Oxford Internet Institute, University of Oxford

Assignment Cover Sheet

Candidate Number	1037061
Assignment	Python for Social Data Science
Term	Trinity Term 2019
Title/Question	Examining the Effects of Diagnosis on OCD Sufferers' Reporting of Intrusive Thoughts in a Community Forum
Word Count	11802

By placing a tick in this box ⊠ I hereby certify as follows:

(a) This thesis or coursework is entirely my own work, except where acknowledgments of other sources are given. I also confirm that this coursework has not been submitted, wholly or substantially, to another examination at this or any other University or educational institution;

(b) I have read and understood the Education Committee's information and guidance on academic good practice and plagiarism at `https://www.ox.ac.uk/students/academic/guidance/skills?wssl=1`.

(c) I agree that my work may be checked for plagiarism using Turnitin software and have read the Notice to Candidates which can be seen at: `http://www.admin.ox.ac.uk/proctors/turnitin2w.shtml`, and that I agree to my work being screened and used as explained in that Notice;

(d) I have clearly indicated (with appropriate references) the presence of all material I have paraphrased, quoted or used from other sources, including any diagrams, charts, tables or graphs.

(e) I have acknowledged appropriately any assistance I have received in addition to that provided by my [tutor/supervisor/adviser].

(f) I have not sought assistance from a professional agency;

(g) I understand that any false claims for this work will be reported to the Proctors and may be penalized in accordance with the University regulations.

Please remember:

- To attach a second relevant cover sheet if you have a disability such as dyslexia or dyspraxia. These are available from the Higher Degrees Office, but the Disability Advisory Service will be able to guide you.

University of Oxford

Examining the Effects of Diagnosis on OCD Sufferers' Reporting of Intrusive Thoughts in a Community Forum

1037061, Pembroke College

Thesis submitted in partial fulfilment of the requirement for the degree of MSc in Social Data Science at the Oxford Internet Institute at the University of Oxford

Trinity Term 2019

11802 words

Abstract

Intrusive thoughts are a common symptom of obsessive-compulsive disorder (OCD), a mental health condition which affects an estimated 2% of the global population. This work explores the content of individuals' intrusive thoughts as they describe them on an online community forum dedicated to OCD, and in particular how an individual's having been formally diagnosed with the condition alters their reporting of intrusive thought symptoms as compared to other individuals who report never having been formally diagnosed. From structural and semantic differences in symptom reporting, I attempt to build statistical classifiers to distinguish individuals according to their diagnostic status. The results paint a picture of reporting of intrusive thought symptoms in an online context and suggest that significant differences exist in individuals' reporting behavior and symptom conceptualization in the course of being diagnosed with and treated for OCD.

Keywords: obsessive-compulsive disorder, OCD, mental health, self-reporting, mental health diagnostics, online forums, topic modelling, text classification

1 Introduction

Internet platforms which facilitate public discourse on stigmatic health concerns are an increasingly popular medium for individuals seeking information and social support surrounding mental health (De Choudhury & De, 2014). Recent research has delved into how a variety of mental health conditions are manifested in discourses on such platforms, including major depressive disorder (De Choudhury, Gamon, Counts, & Horvitz, 2013), post-traumatic stress disorder (Reece et al., 2017), and others (Coppersmith, Dredze, & Harman, 2014). These studies have focused on generalized signals corresponding with mental health conditions rather than their specific symptomologies. Moreover, they have used data collected only from large social media platforms such as Twitter and Reddit. Coupled with this selective orientation, their results suggest that more research is warranted on less-studied mental health conditions, using data from other online platforms which facilitate discourse on mental health, and on comparatively less-studied mental health conditions.

This study examines how a specific mental health symptom known as "intrusive thoughts" is reported on an online forum where individuals discuss their experiences of living with obsessive-compulsive disorder (OCD), a mental health condition which affects an estimated 2% of the global population (Simpson, 2017; Rachman & de Silva, 2009). In addition to describing the ways in which individuals report their

intrusive thoughts online, the study also tests the hypothesis that an individual's diagnostic status—specifically, whether or not they report having been diagnosed with OCD—has a significant impact on their reporting of intrusive thoughts, and by extension, their conceptualization of those thoughts. Using analytical techniques to examine the textual structure and meaning of diagnosed and undiagnosed posts, my findings show that significant differences do exist in the reporting patterns of the two groups. The substance of these differences suggests corresponding changes in the ways OCD sufferers conceive of their thoughts in the course of being diagnosed and treated for the condition, and in ways commensurate with the goals of diagnosis and treatment described in the OCD literature. Transitively, I show that signals indicative of these changes exist in data drawn from an understudied type of platform for mental health online.

I believe the contribution of this study is twofold. First, its results paint a picture of how individuals' reporting of intrusive thoughts changes in the course of diagnosis and treatment for OCD. These changes indicate that interaction with mental health practitioners and treatment courses lead patients to re-conceptualize ego-dystonic thoughts such that they feel "de-essentialized" from them: The bad thoughts come to be understood as things that happen to them as if by accident, as opposed to things which they cause, or otherwise happen due to some essential quality of their identity. Undiagnosed individuals, on the other hand, describe their intrusive thoughts in a more subjective, "essentialized" manner, as though the thoughts reflect things they do or are. They also frequently report their thoughts in ways intended to elicit diagnosis from their audience. While these findings are of themselves unsurprising in relation to the body of research on diagnosis and treatment for OCD, they have not been observed (to the best of my knowledge) in the context of discourses in online social data. Second, the study's contribution is to demonstrate that signals evocative of these findings are detectable on a relatively small community forum dedicated to a specific symptom of a defined mental health condition. This suggests that similar platforms could provide fruitful sources of data for future studies on mental health.

The study proceeds in five parts. The first section provides background information about intrusive thoughts, obsessions and compulsions, as well as diagnostic criteria for OCD. It also discusses related research on expressions of mental health online using social data. The second section describes the use of online forums for collecting data on discourses on stigmatic health concerns, and the particular forum used as a data source for this research. It also describes my own specific methodology used for data collection and labelling. In the third section I present

my methods for analyzing reporting of intrusive thoughts on both structural and semantic levels, illustrating how significant differences were identified in the reporting of diagnosed and undiagnosed individuals. The fourth section builds on these differences, presenting statistical classifiers aimed at identifying forum posts based on their authors' diagnostic statuses. In the fifth section, I discuss these findings in the broader context of research on mental health online before concluding with a critical review of the limitations of this study, as well as inducements for future research.

2 Background

2.1 What are Intrusive Thoughts?

Intrusive thoughts are unwanted thoughts or images that cause marked anxiety and distress (International OCD Foundation, 2014). An intrusive thought can be about harming a loved one, performing a forbidden sexual act, or holding a blasphemous religious belief. In fact, intrusive thoughts may be about any subject that a person may find distressing (Osborn, 1998). Baer (2001) catalogued 91 common intrusive thoughts in individuals in the United States, including fears that they may be or may become a pedophile; worries about pushing a commuter in front of an oncoming train; fears that they might secretly be gay (or straight); worrying about shouting obscenities in public; fearing that they may run over a jogger while driving; and worries about imagining blasphemous sexual images about Jesus or the Virgin Mary (p. 125-128). What all intrusive thoughts have in common is that they are unwanted, and that they cause at least some anxiety or distress. In other words, for a thought to be intrusive, it needs to be ego-dystonic—meaning that it is inconsistent with or opposed to the needs and goals of the person thinking it (Veale, 2004).

Almost all humans admit to having such thoughts (Rachman & de Silva, 1978; Clark & Radomsky, 2014). Rachman & de Silva (1978) found that 85% of otherwise healthy British university students had intrusive thoughts about harm, sexuality, contamination, religion, or some combination thereof. Some of the students admitted to having ten or more such thoughts per week (Rachman & de Silva, 1978, p. 235). However, the vast majority of these subjects also reported that they found these intrusive thoughts easy to dismiss. This and subsequent studies have found that almost all people are occasionally visited by unwanted, disturbing, intrusive thoughts, but are not deeply affected by them (cf. Clark & Radomsky, 2014).

2.2 How Intrusive Thoughts Relate to OCD

Some people are debilitated by their intrusive thoughts. In the majority of such cases, mental health experts agree that the appropriate diagnosis is OCD (International OCD Foundation, 2014). Baer (2001) wrote:

> Bad thoughts—when they are severe they are called obsessions—may cost people the most important things in their lives: Some cannot bear to be around their own children; others cannot have relationships; and others are so paralyzed they cannot perform simple everyday activities ... Many contemplate suicide at some time. These are obsessions of clinical severity and require treatment. (p. 5)

The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) defines obsessive-compulsive disorder (OCD) in terms of obsessions and compulsions. As Baer (2001) noted, obsessions are closely related to intrusive thoughts, as they are characterized by, "... recurrent and persistent thoughts, urges, or impulses that are experienced, at some time during the disturbance, as intrusive and unwanted, and that in most individuals cause marked anxiety or distress" (American Psychological Association, 2013). The sufferer attempts to suppress or neutralize such thoughts or urges with some other thought or action, also known as a compulsion. Compulsions may include repetitive physical behaviors such as hand washing or checking as well as mental acts such as praying or self-reassuring. However, these mental and physical behaviors are not effective in neutralizing or preventing their associated obsessions, or are otherwise "clearly excessive" (APA, 2013).

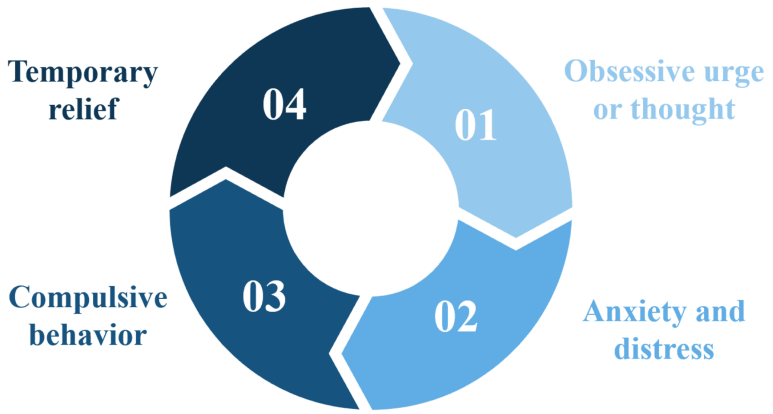

Figure 1: The OCD Cycle

Some experts explain OCD conceptually in terms of the "OCD cycle," depicted in Figure 1, which aims to portray in simple visual form the cognitive-behavioral

model of the condition. Individuals experience an obsessive urge or intrusive thought which causes them significant anxiety and distress, such as "What if I hurt a loved one?" They perform a compulsive behavior, either mental or physical, in an attempt to neutralize the thought, such as "I'll avoid my family and make sure I don't get too close to them" or "I'll make sure there are no knives nearby when I'm with them." The compulsive behavior tends to provide temporary relief from the anxiety. However, for OCD sufferers, enacting such compulsions actually compounds their obsessions and anxiety over time (Rachman & de Silva, 2009). They are in fact conditioning themselves and their brains to believe that their compulsive behavior is responsible for preventing the dreaded outcome. In this way the compulsive behaviors empower the obsession and cause more intrusive thoughts, creating a self-reinforcing cycle.

