-Andrae et al., 2009). These differences have been thoroughly examined and will be covered in more detail below (Van Roy et al., 2010). Furthermore, while the use of multiple informants in mental health assessment is thought to enhance our understanding of the psychological functioning of children, particularly in the infant population, we are still in the process of discovering how to effectively utilize this wealth of information (Reyes, 2013).

Informant Discrepancies

The issue of informant discrepancies is particularly pertinent when interpreting study findings in the field of developmental psychopathology. A significant portion of the evidence about prevalence rates of psychological disorders, classification of diagnosis, effectiveness of interventions for children is derived from reports by multiple informants (Weisz et al., 2005). For example, the prevalence rates of conduct and oppositional defiant disorders in community samples range from 1.6% to 10.2%, depending on whether parent or teacher ratings are relied on to classify disorder in the child or whether both are considered simultaneously (Offord et al., 1996). Prevalence of classification of disorder ranges widely in clinic samples as well. When relying on parent or teacher ratings, or combining information from both, prevalence of conduct disorder ranges from 9.7% to 23%, and emotional disorder (anxiety, depression) ranges from 10.3% to 36.2% (MacLeod et al., 1999) (De Los Reyes & Kazdin, 2005). Furthermore, depending on the informant, it is typical to find inconsistent results from controlled studies evaluating psychological therapies (De Los Reyes & Kazdin, 2005). Understanding these discrepancies is crucial for accurately assessing intervention outcomes and advancing research in developmental psychopathology.

The phenomenon of informant discrepancy has been recognized for nearly 70 years, dating back to Lapouse and Monk's work in 1958. Achenbach, McConaughy, and Howell (1987) conducted a seminal analysis of 119 studies investigating these informant inconsistencies. Their key findings included: (a) reports of the same behavior by different informants generally show low to moderate agreement; (b) the reports of two informants observing children in the same setting are more similar than those of two informants observing children in different settings; (c) there is greater agreement between informants' reports for younger children compared to older ones; and (d) reports of externalizing behaviors like aggression show higher consistency than those of internalizing behaviors such as anxiety. They concluded by stating, "Different informants are needed for different situations. . . there is no royal road or preeminent gold standard for phenomena that are inevitably affected by assessment procedures and other situational variables" (p. 227–228). Consequently, the primary objectives of the informant discrepancies research summarized by Achenbach et al. (1987) were to outline the extent of informant discrepancies, identify the informant pairs (e.g., parent and child, teacher and parent) with the greatest discrepancies, and pinpoint the behavioral domains where these discrepancies were most pronounced.

Research conducted since 1987 has primarily focused on understanding the characteristics associated with informant discrepancies, potential biases in informants' reports, and the impact these discrepancies have on the conclusions drawn from research studies. (Reyes, 2013)

Regarding children, it's possible that the particular domains being taken into consideration affect how accurate proxy ratings are. Therefore, gathering information about a child’s functioning typically involves input from multiple informants (Achenbach, 2006).

Modelling informant discrepancies

It would be a mistake to conclude that no work since Achenbach et al. (1987) has improved our understanding of informant discrepancies. Prior theoretical and methodological work has provided researchers guidance on how to statistically model multi-informant reports. For example, Kraemer et al. (2003) modeled informant discrepancies as a function of three components: (a) variation in the perspectives informants have of children’s behavior (e.g., self vs. other), (b) variation in children’s behavior across meaningful settings (e.g., home vs. school), and (c) and the extent to which the behaviors assessed are consistently expressed across informants’ perspectives and settings. Thus, in the model proposed by Kraemer et al. (2003), parents’ reports are modeled as arising from an observer perspective of children’s behavior expressed in the home context, teachers’ reports as arising from an observer perspective of behavior expressed in a nonhome context (i.e., school), and children’s reports as arising from a self perspective of behavior expressed in both home and nonhome contexts. The implication of this methodological approach is that when informants provide discrepant reports, these discrepancies may point to a meaningful interaction between two factors. The first is that informants vary systematically in where they observe children’s behavior. The second is that children vary systematically in where they express the behaviors being assessed. Recent theoretical and methodological work has also focused on how informant discrepancies may be incorporated within frameworks seeking to explain how youth maladjustment develops. For example, recent theoretical work has illuminated how discrepancies between parent and youth reports of youth victimization may represent features of the relationship between parents and youths that increase risk for poor youth outcomes (Goodman et al., 2010). In addition, in the behavior genetics literature, researchers have taken structural equation modeling approaches to managing multi-informant reports that allow for testing whether nonshared variance among informants’ reports reflects unique information being contributed by informants’ reports rather than mere measurement error (Baker, Jacobson Raine, Lozano, & Bezdjian, 2007; Bartels, Boomsma, Hudziak, van Beijsterveldt, & van den Oord, 2007; Derks et al., 2006; Tackett, Waldman & Lahey, 2009). In this literature, informant discrepancies are modeled not as “systematic error terms” but as valuable systematic variation. Specifically, variation unique to specific informants’ reports represents contextual differences in the contributions of one’s environment as well as genetic predispositions to variance in expressions of specific behaviors. In other words, researchers have partitioned variance to reflect genetic influences on behavior, shared environmental influences, and nonshared environmental influences, in addition to measurement error (Bartels et al., 2007). As a consequence, informant discrepan cies in the behavior genetics literature yield insights as to the extent to which children’s environments affect their behavior differently, depending on where the informants providing reports observe children’s behavior (e.g., Derks et al., 2006). Sound theoretical and methodological models of informant discrepancies are necessary to advance understanding of informant discrepancies. Recent work has made great strides in this regard. At the same time, such modeling may be insufficient to gain a deeper understanding of how to conduct multi-informant assessments and interpret the informant discrepancies that often arise from these assessments. Two observations of the clinical literature support this contention… (Reyes, 2013)