OCD is often associated with overt physical compulsions, such as hand-washing or rearranging things. For nearly as long as the condition has been recognized by the psychiatric community, the disorder was characterized exclusively in terms of such physical manifestations (Williams et al., 2011). However, a sea-change in OCD occurred in the second half of the 20th century, when practitioners began to conceive of intrusive thoughts and repetitive mental behaviors as falling under the OCD umbrella (Phillipson, 1989). The class of sufferers affected mainly by intrusive thoughts and mental compulsions was later codified by Phillipson (1989) as "Pure-Obsessional" OCD sufferers:

> "The 'Pure-O' is manifested by a two-part process: the originating unwanted thought (spike) and the mental activity which attempts to escape, solve, or undo the spike ... It is not unusual for the 'Pure-O' sufferer to spend eight hours a day in rumination, trying to find a way to escape. (p. 1)

To note, some OCD experts have criticized the concept of Pure-O as unhelpful for underestimating the importance of compulsive mental rituals in these cases (cf. Williams et al., 2011). In other words, it is not that sufferers merely have "pure" obsessions. The performance of an accompanying mental activity to alleviate or "escape" from the anxiety constitutes a compulsion in the same way that handwashing or rearranging do. Nevertheless, the concept of Pure-O has been helpful for many individuals in understanding the link between intrusive thoughts and OCD.

2.3 Diagnosing OCD

Diagnosis of the condition is often made as a function of scores on screening tests, most notably the Yale-Brown Obsessive Compulsive Scale (Goodman, Price, & Rasmussen, 1989). The "Y-BOCS" is a 10-item clinician-administered test providing five rating dimensions of obsessions and compulsions, including time spent or occupied; interference with functioning or relationships; degree of distress; resistance; and control (i.e., success in resistance). However, although the Y-BOCS is sometimes used as a sole diagnostic criterion, it is important to note that its creators intended for the scale to rate symptom severity and not to establish diagnosis (Goodman, Price, & Rasmussen, 1989, p. 1008).

Work delving into the causes, pathologies, and epidemiology of OCD has fundamentally changed the way the disorder is treated both pharmacologically and therapeutically. For example, the development of Exposure and Response Prevention (ERP)—the widely accepted best treatment for OCD—has grown out of increased understanding of the cognitive behavioral dynamics of the condition (Lack, 2013; Abramowitz, Blakey, Reuman, & Bucholz, 2018). In fact, ERP is a form of cognitive behavioral therapy which effectively seeks to break the OCD cycle of intrusive thoughts and compulsive responses by forcing patients to do exactly the opposite of what they believe will help them. Patients are gradually exposed to the subject matter of their obsessions and intrusive thoughts, through stories, images, and a variety of other means. The patient is then discouraged or prevented from performing any of their compulsive behaviors in response to the anxiety (hence, "response prevention"). Gradually, patients break the conditioned cycle of obsessing and compulsing, and the vast majority experience far fewer intrusive thoughts (cf. Hezel & Simpson, 2019).

The experience of being diagnosed with a mental health condition like OCD often affects both an individual's conceptualization of their own condition as well as their self-identification with respect to society at large. In the first case, a diagnosis can shift an individual's view from believing their feelings to be the result of some sort of "moral failing" to the view that their experiences are attributable to some external cause (Shafran, Watkins, & Charman, 1996). Simply the understanding that one's own suffering "has a name," can be deeply relieving to some people with OCD. Also, in some cases, research has found that a patient's shifting beliefs towards a biological conception of mental illness—the view they have a "brain disease" as opposed to some inherent "weakness of character"—can reduce feelings of alienation and stigmatization (Andersson & Harkness, 2018). This effect may be particularly pronounced in OCD, as individuals with intrusive thoughts of-

ten believe those thoughts are the result of some profound moral failing (Shafran, Watkins, & Charman, 1996). After all, people generally believe they have control over their thoughts, and if those thoughts center on taboo sexual, religious, or violent topics, then it would be easy to believe they have a bearing on the thinker's moral character (Beadel, Green, Hosseinbor, & Teachman, 2013). However, in the context of diagnosing OCD, many clinicians seek to drive home in their patients the understanding that their thoughts don't represent their true interests or intentions. Put simply, "you are not your thoughts" (IOCDF, 2017). Therefore we may expect that individuals who have been diagnosed may conceive of their thoughts as less essential to themselves than undiagnosed individuals.

Being diagnosed with OCD or other mental health conditions can change how individuals view themselves as members of society in ways that go beyond just the content or nature of their intrusive thoughts. Goffman (1963) argued that individuals with mental health conditions are often led to believe they have some "essential" attribute that distinguishes them from the rest of society. Such individuals seek to manage impressions of themselves to reconcile with their perceived shortcomings, oftentimes by concealing aspects of themselves that they believe depart from acceptable standards. In this view, while diagnosis of a mental health condition may provide an attribution for subjects' personal difficulties, it may also serve as a label setting them apart from other people. The meaning of these labels, and their ability to separate people in a community or society can change significantly over time. Foucault's (1963) influential study showed how the categories of madness and insanity had changed in Modern history. And indeed, empirical evidence suggests that the level of "otherness" associated with mental health conditions has diminished in places like the United States and Great Britain since the second half of the twentieth century (Phelan, Link, Stueve, & Pescosolido, 2000). Still, the lasting effects of stigmatization can drive sufferers of mental health conditions to form communities over their conditions, and the Internet provides a uniquely accessible and anonymous place for such communities to flourish.

2.4 Mental Health and Social Data

Is it possible to identify individuals suffering from mental health conditions based on their online, social data? In recent years there has been growing interest in this question and, more broadly, the ways in which mental health conditions are manifested in online contexts, yet insight into these topics with respect to OCD is limited at best. Anecdotally, medical health practitioners in the United States and Great Britain have been known to suggest that upwards of half of all cases of OCD

have a significant online component: "People with all sorts of OCD problems use the Internet as a means of compulsive checking or reassurance seeking about their obsessional fears and worries," remarked one expert (Tait, 2016). Some ways that these behaviors may manifest include "compulsive Googling", where individuals will spend hours each day searching for stories and articles related to their obsessions in an attempt to assuage their fears (OCD Life, 2017). And indeed another form of this phenomenon could be individuals frequenting online community forums seeking reassurance from other visitors. Yet I have been unable to identify any systematic study of the rates at which this compulsive Internet usage occurs.

Considerably more attention has been devoted to other mental health conditions and in particular identifying individuals suffering from them using social data. De Choudhury, Gamon, Counts, & Horvitz (2013) was among the first promising works in this regard. The authors of that study identified ways in which natural language processing could be used on social media data of various kinds to examine mental health dynamics in individuals and populations. The study applied these methods to the Twitter data for 476 users previously diagnosed with clinical depression and identified significant trends in content and periodicity of those users' Tweets in the lead-up to their reported date of diagnosis. The study also presented a support vector machine classifier that achieved around 72% accuracy in predicting whether a Tweet had occurred before or after the diagnosis date. Coppersmith, Dredze, & Harman (2014) employed similar statistical classifiers on Twitter data for a larger selection of mental health conditions, where subjects were assigned to various groups from their self-stated reports of diagnosis on Twitter. Their classification approaches made use of a variety of structural and platform-related factors, in addition to the content of users' tweets, and in many cases were able to successfully differentiate undiagnosed users from users diagnosed with mental health conditions. Most recently, Reece et al. (2017) collected data from 204 Twitter users recruited through Amazon's Mturk platform, creating a dataset containing 279,951 tweets. Participants also provided information on their first clinical diagnosis of either depression or PTSD. They achieved 88.2% accuracy using a random forest classifier to identify individuals diagnosed with PTSD or depression based on the Tweets.

To note, the goals of those studies were not to formally diagnose individuals with a mental health condition, or even to develop automated tools to make such diagnoses. Rather, the stated objectives revolve around advancing our collective understanding of the conditions in question, as well as understanding their manifestation(s) online. To claim that automated tools may diagnose people with mental health conditions is both scientifically misleading and also potentially morally haz-

ardous: Individuals with mental health concerns may rely on online tools claiming to "diagnose" them with specific mental health conditions, often making them less likely to seek out accredited mental health diagnosis and treatment (Pillay, 2010; Lupton & Jutel, 2015).

The methods of this report share some aspects of those used in the studies mentioned above, with an important difference regarding its guiding research question. Rather than seeking to identify individuals with verified mental health diagnoses based on their social media data, this study instead probed: "What differences, if any, exist in posts about intrusive thoughts written by individuals who chose to self-report their diagnostic status on an online forum?" In addition, as I will describe further in the next section, my data was collected from a platform other than Twitter or Reddit, and without the use of crowd-working platforms like MTurk. Accordingly, I identified and categorized users based on their self-reported diagnostic status in much the same way as Coppersmith, Dredze, & Harman (2014). This means that my data has no formal control group separating individuals who have been diagnosed with OCD from those who have not. This aspect of my study design presents opportunities and limitations, which I will bring up in the course of my analyses as well as my discussion section.

3 Methods

3.1 Data from a Community Forum

Individuals affected by mental health problems often seek out online communities devoted to those conditions and the experiences of living with them (Henderson, Evans-Lacko, & Thornicroft, 2013; De Choudhury & De, 2014). For one, the symptoms of common mental health problems such as OCD can leave individuals feeling isolated, introspective, and desperate to share their thoughts and feelings with others. Simultaneously, these same individuals may be profoundly ashamed of these feelings due to a lack of understanding of the features of mental illnesses and a keen sensitivity to the stigma surrounding them. Moreover, many individuals with mental health concerns lack access to mental healthcare outright. Henderson, Evans-Lacko, & Thornicroft (2013) estimated that 70% of people with mental illness worldwide receive no treatment from health care professionals.

Online community forums facilitate communication and networking for communities on the Internet. Discourses on such forums often present ideological expressions of their members, making it possible to identify the community's dominant sentiments or interpretations about specific topics (La Violette & Hogan, 2019).

Joining an online community, and in particular an online community forum, can be an attractive option for individuals suffering with depression, OCD, and other mental health conditions. Community forums for mental health typically offer users the ability to share their thoughts and experiences with relative anonymity and to a largely sympathetic and understanding audience. They are also places where individuals can learn more about specific condition(s), symptoms, and medications, and how they might seek other forms of support offline (De Choudhury & De, 2014). In some cases, individuals may use online mental health forums to elicit diagnoses of mental health conditions, either in addition to or in lieu of actual diagnoses from professionals. Giles and Newbold (2011) observed that many people seeking mental health diagnoses online did so out of fear, mistrust, or lack of access to mental healthcare offline. While the extent of online diagnosis-seeking remains unclear, the rapid growth of "self-diagnosis" mobile apps suggests that an increasing number of individuals are turning to the Internet for health diagnostics (Lupton & Jutel, 2015).

Many people with intrusive thoughts about harm, sexuality, or religion are reluctant to share those thoughts with anyone for fear of being labelled as a psychopath, pedophile, heretic and so forth. In reality, as we have seen, millions of other people struggle with similar thoughts. Baer (2001) wrote:

> Nearly everyone who comes to me for help with bad thoughts thinks he or she is the only person with these thoughts. Yet, if everyone in the United States who suffers from these bad thoughts congregated, they would form the fourth-largest city in the United States. (p. 17)

On a community forum, such individuals can relate their thoughts anonymously and from the comfort of their own home. Some individuals may use these forums to easily seek reassurance or "diagnosis" from a large group of sympathetic listeners, a common compulsive behavior observed in OCD.

I determined that an online community forum such as this would provide an effective platform for collecting data pertaining to individuals' experiences with OCD, intrusive thoughts, and diagnosis of the condition. Data for this research were therefore drawn from an online community forum centered on OCD and administered in English. In cooperation with the forum's maintainers, I used a web scraping tool built with the Requests library in Python to collect posts from the forum. The data are publicly available online and viewable by anyone, however an account is needed to make a posting on the forum. In the context of this work it was agreed with the forum's maintainers that the specific details of the supervising organization and the forum should remain anonymous so as to preserve the forum's status as a safe space for individuals to discuss their experiences of intrusive thoughts. Posts were

collected from the forum's inception through the beginning of June 2019. The data associated with each post included its URL path in the forum, the post title, the post author's username, the time of posting in Coordinated Universal Time, and the raw contents of the post.

3.2 Annotation

In data science, annotation refers to the task of labelling data such that it can be used for comparative, "supervised" tasks such as statistical classification (Stubbs & Pustejovsky, 2012). Annotation was used in the context of this research for identifying which posts were authored by individuals with a formal diagnosis of OCD and which were authored by individuals who reported having never been diagnosed with the condition. While it may have been possible to use an automated tool to label some relevant posts successfully, a large number of other posts would inevitably be mislabelled. For example, searching for specific phrases like "haven't been diagnosed" versus "was diagnosed" in the raw texts would inevitably capture individuals describing diagnoses of other conditions besides OCD. Moreover, in some posts the authors described a conflicting diagnosis coming from multiple mental health professionals. Simply put, an automated tool would not be reliable enough for labelling the data. I determined therefore that manual annotation was the best means of delineating posts according to the diagnostic status of the author. For this task, the entire forum dataset was filtered for posts making mention of frequent terms related to diagnosis, including "diagnosis", "diagnosed", "psychiatrist", "psychologist", "medication", and common misspellings of those terms. Important to note is that the word "OCD" was not included in the filtering, as a multitude of posts made mention of the condition without any reference to its diagnosis. Filtering in this manner resulted in a sample of approximately 1,150 posts authored by 782 unique authors.

I then annotated the filtered sample by reading the posts individually and categorizing the authors as "Diagnosed", "Undiagnosed", or in some cases "Unsure." For a post to be labeled as "Diagnosed" there had to be explicit reference to diagnosis of OCD by a mental health professional—either a psychiatrist, psychologist, or therapist. Individuals who claimed to be self-diagnosed were placed in the "Undiagnosed" group. This included individuals claiming to have self-administered diagnostic tests for OCD such as the Y-BOCS and scored "highly enough" to be diagnosed. Cases where the author mentioned conflicting diagnosis (i.e., one professional had diagnosed them whereas another disagreed with that diagnosis) were labeled as "Unsure." Although strictly speaking such individuals may have been

diagnosed with OCD at one point, I determined that the existence of a conflicting diagnosis should mean that their diagnostic status was not clear enough to be placed in the "Diagnosed" category.

Ultimately, the complete labelled dataset yielded roughly 52.4% of posts belonging to the "Diagnosed" class, 42.4% belonging to the "Undiagnosed" class, and 4.3% belonging to the "Unsure" class, with the following group frequencies:

Table 1: Class Counts

Diagnosed	578
Undiagnosed	468
Unsure	58

3.3 Structural Differences Between Groups

3.3.1 Post Lengths

Exploratory analysis of the two corpora commenced with aggregation and comparison of post lengths. Although text length is often measured by word count, I determined character counts were more appropriate for length measurement with the scraped texts. This is primarily because any attempt to split the raw texts into individual words could result in missed or extra words depending on stylistic differences among a sample of many different authors. On the other hand, counting characters would provide a more consistent, if less granular picture of how the posts were distributed in terms of length. The raw texts were therefore processed as strings and measured by Python's built-in length function.

There was a notable difference in post length between the groups as measured in total post characters, depicted in Figure 2. Both distributions were unimodal and skewed right, indicating that the majority of posts fell into the range of zero to 5,000 characters. However, the tail of the undiagnosed authors' distribution extended further than the diagnosed group, indicating that this group had produced more very long posts. The result was an average length of 3,043 characters for undiagnosed posters compared to 2,058 among diagnosed posters. Several extremely long posts among undiagnosed users also contributed to the group's higher standard deviation as compared to the diagnosed group, 3,338 and 1,810, respectively. The larger mean and standard deviation observed in the undiagnosed group is largely attributable to a number of particularly long posts authored by that group of posters. For example,

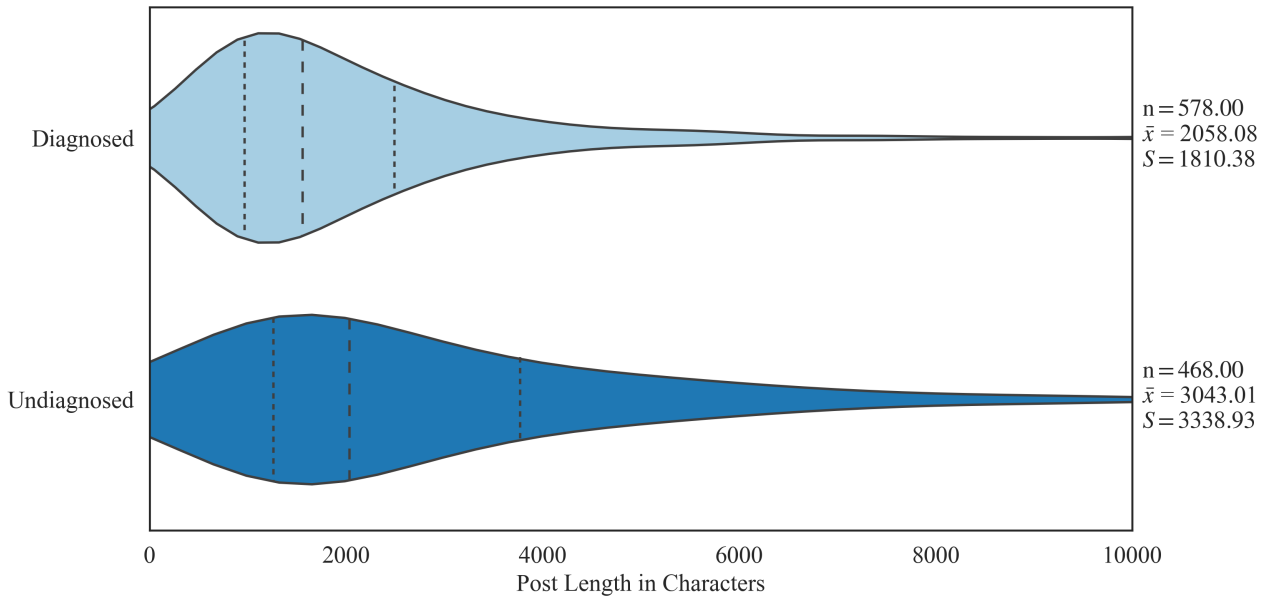

Figure 2: Post Length Distributions

whereas just one post by diagnosed author was longer than 15,000 characters, nine posts authored by undiagnosed authors exceeded that length.

All other things being equal, the difference in post lengths between the groups suggests that the undiagnosed posters had comparatively more information that they wanted to convey in their texts on the forum. An alternative interpretation, albeit contingent on additional information, might suggest that groups sought to express similar amounts of information, but that one of the classes of posters did so with fewer characters—perhaps due to stylistic or linguistic differences in their writing styles. However, the fact that undiagnosed posters authored more of the "very long" posts suggests that the differences in post length had more to do with the amount of information that authors' wished to transmit, as opposed to this alternative explanation regarding linguistic differences.

3.3.2 Parts of Speech

A second exploratory task aimed at interrogating the structures of the raw forum texts was examining differences in the usage of various parts of speech in the two groups of posts. In linguistics, parts of speech are categories of words assigned in accordance with those words' syntactic functions, such as nouns, verbs, pronouns, and so forth (Nugues, 2006, ch. 5). Past research aimed at identifying significant differences in the usage of parts of speech across texts suggests that these differences

can lend insight into the governing structural and semantic dynamics of those texts (Rayson & Garside, 2000). I therefore sought to compare the frequency of various parts of speech used in the two classes of posts.

The first step in this task was actually accumulating counts of the parts of speech used in each forum post, and then aggregating those counts across the two groups. Here again, the textual data used was simply the tokenized (i.e., split word by word) raw texts scraped from the forum, having undergone no forms of preprocessing. Preprocessing would inevitably remove or alter words in the texts, and in this case I was interested in comparing all words as they appeared in the original writing.

Tagging the words contained in every post according to their associated part of speech was accomplished using the NLTK package in Python and in particular the module's Averaged Perceptron Tagger (Honnibal, 2013). The Averaged Perceptron Tagger is a part of speech tagger in NLTK, representing a neural network that has been pre-trained on a large corpus of written English to identify parts of speech in the context of raw texts. The tagger can be applied to new texts to categorize tokens into one of 45 possible parts of speech. Once the Tagger was applied to each post, I summed the counts for each part of speech category across the diagnosed and undiagnosed groups of text, respectively. Overall this process yielded parts of speech counts for each class, which I then normalized simply as the proportion of tokens represented by a given part of speech relative to the total number of tokens in the class. In other words, these proportions represented what percentage of all the words used in each group represented each part of speech. Next, as a simple means of identifying significant differences in the parts of speech usage between the groups, I performed a Two Proportion Z-Test for each part of speech and its relative frequency across diagnosed and undiagnosed classes. Because significance tests were performed individually for each of the 45 parts of speech, there was a high chance that one or more of the tests would produce a type I error on the significance level of 0.05, or even 0.01. I therefore tested for significant differences using an α of 0.001, often referred to as the threshold for a "highly significant" result.

Figure 3 displays the distributions of non-punctuational parts of speech representing at least 5,000 tokens (roughly 2%) of the total tokens for each class. The x-axis represents the percentage of all tokens belonging to a group (i.e., undiagnosed or diagnosed) that were tagged as each of those parts of speech. For example, "Prepositions or Conjunctions" represented around 2% of all words used by both diagnosed and undiagnosed posters. Bars in red represent parts of speech where the associated test statistic exceeded the threshold for significance on $\alpha = 0.001$, where the side the red bar is on represents the group that used that part of speech more.