Furthermore, studies in the 1970s indicated that girls tend to be more reliable informants than boys [14]. (Caqueo-Urízar et al., 2022)

Additionally, there are certain contextual factors that associate with higher agreement or discrepancy between reports. It has been observed that when there is a higher socioeconomic level, parents tend to underestimate the problems related to their children’s mental health, while in lower socioeconomic levels, the opposite happens [22]. Furthermore, family factors, such as parenting style, lack of communication, and conflict between parents and children, have been associated with higher levels of discrepancy in both reports [23–32], while family cohesion and parental acceptance have shown fewer discrepancies [9,25]. Traditionally, these observed differences have been interpreted as a function of measurement errors and informant bias [26]. However, such discrepancies may be significant to understand the nature and course of child and adolescent psychopathology, as they may reflect underlying family problems, which potentially contribute to the development of psychopathologies [27]. (Caqueo-Urízar et al., 2022)

Parent Rating

Moreover, when externalized problems are analyzed, it seems that parents tend to be more precise than their children; however, when internalized symptoms are analyzed, there seems to be less agreement about which group reports symptoms better [15]. This disagreement stems from the fact that children value their behavior more positively than parents do [10]; although, it becomes more complicated towards adolescence, where they seem to report poorer health than parents, especially in emotional health [16]. Various studies support that these discrepancies are based on an underestimation of anxious and depressive symptoms by the parents [17–20]. (Caqueo-Urízar et al., 2022)

However, when externalizing symptoms such as defiant behavior or hyperactivity are under evaluation, some studies have found a higher level of agreement between both groups [9,21]. (Caqueo-Urízar et al., 2022)

In that sense, a meta-analysis carried out by Los Reyes et al. [4] including 341 studies published between 1989 and 2014, observed low-to-moderate correspondence between children’s self-report and parents’ report (mean internalizing: r = 0.25; mean externalizing: r = 0.30; mean overall: r = 0.28). (Caqueo-Urízar et al., 2022)

Research indicates discrepancies and varying accuracy in symptom reporting, with no clear consensus. Child reports internalized symptoms more accurately, while parents tend to be more precise in identifying externalized (Silverman & Eisen, 1992).

However, the reliability of parent reports for assessing children's experiences, especially for non-observable functions like emotions, has been questioned (Eiser & Morse, 2001). Parental assessments often differ from children’s self-perceptions, potentially due to biases, superficial observations, or the nature of the parent-child relationship. Conversely, children frequently lack objective self-perception (Barrett et al., 1991; Martin et al., 2004).

Agreement as a test of validity for multiple informant

It can be challenging for clinicians to make a diagnosis if data from different sources are conflicting. When assessing children’s behavioral and emotional problems, informants tend to disagree [30,31]. Disagreement between children and their parents about target problems can be problematic when it comes to setting treatment goals, which can, ultimately, lead to poorer treatment outcomes [35]. Studying moderators of parent–youth agreement may facilitate diagnostic processes [36]. Research suggests that agreement between parents and children is related to factors such as a child’s age [31,35], gender [37,38] and type and severity of disorder – especially anxiety and depression [39] – and parental psychopathology [39–42]. Studying those moderators may guide clinicians in assessing which reports have greater veracity [35,42,43]. (Jónsdóttir et al., 2022)

MRI

Voxel based morphometry

Meta-analyses of voxelbased morphometry (VBM) studies have found altered volumes in frontal and striatal regions, and in a broader range of structures, including parietal and limbic brain areas.3,4 A recent VBM mega-analysis on pooled raw magnetic resonance imaging (MRI) data from our multisite OCD Brain Imaging Consortium (OBIC), revealed significantly smaller volumes of frontal grey matter and white matter bilaterally, as well as group6age interactions in frontostriatal and limbic regions.5 Of particular interest in the meta-analyses and VBM mega-analysis is the finding of decreased grey matter volume in the dorsomedial prefrontal cortex,6,7 a region responsible for performance monitoring and emotional processing, processes which are affected in OCD.8 (Fouche et al., 2017)

Although we performed a VBM analysis on this data previously, other techniques such as surface-based methods can provide complementary information. Whereas VBM measures grey matter volume or density, surface-based methods such as FreeSurfer can calculate morphometric attributes in the native space of the participant, and allow a determination of cortical thickness.9

In addition, segmentation in VBM is suboptimal for some subcortical areas, such as pallidum and thalamus, because of the lack of clear grey–white contrast of these structures.