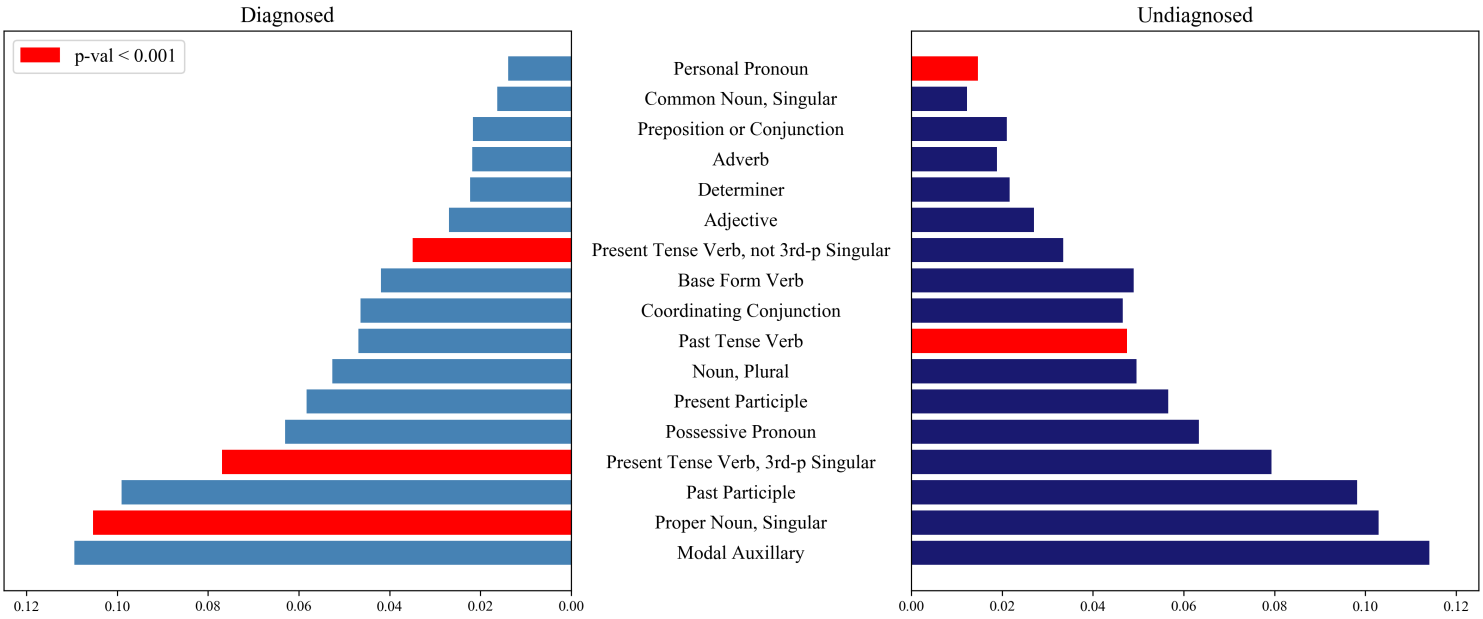

Figure 3: Percent of all Tokens Belonging to Various Parts of Speech

For example, undiagnosed authors used significantly more personal pronouns than diagnosed authors.

From these results, we observe that undiagnosed posters used relatively more past tense verbs and personal pronouns in their writing, whereas diagnosed posters used relatively more present tense verbs. Ostensibly, this would indicate that the focus of the undiagnosed posts skewed more towards events in the past. Moreover, the higher relative frequency of personal pronouns such as "I", "He", and "She", suggests that the undiagnosed posters referred more to these subjects—perhaps themselves or individuals close to them. The presence of more past tense verbs and personal and self-referencing pronouns is often associated with a storytelling and narrative writing style (Alzahrani, 2018). Therefore, the larger relative frequencies of these parts of speech in the undiagnosed texts point to a higher degree of narration among undiagnosed posters, as these words are typically used after a particular figure is introduced in the context of a story being told about the past (e.g., "Martha came to my house. She knocked on the door."). On the other hand, we might infer that the relative importance of present tense verbs among diagnosed posters indicates a greater importance assigned to the present.

Combining these grammatical differences with the distribution of post lengths, a picture begins to emerge on the structural differences between the posts of authors

indicating that they had been diagnosed with OCD and their counterparts reporting having never been diagnosed. The undiagnosed posts tended to be longer, and the information they sought to express tended to be oriented towards past events and personal entities, likely conveyed in the form of a more narrative, story-telling style. Following Holmes (2000), this style is often associated with patient dialogue in psychiatric clinical practice: "In [this] context, a story is a sequence of events centring on a suffering patient. It consists of alternating episodes of 'what was done to me/what I felt', and 'how I reacted/responded/felt about it in return'." We may therefore infer that a large number of the undiagnosed posts adapted this style in their online writing, telling stories in the course of asking "What's wrong with me?" or otherwise trying to elicit a diagnosis from their readers.

3.4 Semantic Differences Between Groups

3.4.1 Preprocessing

Semantics is the linguistic and philosophical study of meaning in language. Having observed structural differences in the lengths and grammatical dynamics of the two classes of text, I next turned to semantical analysis of the two groups. By examining the possibility of semantic differences among diagnosed and undiagnosed class posts, I sought to better understand analogous differences in meaning between the groups. In other words, what things are the two types of posters talking about and how?

A common strategy applied in semantic analyses of corpora is preprocessing of the textual data therein. While the components of this preprocessing may vary depending on the kinds of texts under analysis, common strategies include stemming of texts, lemmatization of texts, and removal of "stopwords" with little semantic significance. Several of these strategies were applied to the raw post data with the Python modules NLTK and Gensim. First, the raw texts were tokenized, meaning they were broken into individual components, usually words. The resulting lists contain the individual linguistic components of the original texts, which are then filtered for the removal of stopwords. In natural language processing, stop words are commonly used words deemed to have little semantic relevance, such as articles, pronouns, and prepositions. In this case, the particular set of stop words removed were those found in the default Gensim stop words set (Rehůřek, 2009). The rationale for removing stop words like "a", "the", and "in", is that while such terms matter in analyzing the grammatical structure of texts, as in the case of parts of speech analysis, they tend to have little bearing on texts' meaning, making it more difficult to spot semantic signals therein (Munková, Munk, & Vozár, 2013). Finally,

having tokenized the texts and removed stop words, the tokens were lemmatized, a process whereby each token is reduced to its lemma—its non-inflectional form. The idea here is to reduce a word to its dictionary form such that it can be analyzed consistently. So for example, the tokens "studies" or "studying" are reduced to their non-inflectional form, "study." Lemmatization for the forum texts was accomplished with NLTK's WordNet Lemmatizer function. WordNet is a large lexical database of English words containing comprehensive information on words, their inflectional forms, and their lemma (Fellbaum, 2005).

While some information contained in the texts is inevitably lost as a result of these preprocessing approaches, they function to make the texts more conducive to semantic analysis. By tokenizing, removing stop words, and lemmatizing the texts it becomes more feasible to identify the central meanings they contain.

3.4.2 Word Frequencies

A simple approach to recognizing semantic differences across corpora is to identify impactful words with significantly different usages across groups. Differences in word frequency can offer insight into different topics being discussed, as well as the relative importance of certain topics or concepts in one corpus versus another (Säily & Suomela, 2017). There is an ongoing debate in the literature regarding the appropriate statistical tests for significance when comparing word frequencies. Kilgarrif (1996) and others have pointed out that parametric statistical tests—which typically rely on the assumption that data is normally distributed—are inappropriate for written language, because words used in normal language are almost never normally distributed. Therefore, using parametric tests for words frequencies across corpora often leads to overestimated significance (Kilgarrif 2007; Bestgen 2014). In order to avoid overestimating the significance of differences in the relative word frequencies I chose to apply a non-parametric Chi-square test of independence of variables, and also to set the level of significance at the level of $\alpha = 0.001$.

I produced a two-way contingency table for each word across both corpora where the word appeared at least 15 times in both groups. The contingency tables contained the number of times a specific word appeared in the diagnosed texts as well as the undiagnosed texts, and also the sum of all tokens used by each group. Table 2 depicts one of the two-way contingency tables used in the Chi-square test.

Applying the Chi-square contingency test to all of the tokens, I could then examine significant differences in their relative frequencies across corpora. Figure 4 depicts 30 words found to be significant on $\alpha = 0.001$. These 30 words had the largest test statistic magnitude (i.e., the smallest p-values)—15 belonging to the

Table 2: Two-way Contingency for Token "OCD"

	Diagnosed	Undiagnosed
Count	1701	1059
Total	79186	93293

undiagnosed posts and 15 belonging to the diagnosed posts. The bar widths represent the percent difference in relative frequency for each word. In other words, a bar with width of one represents a token used twice as frequently as the other.

There were significant differences among relative frequencies, with some more expected than others. Perhaps not surprisingly, the term "OCD" occurs twice as frequently in the diagnosed group than the undiagnosed group. The individuals in this group appear to be more likely to use the term itself in describing their condition. Interesting to note, though, is that the undiagnosed group was significantly more likely to use other acronyms to describe their condition, such as "HOCD" and "POCD" (not seen in the figure). These acronyms refer to OCD characterized by specific intrusive thought content, namely "Homosexual OCD" and "Pedophilia OCD." In other words, someone may describe themselves as suffering from "HOCD" if their intrusive thoughts center around thoughts like "How can I know if I'm gay or I'm straight?" and so forth (cf. Penzel, 2007). Important to note is that these terms are generally not used by medical health professionals in much the same way that "purely obsessional OCD" is rarely used (and often criticized as an unhelpful label). Nevertheless these terms appear in self-help articles for OCD online and in many colloquial references to intrusive thoughts. Therefore it makes sense that the undiagnosed group should use these acronyms more frequently when characterizing what they believe to be their condition. On the other hand, the diagnosed posters were significantly more likely to describe their condition simply as "OCD" and with a specific "theme," such as "harm", and "paedophile." A clinician or OCD specialist is much more likely to describe the condition in these terms like "theme" and "urge" as opposed to terms like "HOCD" and "POCD." Unsurprisingly, the diagnosed group also exhibits greater usage of medical and therapeutic terms, including "CBT," which stands for cognitive-behavioral therapy, "medication," and "psychiatrist," than the undiagnosed group.

One unexpected result in the frequency differences was the presence of more "youthful" terms as well as sexual terms in the undiagnosed group. "Boyfriend", "friend", and "sister" appear as significantly more frequent among undiagnosed posters, and so do terms related to sexuality and orientation like "gay", "porn",

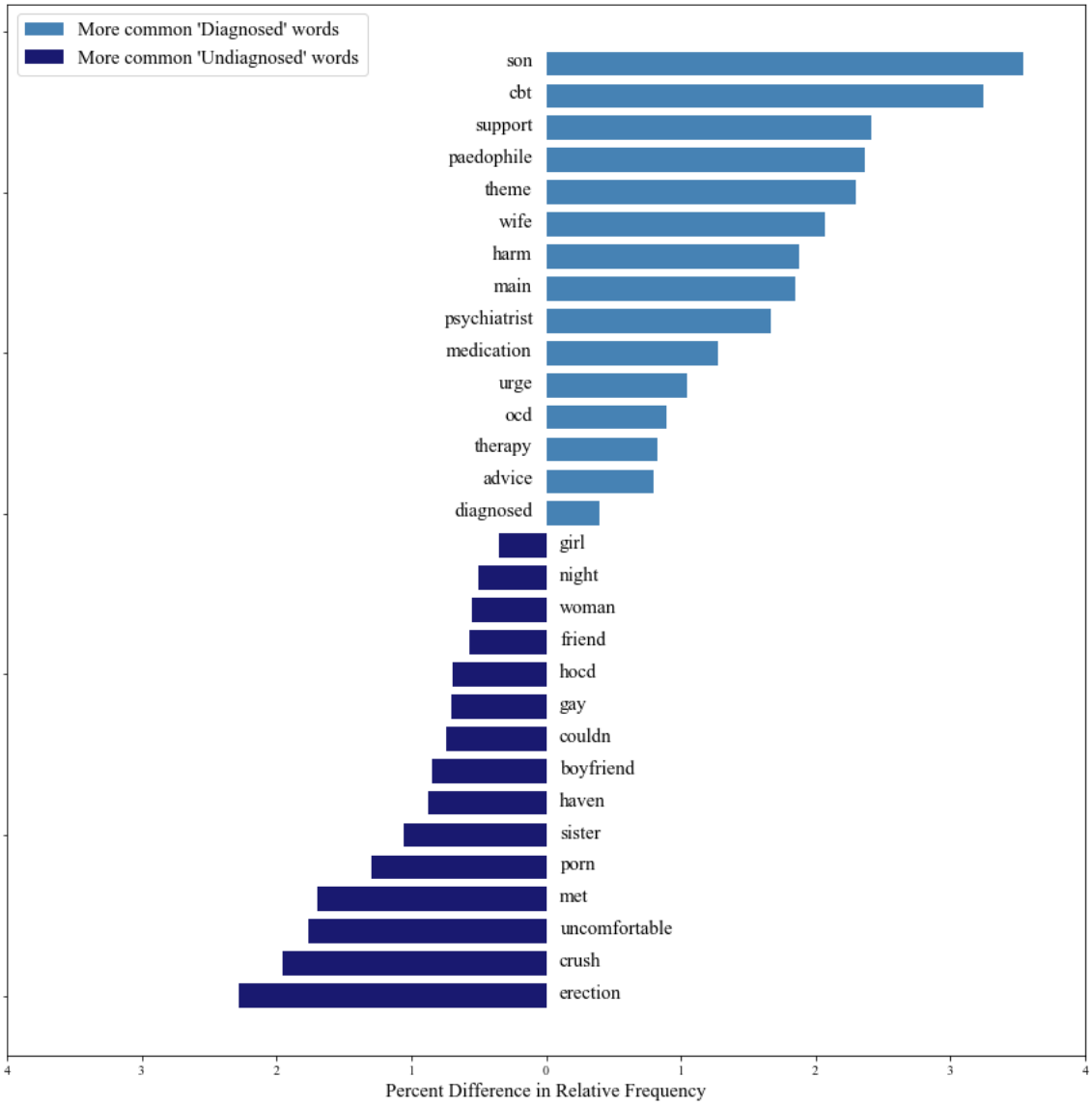

Figure 4: Significant Differences in Word Usage

and "erection." On the other hand, the diagnosed group hosted more mentions of "son", "wife", and "harm." We might infer a few things from these findings. For one, it makes sense that younger people should be less likely to be diagnosed with OCD simply because they have been alive for less time and therefore have had less time to be diagnosed (Chowdhury, Frampton, & Heyman, 2004; Storch, 2007). More

pertinent to most cases, however, is that OCD frequently goes undetected. One study estimated that the average OCD sufferer had 10 or more years go by from the time their symptoms began to their diagnosis (Stein et al., 2009). The other salient fact is that younger people may be expected to have fewer resources and knowledge to seek out mental health support as compared to older individuals. If an older person suspects they might have a mental health condition like OCD they are often more poised to make an appointment with a practitioner (MacDonald, Fainman-Adelman, Anderson, & Iyer, 2018).

Finally, there is evidence that some intrusive thoughts may be less recognizable as intrusive thoughts by mental health practitioners, or otherwise considered non-clinical (Rachman & de Silva, 1978). A young person presenting with relentless doubts about "the possibility of being gay" is less likely to be associated with a mental health condition than an adult describing thoughts about harm against their wife or child. It is also possible that intrusive thought themes relating to sexual orientation are more prevalent in younger sufferers. Intrusive thought content is known to vary for many sufferers over time, with many individual thoughts moving from theme to theme depending on their age and life circumstances. Anecdotal cases indicate that younger people may be more likely to experience these doubts surrounding their sexuality, as they are in a more sexually formative period of their lives (Igartua, 2015; Baer, 2001, p. 10). Older individuals with families may be more fixated with themes of harm, particularly to their spouses or children either as physical or sexual violence.

3.4.3 Topic Models

Significant differences in word frequencies can illustrate some differences in meaning across groups of text. However, an intrinsic limitation in this approach is that it treats individual words as though they exist in isolation of one another; understanding if a word appears more in one group versus another tells us nothing about its relationship to other words. As part of my semantic analysis of diagnosed and undiagnosed texts I sought to examine how words appeared together, and if identifiable "clusters" of words might correspond with the underlying meanings of the texts.

For this task I chose to use a Latent Dirichlet allocation, a topic modelling approach which aims to capture abstract "topics" contained in text based on probabilistic relationships among words (Blei, Ng, & Jordan, 2003). In other words, the LDA generates a probability distribution of words associated with each topic in a group of topics (cf. Blei, Ng, & Jordan, 2003, p. 4). To note, what constitutes a "topic" is ultimately subject to human interpretation and construction, and

therefore LDA relies on assumptions about the nature of topics in a corpus. In particular, the model relies on an assumption of the number of topics over which to generate word-probability distributions. Having observed significant differences in word frequencies which hinted at different topics in the forum texts, I had some knowledge of broad topic themes such as sexual orientation as well as psychiatry and medication. I determined that this prior knowledge made the LDA a good fit for the topic modelling task.

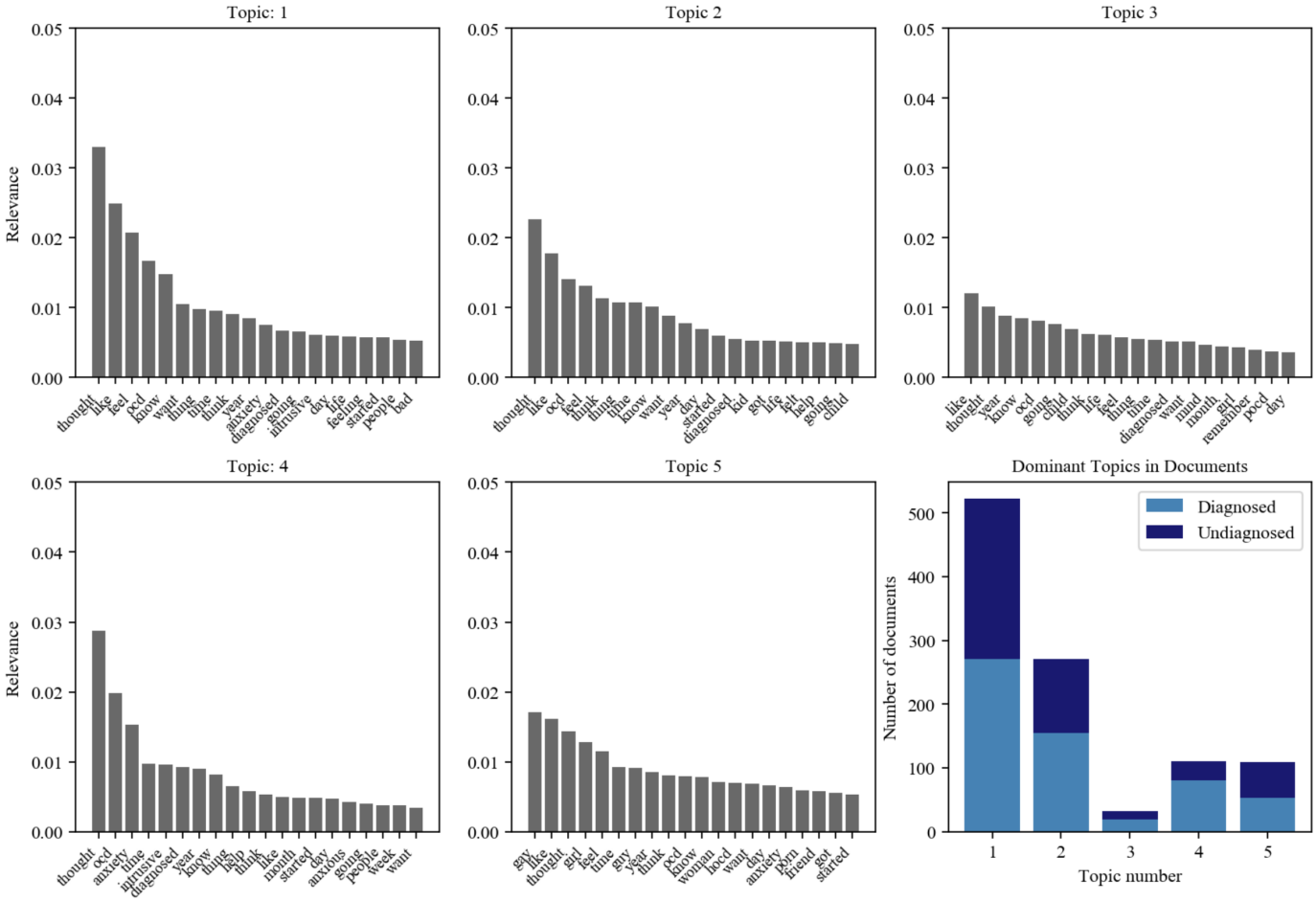

Figure 5: LDA Topic Distributions

I conducted an LDA over the entire corpus of diagnosed and undiagnosed texts, testing the allocation over different numbers of topics. Choosing too small a number yielded muddled word-probability distributions containing terms which seemed to be associated with multiple topics. On the other hand, choosing too large a number yielded either incoherent or highly specific topics that may have been associated with just a few posts. I found that 5 topics represented a "sweet spot" in the topic count tuning, as performing the LDA with this number consistently returned a coherent set of distinct topics, where each topic was dominant in at least some

documents from both undiagnosed and diagnosed texts. Figure 5. shows the 20 most relevant terms for each of the generated topics, where relevance was calculated in part according to a term's probability of occurring in the topic with respect to other terms (cf. Sievert & Shirley, 2015, p. 66). The bottom right subplot contains the distribution of documents of both classes according to their dominant document topic as scored according to the cumulative relevances of the words they contained.

The topics all contained very similar relevant terms. Words like "thought," "like," and "ocd," were consistently among the topics' most relevant terms with a slight exception of Topic 5, where the most relevant term was "gay." Looking further down the relevance scale, we begin to see terms indicative of different OCD themes relevant to each topic. Topic 3, for example, was the only one to include the term "pocd," and placed a relatively high relevance on the term "child" as compared to other topics. This would seem to indicate that posts where this topic was dominant centered on intrusive thoughts surrounding pedophilia, a theme borne out in some of the significant "diagnosed" terms identified in the previous section. Accordingly, most of the posts in which Topic 3 was dominant were authored by diagnosed posters (although these posts only represented around 3% of the entire corpus).

Topic 5 was dominant in about 10% of all posts, and showed keywords relevant to the theme of "Homosexual OCD," which were associated with the undiagnosed group in the analysis of word differences. This was the only topic to include in its most relevant terms "hocd," "porn," and "gay," and appears most dissimilar from the other topics. In the same way that these terms were identified as significantly more frequent among undiagnosed posters, we also observe this as the only topic which was dominant in slightly more undiagnosed posts than diagnosed posts, even with the overall class imbalance.

Despite these findings in topics 3 and 5, it should be noted that the most dominant topics by far were topics 1 and 2, which both showed roughly similar keywords as well as distributions in the diagnosed and undiagnosed classes. This result suggests that the bulk of the posts on the forum—whether they were authored by diagnosed or undiagnosed posters—were talking about the same things. In particular, the result suggests that the posts were largely about the experience of living with intrusive thoughts and being diagnosed with OCD, including details about when a person was diagnosed and how they feel about their condition. This is as opposed to posts where the dominant topic was the content of the thoughts themselves, which appear to have been more common in the undiagnosed posts based on the findings of Topic 5.

Prima facie, the topic modeling approach may appear largely as a "null finding,"

based on the relative inability of the LDA model to separate topics into distinct semantic groups. However, the result is significant when considered in light of previously observed differences in structure and word usage of the texts. Although there were significant differences in the ways in which diagnosed and undiagnosed posts were structured, as well as the words they contained, the prevailing topics across the entire corpus were largely consistent. Put differently, diagnosed and undiagnosed posters were talking largely about the same things; what did differ between the two classes was how they talked about intrusive thoughts.