In recent years, magnetic resonance imaging (MRI) studies have provided numerous evidence of functional and structural abnormalities in various brain regions in OCD, mainly within the cortico-striato-thalamo-cortical (CSTC) circuit [8–12] (Hu et al., 2017). The success of neuroimaging in revealing the neural correlates of OCD has raised hopes of using MRI indices to discriminate OCD patients and the healthy.

volume of gray matter (VGM), cortical thickness and sulcal depth were extracted from T1-weighted images as the structural neuroimaging markers.

These MRI indices were used because previous studies have successfully revealed altered VGM [34–36] and cortical thickness [37–39] in various brain regions, including the traditional CSTC circuit and newly reported regions, such as the occipital, parietal, and temporal lobes and the cerebellum, in patients with OCD. sulcal depth may provide valuable information for classification since previous studies indicate the association of altered sulcus morphology and psychotic disease [40–42].

It is also important to explain the contribution of the MRI markers when constructing the OCD diagnosis models. However, due to the “black box” problem of machine learning models, such as SVM, previous studies seldom explored the contribution of the MRI markers used in the classification models. The Shapley value is a fair profit allocation among many stakeholders depending on their contribution and was derived from the name of the economist who introduced it. By using the idea of the Shapley value, approaches were proposed to interpret the predictions from any “black box” model [49, 50]. The key component of general explanations is the contributions (equivalent to the Shapley value) of individual input features. A prediction is explained by assigning to each feature a number which denote its influence. For each feature, such contributions can be aggregated to plot the feature’s average contribution against the feature’s value. This provides an overview of the model and explanation of the predictions.(Huang et al., 2023)

the ultimate conclusions drawn from this empirical work and thus recommendations to researchers and practitioners mirror those that Achenbach et al. (1987) provided over 25 years ago. In other words, recent empirical work echoes findings long known from prior meta-analytic reviews, begging the question, Have 25 years of empirical work on informant discrepancies improved how researchers and practitioners conduct multi-informant assessments? (Reyes, 2013)

Machine Learning

 XGBoost, renowned for its capability to model complex interactions within datasets, provides a robust framework for synthesizing high-dimensional neuroimaging data with behavioral and demographic variables. (GPT)

The present study

This thesis aims to address these gaps by utilizing the Adolescent Brain Cognitive Development (ABCD) dataset to investigate the intersections of parental versus child reports, structural MRI markers, and critical demographic variables in understanding OCD symptomatology. Employing the advanced machine learning model XGBoost, this study seeks to develop a robust predictive framework for OCD diagnosis that acknowledges informant discrepancies and incorporates neuroimaging insights. By combining sMRI data with behavioral reports, this research endeavors to present a more comprehensive picture of pediatric OCD, thus informing more nuanced clinical assessments and interventions.

This thesis seeks to address these gaps by leveraging the Adolescent Brain Cognitive Development (ABCD) dataset to explore the intersection of parental versus child reports, structural MRI markers, and key demographic variables in understanding OCD symptomatology. Employing a state-of-the-art machine learning model, XGBoost, this study aims to develop a robust predictive framework for OCD diagnosis that acknowledges informant discrepancies and incorporates neuroimaging insights. By doing so, this research endeavors to offer a more comprehensive picture of pediatric OCD, thereby informing more nuanced clinical assessments and interventions. (GPT) Advancements in neuroimaging have illuminated the potential of structural MRI (sMRI) to reveal underlying neurobiological correlates of OCD. Alterations in the cortico-striato-thalamo-cortical circuit and other associated brain regions point toward objective biomarkers that could complement traditional assessment methods. By combining sMRI data with behavioral reports, researchers can deepen their understanding of OCD's etiology and progression. (GPT)

Research question: To what extent does structural brain data explain the variation in anxiety symptoms as reported by youths versus their parents?

Model building focus on a phenomena that is already well established to see whether machine learning can detect the same features of informer discrepancy on ABCD dataset

Use MRI to predict youth score vs. parent score on four dimensions.

Have difference in score as an independent variable.

Hypothesis

There will be a significant difference in the prediction accuracy of structural brain data between self-reported and parent-reported anxiety symptoms in adolescents with GAD, with an expectation of higher accuracy for self-reported symptoms.

The purpose of this study was to examine agreement between youths and their parents regarding psychiatric problems, at a symptom level, using the CBCL. There are certain limitations of measuring parent–child agreement using the diagnostic approach, such as the loss of information (e.g. severity or magnitude of disagreement) when data are dichotomized. Another approach is to measure symptom agreement, which would make a quantitative distinction between parent’s and child’s ratings [26]. Thus, we will also examine agreement at a symptom level using the Achenbach System of Empirically Based Assessment (ASEBA) in an Icelandic outpatient clinical sample, as well as studying the influence of age, gender, attention-deficit/ hyperactivity, anxiety disorder and depres depressive disorder on parent–youth agreement. (Jónsdóttir et al., 2022)