3.5 Binary Group Classification

3.5.1 Statistical Classification of Text

In statistics and machine learning, classifiers are algorithms aimed at distinguishing two or more classes of entities based on the features those entities exhibit. When classifiers are used to identify classes associated with documents of textual data, they often use features representing the words contained in the texts and other grammatical or structural attributes of the documents. Classification of textual data can help identify the most important features that differentiate different document classes, and more generally to gauge the separability of the classes using those features (Bengfort, Ojeda, & Bilbro, 2018, ch. 4).

3.5.2 TFIDF Vectorization

In text classification tasks, vectorization is the process of rendering textual data in a quantitative format such that it can be processed by classification algorithms. In Natural Language Processing, a common approach to building such a representation is known as TFIDF vectorization (Bengfort, Ojeda, & Bilbro, 2018, ch. 3). TFIDF stands for "Token frequency inverse document frequency" and is among the most popular vectorizing tools used in research (Beel, Breitinger, Gipp, & Langer, 2016). TFIDF seeks to represent each token in a document as a weighted proportion of its occurrences in the document (term frequency) with respect to (inverse) the term's importance in the overall corpus (document frequency).

Using the Python library sk-learn I accomplished the vectorization on the processed forum posts (i.e. the lemmatized tokens with stopwords removed). This process produced a matrix of TFIDF vectors with each row vector representing a single post from the labelled sample. In this case I removed one token from the TFIDF vectors, which was the term "haven" and the lemma of the word "haven't." This token appeared more frequently in the diagnosed group and on preliminary

classification tests appeared routinely in the top five feature importances. The objective here is to try to make a prediction on a poster's reported diagnostic status based on tokens used, rather than the phrase where they actually describe their diagnostic status, like "I haven't been diagnosed"). To the greatest extent possible I wanted to prevent the classifiers from picking up on the actual diagnostic claims in the posts, which is why "haven" was removed.

3.5.3 Model Fitting

A number of machine learning techniques have been shown to perform especially well in text classification tasks. However, the results of classification algorithms depend heavily on the underlying texts to which they are applied and can vary widely (Kowsari et al., 2019). Therefore, the next step was fitting several models and comparing their performance. I chose three families of algorithms which take different approaches to the binary text classification task: random forests, support vector machines, and logistic regression models (Kowsari et al., 2019). Choosing algorithms from different families could lend insight into the underlying texts, depending on which family performed best.

- Random forests are collections of decision trees which each use randomized subsets of training data to predict the data's correct class. In the text classification context, the individual trees (also called estimators) attempt to differentiate the data based on the features of the text, such as word counts. The "Forest" is therefore an ensemble of individual decision trees where the number of trees can be chosen prior to model training (Ho, 1995).

- Support vector machines represent data as points in space, attempting to find the largest margin between groups of points where the margins represent class boundaries for the data. Applied to textual data, SVMs represent documents as points in a higher—dimensional space and attempt to separate the margin between documents according to specific tuning parameters. The class of new documents is predicted based on their location with respect to the class boundaries (Cortes & Vapnik, 1995).

- Logistic regressions use the logistic function to generate a probability that data belongs to one of two or more classes. Data points are represented as linear combinations of their features, known as the "log-odds." In the binary classification task, the logistic function is used to convert these log-odds into probabilities that a given data point belongs to each of the two classes. A final

classification is made depending on where a data point's probability exists between 0 and 1 (Cramer, 2003).

I tested each of these methods for identifying the diagnostic status of post authors based on the content of the posts themselves, represented as TFIDF vectors. For each algorithm I performed a train-test split in sk-learn, where 20% of the data was reserved for testing and the remaining 80% for training. To ensure results were robust, I also tested each algorithm using stratified K-fold cross validation, an approach which performs the classification task on different stratified subsets of the data (i.e., there is a consistent proportion of items belonging to each class—in this case diagnosed and undiagnosed—in each subset).

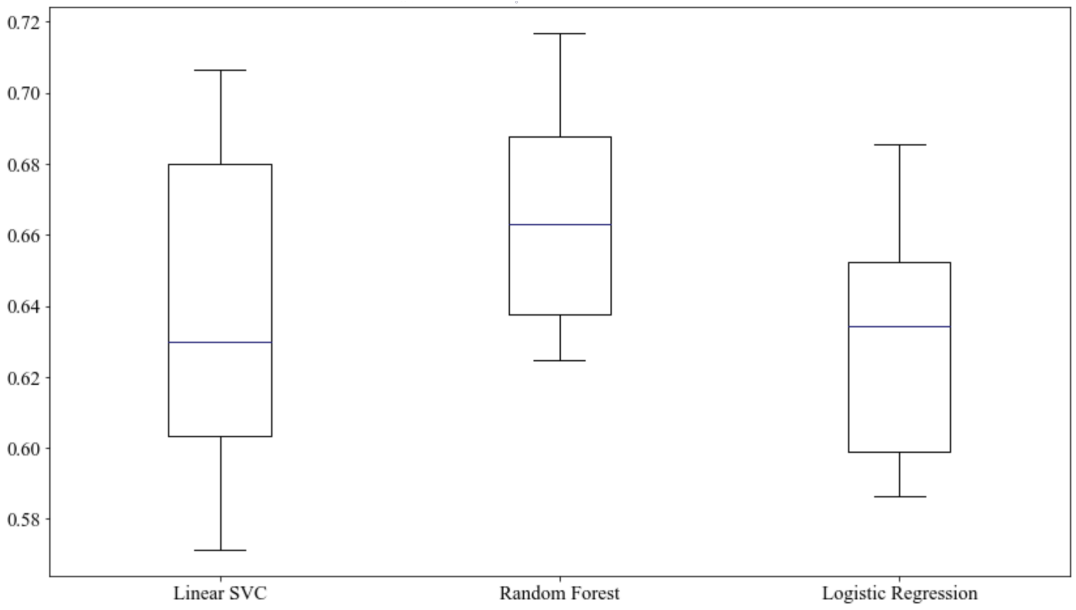

Figure 6: F1 Scores over 10 Stratified Folds

Figure 6 depicts F1 scores for each of the three algorithms across 10 stratified folds. The model with the highest overall F1 scores was the random forest, which achieved a median F1 score of close to 0.66 when tuned to 250 estimators on 500 TFIDF features. I next examined the performance of a sample random forest model tuned to the same parameters, displayed in Figure 6. The left plot displays feature importances for a classification resulting in an accuracy score of 0.7429. The right plot displays the confusion matrix for the same classification, illustrating the number of correctly labelled diagnosed posts, incorrectly labelled diagnosed posts, and so forth. To note, the random forest feature importances provide information on the features' relevance, but not their directionality. In other words, a high feature importance tells us that a particular token was prominent in the algorithm's internal

branching strategy, but not the "direction" of its importance—whether it led the model to predict diagnosed or undiagnosed authors (Reece et al., 2017). That being said, we may gain some information on the directionality of the features from the distributions of the individual tokens (i.e., whether they appeared more in diagnosed texts or vice-versa).

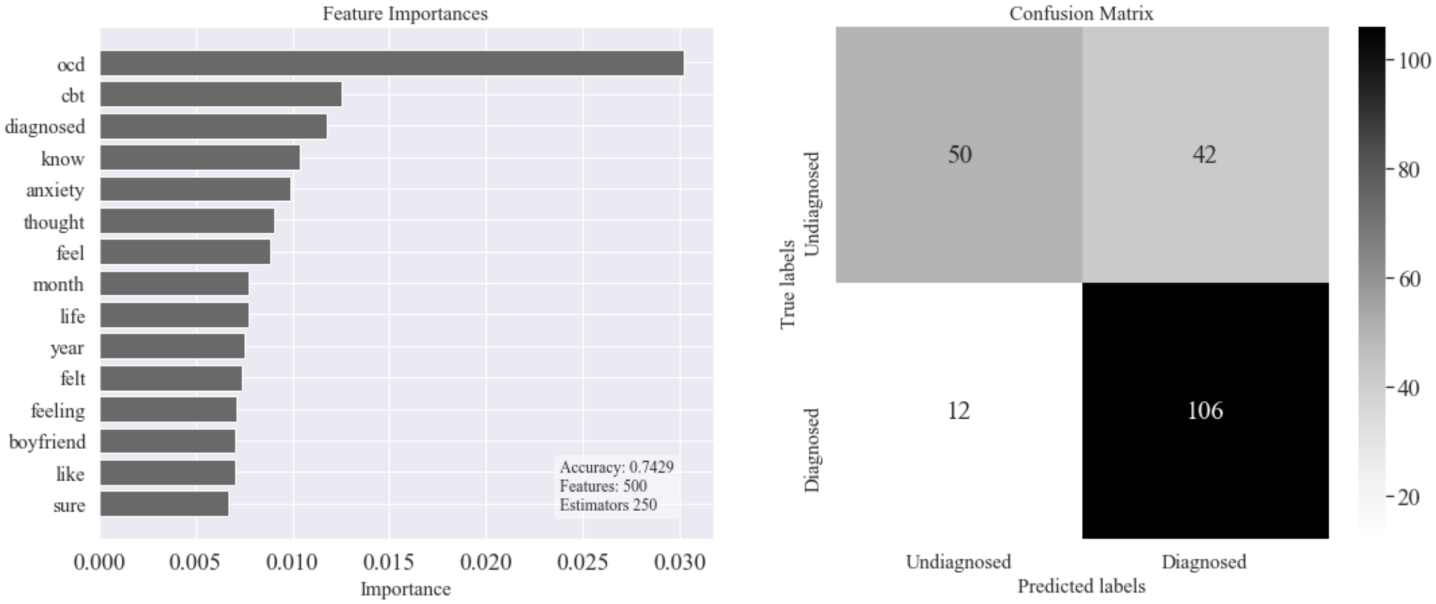

Figure 7: Random Forest Results

The most important feature by far within the random forest was the token "OCD" itself, with an importance more than twice as large as the second most important token "cbt" (referring to cognitive behavioral therapy). This lends some insight into why the random forest outperformed the support vector classifier and logistic regression models: The branching strategies within its trees could rely more heavily on "give-away" terms and their co-occurrences than the distance and linear combination strategies of the other models. In some sense, the random forest was the model which could best "overfit" the data. Indeed, the large importances for "OCD" and "cbt" correspond with the fact that both tokens were relatively common across both diagnosed and undiagnosed posts, but were significantly more likely to appear together in the diagnosed posts. To a lesser extent the high importance of "OCD" may have also been attributable to the presence of less important features like "HOCD," "POCD," and "ROCD," which corresponded more with undiagnosed posters. In other words, members of the undiagnosed group may have used other acronyms to describe the condition, like "HOCD," in lieu of simply "OCD." A similar effect could have applied to a much lesser extent for the term "diagnosed" which was more common in the diagnosed group, as opposed to "diagnosis," which was

slightly more common in the undiagnosed group. Table 3 displays normalized token counts for these terms in the two classes, where the normalizing constant was one plus the percent difference between the number of posts in each group (i.e., undiagnosed counts are scaled up by 19% to account for the class imbalance, whereas diagnosed counts are raw counts).

Table 3: Normalized Token Counts

	Diagnosed	Undiagnosed
"ocd"	1701	1261
"hocd"	93	220
"pocd"	165	223
"rocd"	37	60
"diagnosed"	645	646
"diagnosis"	35	46

The confusion matrix for the classifier shows that the random forest made most of its classifications as diagnosis positive and resultantly suffered from a relatively large number of false positives. The high rate of "diagnosed" classifications was likely the result of both the natural imbalance of classes in the data, as well as the high feature importances on words more closely associated with the diagnosed texts like "ocd" and "diagnosed" itself. To note, different approaches to vectorizing the data and further tuning of the classifiers may have produced higher F1 scores. In this context, however, the goal was not so much achieving the highest possible classification accuracy as determining the extent to which the data were separable by a binary classifier given basic preprocessing and tuning.

4 Discussion

In light of the structural and semantic differences between the diagnosed and undiagnosed groups, and the ability of several text classification models to separate the groups with relative accuracy, I conclude that significant differences exist in the ways diagnosed and undiagnosed individuals reported their intrusive thoughts in the online forum. The average post lengths across the two groups indicated that undiagnosed individuals sought to express more information per-post. These undiagnosed posters also tended to use more past tense verbs, as well as personal pronouns in their writing, which is a strong indication that their posts were of a more narrative or "storytelling" character than their diagnosed counterparts. This style is commensurate with narrative discourse among patients in clinical psychiatry, indicating

that many undiagnosed posters aimed to elicit diagnosis from their readers. On the other hand, already diagnosed posters' usage of significantly more present tense verbs indicated that their writing style was more expository of present ideas and actions.

Semantic analyses including significance testing for differences in usage of specific words as well as latent Dirichlet allocation of documents across the entire corpus produced more mixed results. On one hand, significant word differences suggested that undiagnosed posters were younger people describing bad thoughts about the possibility of being gay, straight, or unable to determine their sexual orientation. Diagnosed posters appeared to be comparatively older individuals who used more medicalized terms to describe their intrusive thoughts, and were more direct in referring to their condition as "OCD." On the other hand, an LDA was largely unable to separate the corpus' posts according to coherent topics, instead suggesting that the majority of posts revolved more around "meta" descriptions of the experience of living with intrusive thoughts and OCD than the visceral content of the authors' intrusive thoughts. So for example, a poster might be more likely to describe an intrusive thought as "That's when I had a thought that made me think—what if I hurt my son?" as opposed to "I saw a knife on the table and saw an image of myself stabbing my son." This tendency towards meta-description of intrusive thoughts is unsurprising considering that just writing such thoughts down can cause significant distress to OCD sufferers. In fact, repeatedly writing down bad thoughts is a common exposure strategy used in exposure and response prevention therapy (Hezel & Simpson, 2019). A seeming exception to the tendency towards meta-description was writing on the topic of "Homosexual OCD," where it appeared that authors were describing stories related to their intrusive thoughts about being gay, straight, and so forth in comparatively visceral terms. This topic was dominant in about 10% of all posts in the corpus, the majority of which were authored by undiagnosed posters. Testing with a variety of classification algorithms revealed that a minimally-tuned random forest model could predict the diagnostic status of the author of a post based on vectorized representation of its contents around 75% of the time. However, the classifier relied heavily on the term "OCD" itself, which could be interpreted as a form of data leakage into the training dataset (Brownlee, 2016).

To the best of my knowledge, these results represent the first systematic study of the online reporting of intrusive thoughts as they relate to OCD. The analyses paint a picture of individuals besieged by recurring bad thoughts about the worst things they can imagine. These are people who go online to share their experi-

ences and seek support from a sympathetic audience and with relative anonymity. Following De Choudhury & He (2014), my results suggest that online community forums offer many sufferers of stigmatic mental health problems a uniquely safe and supportive environment for information and support-seeking. And similar to (De Choudhury, Gamon, Counts, & Horvitz (2013), Reece et al., (2017), and (Coppersmith, Dredze, & Harman, 2014) my results point to significant differences in the behavior of individuals diagnosed with specific mental health conditions on social media as compared to undiagnosed individuals. Unlike those studies, however, I tried to identify these differences across diagnosed and undiagnosed individuals who in both cases were describing their experiences of what they believed to be their intrusive thoughts.

One contribution of my results is to portray what these differences looked like. In the broadest terms, diagnosed individuals and undiagnosed individuals spoke about largely the same things in their posts—bad thoughts, bad feelings, and requests for support, information, and in some cases diagnoses. However, they differed significantly in the ways they went about saying those things. Undiagnosed individuals preferred to tell stories related to their bad thoughts and the ways they affected them and the people close to them. The words used by these individuals suggested they were on average younger than their diagnosed counterparts, with many bad thoughts revolving around their sexual orientation at a formative point in their lives. Diagnosed authors also spoke about similar intrusive thoughts, as well as many other topics such as pedophilia and harm. They did so with a more present-oriented style, as well as words indicating a greater distance from the content of the thoughts and their identities overall. In this view, "I suffer from intrusive thoughts with themes about homosexuality and I'm seeking advice" would be a phrase more characteristic of a diagnosed poster, whereas "I was with my boyfriend recently when I thought 'what if I'm gay?' Do you think I have HOCD?" would be more emblematic of an undiagnosed poster. Following Giles and Newbold (2011), these results suggest that many individuals frequenting online mental health forums have had little or no interaction with mental health professionals and are attempting to elicit diagnoses from other forum visitors. This may be particularly likely among younger visitors to such forums whose general reliance on the Internet for information has led them to believe that an online diagnosis may substitute for one made by a mental health professional (Lupton & Jutel, 2015). This effect may also be compounded by a lack of access to or information about mental healthcare, or because of stigma associated with it ((MacDonald, Fainman-Adelman, Anderson, & Iyer, 2018). In any case, online communities centered on mental health should make clear to visitors that only

a licensed professional can diagnose them with a mental health condition, providing instructions for seeking out such a professional offline.

Following from Baer (2001), Shafran, Watkins, & Charman, (1996) and Hezel & Simpson (2019), my results align with what might be expected from past research on treatment outcomes in OCD. Simply put, diagnosed individuals are more likely to have been treated for the condition with pharmacological and therapeutic interventions. Through practices like cognitive-behavioral therapy, they have learned to disassociate their bad thoughts from themselves. "You are not your thoughts!" wrote the International OCD Foundation (2017)—a phrase used by many an exposure and response prevention therapist.

A concurrent contribution of this research is to show that signals related to the diagnosis and treatment of OCD (and the lack thereof) are discernible in a small sample of discourses drawn from an online mental health forum. The implications of this finding are two-fold. First, continuing research into larger, more variegated, social datasets pertaining to OCD may yield further insights into the nature of the condition and the ways it is diagnosed and treated. Second, online mental health forums revolving around other mental health conditions—and not just major social media platforms—may prove fruitful data sources for future research on mental health more generally. This may be of particular import to researchers working to advance the United States National Institute of Mental Health's new Research Domain Criteria Inititive (RDoC), which aims to "Develop, for research purposes, new ways of classifying mental disorders based on dimensions of observable behavior and neurobiological measures," including through patient self-reports (National Institute of Mental Health, 2008).

4.1 Limitations

My results are tempered by several limitations in my research design and methods. First, The nature of the dataset used in this research diminishes the findings' generalizability. For one, the data were relatively small when compared to other studies on online textual data, and also collected from just one platform. To note, the sample size of posters was comparable with those in other studies like De Choudhury, Gamon, Counts, & Horvitz (2013) and Reece et al., (2017), but in both these cases the authors point to small sample sizes as a pitfall to mental health research of this kind.

Platform effects are also known to have a major impact on online discourses, and in particular what information people wish to share and how they wish to share it (Aragón, Gómez, & Kaltenbrunner, 2017; La Violette & Hogan, 2019). Normally,

a discussion of the specific platform used for data collection would be warranted for this type of research. However, in an effort to protect the anonymity of posters and the privacy of the forum this report did not go into detail in this regard. I therefore did not acknowledge what influences the specific platform may have had on the forum posts. Relatedly, collecting data from just this platform meant that my results were drawn from posts written by English-speaking authors with access to the Internet and the savvy to use an online forum. It is likely therefore that the population of posters was skewed socially and geographically, presumably to middle class or affluent regions of the Global North (Graham, Hogan, Straumann, & Medhat, 2014). Intrusive thought symptoms and OCD differ cross-culturally (Williams & Steever, 2015; Nicolini, Salin-Pascual, Cabrera, & Lanzagorta, 2018), but this research did not factor such cultural differences in its design.

Another limitation pertaining to the internal validity of this research was my inability to confirm whether or not posters in my sample were accurate in reporting their diagnostic status. Some users are likely to have been equivocal about their diagnostic statuses, while others may have misrepresented them. For example, an individual who was diagnosed with major depression but whose psychiatrist mentioned "intrusive thoughts" (also a symptom of major depression) may have become convinced that they were suffering with OCD instead and gone on to report having been diagnosed with OCD on the forum. There will also have been extreme heterogeneity in the context and nature of individuals' diagnoses. A "diagnosis" from somebody other than a mental health practitioner should not count as a diagnosis according to the formal diagnostic criteria for OCD (APA, 2013), although a number of posters would have reported it as such. Taken together, these factors pare down the reliability of my working distribution of diagnostic statuses with respect to their true "ground truth" labels.

Finally, I consider a major limitation of this research to be the lack of information about the temporal context of posters' diagnostic statuses. My approaches to data collection and annotation could not make provision for discerning when a diagnosis had been made with respect to the posters' writing. The time that had elapsed since an individual's diagnosis could have major impacts on their symptom reporting, especially depending on the nature and continuity of care they had received (if any) in the interim (Joyce et al., 2004). Consider the case of an individual who was diagnosed with OCD years or even decades before they wrote something on the forum but received no additional mental healthcare since. Due to the amount of interceding time, we may expect such an individual to use terms more similar to individuals who were undiagnosed when they discussed their intrusive thoughts.

In as much as these limitations temper the findings of this study, they should not belie its contributions. The results provide new insight into differentiating factors in online reporting of intrusive thought reporting between diagnosed and undiagnosed individuals, as well as the alignment of those differences with observed diagnosis and treatment outcomes for OCD. The results also demonstrate the value of understudied online forums for research into mental health. As classifications of mental health conditions expand to encapsulate new forms of social data and patient self-reports, data from such forums may be particularly impactful (NIMH, 2008). Future research may seek to account directly for the limitations highlighted above, including direct engagements with the temporal dynamics in diagnosis and symptom reporting, as well as culturally dependent factors pertaining to the content of intrusive thoughts as they are reported online.

Acknowledgements

<Withheld for review>

References

Abramowitz, J. S., Blakey, S. M., Reuman, L., & Buchholz, J. L. (2018). New Directions in the Cognitive-Behavioral Treatment of OCD: Theory, Research, and Practice. *Behavior Therapy*. doi: 10.1016/j.beth.2017.09.002

Alzahrani, H. (2018). Analysis of Parts-of-Speech Distribution and Omission Patterns in The New York Times and The Guardian. *International Journal of Linguistics*. doi: 10.5296/ijl.v10i3.13066

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (5th Edition)*. doi: 10.1176/appi.books.9780890425596.744053

Andersson, M. A., & Harkness, S. K. (2018). When Do Biological Attributions of Mental Illness Reduce Stigma? Using Qualitative Comparative Analysis to Contextualize Attributions. *Society and Mental Health*. doi: 10.1177/2156869317733514

Aragón, P., Gómez, V., & Kaltenbrunner, A. (2017). Detecting Platform Effects in Online Discussions. *Policy and Internet*. doi: 10.1002/poi3.158

Baer, L. (2001). *The Imp of the Mind: Exploring the Silent Epidemic of Obsessive Bad Thoughts*. Plume.

Beadel, J. R., Green, J. S., Hosseinbor, S., & Teachman, B. A. (2013). Influence of age, thought content, and anxiety on suppression of intrusive thoughts. *Journal of Anxiety Disorders*. doi: 10.1016/j.janxdis.2012.12.002

Beel, J., Gipp, B., Langer, S., & Breitinger, C. (2016). Research-paper recommender systems: a literature survey. *International Journal on Digital Libraries*. doi: 10.1007/s00799-015-0156-0

Benjamin, B., Tony, O., & Rebecca, B. (2018). *Applied Text Analysis with Python.* O'Reilly Media.

Blei, D. M., Ng, A. Y., & Edu, J. B. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research.*

Brownlee, J. (2016). *Data Leakage in Machine Learning.* Retrieved from `https://machinelearningmastery.com/data-leakage-machine-learning/`

Choudhury, M. D., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting Depression via Social Media. *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media.*

Chowdhury, U., Frampton, I., & Heyman, I. (2004). Clinical characteristics of young people referred to an obsessive compulsive disorder clinic in the United Kingdom. *Clinical Child Psychology and Psychiatry.* doi: 10.1177/1359104504043922

Clark, D. A., & Radomsky, A. S. (2014). Introduction: A global perspective on unwanted intrusive thoughts. *Journal of Obsessive-Compulsive and Related Disorders.* doi: 10.1016/j.jocrd.2014.02.001

Coppersmith, G., Dredze, M., & Harman, C. (2014). Quantifying Mental Health Signals in Twitter. In *Proceedings of the workshop on computational linguistics and clinical psychology: From linguistic signal to clinical reality.* doi: 10.3115/v1/W14-3207

Cortes, C., & Vapnik, V. (1995). Support-Vector Networks. *Machine Learning.* doi: 10.1023/A:1022627411411

Cramer, J. S. (2003). The origins and development of the logit model. *Logit models from economics and other fields.*

De Choudhury, M., & De, S. (2014). Mental Health Discourse on reddit: Self-Disclosure, Social Support, and Anonymity. *Proceedings of the Eight International AAAI Conference on Weblogs and Social Media.*

Fellbaum, C. (2005). *WordNet: About WordNet.*

Foucault, M. (1963). *Madness and Civilization: A History of Insanity in the Age of Reason.* Penguin. doi: 10.2307/2737615

Foundation, I. O. (2017). *You Are Not Your Thoughts.* Retrieved from `https://iocdf.org/blog/2017/11/10/you-are-not-your-thoughts/`

Giles, D. C., & Newbold, J. (2011). Self- and other-diagnosis in user-led mental health online communities. *Qualitative Health Research.* doi: 10.1177/1049732310381388

Goffman, E. (1963). *Stigma: Notes on the Management of Spoiled Identity.* Penguin. doi: 10.2307/2575995

Goodman, W. K., Price, L. H., Rasmussen, S. A., Mazure, C., Fleischmann, R. L., Hill, C. L., ... Charney, D. S. (1989). The Yale-Brown Obsessive Compulsive Scale: I. Development, Use, and Reliability. *Archives of General Psychiatry.* doi: 10.1001/archpsyc.1989.01810110048007

Graham, M., Hogan, B., Straumann, R. K., & Medhat, A. (2014). Uneven Geographies of User-Generated Information: Patterns of Increasing Informational Poverty. *Annals of the Association of American Geographers.* doi: 10.1080/00045608.2014.910087

Henderson, C., Evans-Lacko, S., & Thornicroft, G. (2013). *Mental illness stigma,*

help seeking, and public health programs. doi: 10.2105/AJPH.2012.301056

Hezel, D., & Simpson, H. (2019). Exposure and response prevention for obsessive-compulsive disorder: A review and new directions. *Indian Journal of Psychiatry*. doi: 10.4103/psychiatry.indianjpsychiatry_516_18

Ho, T. K. (1995). Random decision forests. In *Proceedings of the international conference on document analysis and recognition, icdar.* doi: 10.1109/ICDAR .1995.598994

Holmes, J. (2000). Narrative in psychiatry and psychotherapy: the evidence? *Medical Humanities*. doi: 10.1136/mh.26.2.92

Honnibal, M. (2013). *A Good Part-of-Speech Tagger in about 200 Lines of Python.* Retrieved from `https://explosion.ai/blog/part-of-speech-pos-tagger-in-python`

Igartua, K. J. (2015). Distinguer le processus d'acceptation d'une identité sexuelle minoritaire d'un trouble obsessionnel compulsif avec obsessions sexuelles. *Santé mentale au Québec*. doi: 10.7202/1034915ar

International OCD Foundation. (n.d.). *What is OCD?* Retrieved from `https://iocdf.org/about-ocd/`

Joyce, A. S., Wild, T. C., Adair, C. E., McDougall, G. M., Gordon, A., Costigan, N., ... Barnes, F. (2004). *Continuity of care in mental health services: Toward clarifying the construct.* doi: 10.1177/070674370404900805

Kilgarriff, A. (1996). Comparing word frequencies across corpora: Why chi-square doesn't work, and an improved LOB-Brown comparison. *ALLC-ACH Conference*, 169–172.

Kilgarriff, A. (2007). Comparing Corpora. *International Journal of Corpus Linguistics*. doi: 10.1075/ijcl.6.1.05kil

Kowsari, K., Meimandi, K. J., Heidarysafa, M., Mendu, S., Barnes, L., & Brown, D. (2019). *Text classification algorithms: A survey.* doi: 10.3390/info10040150

Lack, C. W. (2013). Obsessive-compulsive disorder: Evidence-based treatments and future directions for research. *World Journal of Psychiatry*. doi: 10.5498/wjp.v2.i6.86

Lijffijt, J., Nevalainen, T., Säily, T., Papapetrou, P., Puolamäki, K., & Mannila, H. (2016). Significance testing of word frequencies in corpora. *Literary and Linguistic Computing*, *31*(2), 374–397. doi: 10.1093/llc/fqu064

Lupton, D., & Jutel, A. (2015). 'It's like having a physician in your pocket!' A critical analysis of self-diagnosis smartphone apps. *Social Science and Medicine*. doi: 10.1016/j.socscimed.2015.04.004

MacDonald, K., Fainman-Adelman, N., Anderson, K. K., & Iyer, S. N. (2018). *Pathways to mental health services for young people: a systematic review.* doi: 10.1007/s00127-018-1578-y

Moore, P. S., Mariaskin, A., March, J., & Franklin, M. E. (2007). Obsessive-compulsive disorder in children and adolescents: Diagnosis, comorbidity, and developmental factors. *Storch, Eric A [Ed]; Geffken, Gary R [Ed]; Murphy, Tanya K (2007) Handbook of child and adolescent obsessive-compulsive disorder (pp 17-45) xvi, 415 pp Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers; US*.

Munková, D., Munk, M., & Vozár, M. (2013). Data pre-processing evaluation for

text mining: Transaction/sequence model. In *Procedia computer science.* doi: 10.1016/j.procs.2013.05.286

Nicolini, H., Salin-Pascual, R., Cabrera, B., & Lanzagorta, N. (2018). Influence of Culture in Obsessive-compulsive Disorder and Its Treatment. *Current Psychiatry Reviews*. doi: 10.2174/2211556007666180115105935

NIMH. (2008). *The National Institute of Mental Health Strategic Plan* (Tech. Rep.). https://web.archive.org/web/20081217154853/http://www.nimh.nih.gov/about/strategic-planning-reports/index.shtml.

Nugues, P. M. (2014). Words, parts of speech, and morphology. In *Words, parts of speech, and morphology.* Springer. doi: 10.1007/978-3-642-41464-0_6

OCD Life. (2017). *OCD Sufferers: Google is not your friend.* Retrieved from `https://ocdlife.ca/ocd-sufferers-google-is-not-your-friend/`

Osborne, I. (1998). Tormenting Thoughts and Secret Rituals: The Hidden Epidemic of Obsessive-Compulsive Disorder. *Booklist*.

Penzel, F. (2007). *How Do I Know I'm Not Really Gay?*

Phelan, J. C., Link, B. G., Stueve, A., & Pescosolido, B. A. (2006). Public Conceptions of Mental Illness in 1950 and 1996: What Is Mental Illness and Is It to be Feared? *Journal of Health and Social Behavior*. doi: 10.2307/2676305

Phillipson, S. J. (1989). Thinking the Unthinkable. *Center for Cognitive Behavioral Pyschotherapy*. Retrieved from `https://www.ocdonline.com/thinking-the-unthinkable`

Pillay, S. (2010). *The Dangers of Self Diagnosis.* Retrieved from `https://www.psychologytoday.com/gb/blog/debunking-myths-the-mind/201005/the-dangers-self-diagnosis`

Pustejovsky, J., & Stubbs, A. (2012). *Natural language annotation for machine learning.* O'Reilly Media. doi: 1332788036

Rachman, S., & de Silva, P. (1978). Abnormal and normal obsessions. *Behaviour Research and Therapy*. doi: 10.1016/0005-7967(78)90022-0

Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling.. doi: 10.3115/1604683.1604686

Reece, A. G., Reagan, A. J., Lix, K. L., Dodds, P. S., Danforth, C. M., & Langer, E. J. (2017). Forecasting the onset and course of mental illness with Twitter data. *Scientific Reports*. doi: 10.1038/s41598-017-12961-9

Shafran, R., Watkins, E., & Charman, T. (1996). Guilt in obsessive-compulsive disorder. *Journal of Anxiety Disorders*. doi: 10.1016/S0887-6185(96)00026-6

Sievert, C., & Shirley, K. (2015). LDAvis: A method for visualizing and interpreting topics.. doi: 10.3115/v1/w14-3110

Simpson, H. (2017). *Obsessive-compulsive disorder in adults: Epidemiology, pathogenesis, clinical manifestations, course, and diagnosis.* UpToDate. Retrieved from `https://www.uptodate.com/contents/obsessive-compulsive-disorder-in-adults-epidemiology-pathogenesis-clinical-manifestations-course-and-diagnosis`

Stein, D. J., Denys, D., Gloster, A. T., Hollander, E., Leckman, J. F., Rauch, S. L., & Phillips, K. A. (2009). *Obsessive-compulsive Disorder: Diagnostic and Treatment Issues.* doi: 10.1016/j.psc.2009.05.007

Tait, A. (2016). *"It's like stepping into the storm": How OCD can affect your online*

life. Retrieved from https://www.newstatesman.com/science-tech/2016/11/it-s-stepping-storm-how-ocd-can-affect-your-online-life

Veale, D. (2007). Psychopathology of obsessive–compulsive disorder. *Psychiatry*, *6*(6), 225–228.

Violette, J. L., & Hogan, B. (2019). Using platform signals for distinguishing discourses: The Case of Men's Rights and Men's Liberation on Reddit. *Proceedings of the Thirteenth International AAAI Conference on Web and Social Media*.

Williams, M., & Steever, A. (2015). Cultural manifestations of obsessive-compulsive disorder. In *Obsessive-compulsive disorder: Etiology, phenomenology, and treatment* (pp. 63–84).

Williams, M. T., Farris, S. G., Turkheimer, E., Pinto, A., Ozanick, K., Franklin, M. E., ... Foa, E. B. (2011). Myth of the pure obsessional type in obsessive-compulsive disorder. *Depression and Anxiety*. doi: 10.1002/da.20820

Řehůřek, R. (2019). *parsing.preprocessing - Functions to preprocess raw text.* Retrieved from https://radimrehurek.com/gensim/parsing/preprocessing.html