Natural Language Processing Exam, winter 2022 47422 characters (19.8 pages)
Eva Sahlholdt Hansen (201805908) and Liv Tollånes (201905665)

Uncovering Patient Experiences with SSRIs and SNRIs: A BERTopic and Sentiment Analysis of Depression

Abstract

Depression is a mental health disorder defined by longer periods of sadness, hopelessness, and fatigue. Antidepressants such as serotonin reuptake inhibitors (SSRIs) and serotonin and norepinephrine reuptake inhibitors (SNRIs) are part of the treatment course for moderate- to severe cases of the condition. Due to the phenomenological nature of depression, it is important to assess patient experience with antidepressant drugs and the impact of these. Current methods for doing this are time consuming, hard to compare and generalise. Connected to this is what appears to be a research gap in large-scale and systematic assessment of non-restricted accounts of patient experiences with medical treatment of depression. The study at hand attempts to overcome this by using NLP to explore patient experiences in free-text reviews of SSRIs and SNRIs specific to clinical depression. Specifically, we explore whether topic modelling can provide useful insights into which themes commonly occur amongst patients reviewing these drug categories, and whether these vary according to patient satisfaction. The analysis is further nuanced in a sentiment analysis. The results are presented and discussed. Overall, we find that several of the derived topics relate to well-known side effects of SSRIs and SNRIs. Also, positive experiences related to treatment and economic concerns are identified. We also detect different topic distributions for the two drug categories and for their prevalence in positive and negative reviews. These generally align well with the literature. Conclusively, we find that NLP-methods such as topic modelling can provide insight into central aspects of patient experience with drug treatment, albeit with a range of methodological as well as interpretative limitations.

Disclaimer:

Names of the authors who contributed to each section are noted after all subtitles.

Link to GitHub Repository with code and data can be found in the appendix.

Introduction

Clinical depression is a common mental health issue, estimated by the WHO to affect 280 million people globally (World Health Organization, 2021). It is an affective disorder, characterised by prolonged periods of low mood, reduced energy levels, troubled thought patterns, lack of positive emotions, and disturbed sleep patterns (Ruiz et al., 2018; Bear et al., 2016). The affective nature of the condition does however introduce a large variability in how it is experienced. Depression is destructive to life quality, and inflicts high societal and economic costs as well (Sartorius, 2001). The disorder normally lasts between four and twelve months, but recurrence rates are high, especially when left untreated (Bear et al., 2016). For many depressed patients, pharmacological options are an important part of the treatment process. In Denmark, antidepressant use exceeds 400.000 patients every year (Sundhedsdatastyrelsen, 2022). It is therefore important to understand how patients experience effects of these drugs upon their condition, as well as upon their general welfare. These measures might be possible to detect in free-form, written drug reviews using tools from the expanding field of natural language processing (NLP). In the current paper, we utilise state-of-the-art NLP techniques, specifically topic modelling and sentiment analysis, with the aim of assessing central features of how treatment with medically approved drugs for depression is experienced. We address the following questions:

1. Which semantic themes and overall sentiment occur in drug reviews for SSRIs and SNRIs, and do these differ according to the general satisfaction rate?
2. Can NLP-methods be used to derive informative indicators about how different types of medical drugs used to treat depression are experienced by patients?

Theoretical background

Pharmacological treatment: SSRIs and SNRIs **(Liv)**

In the current paper, we focus on the two most commonly prescribed antidepressant types used for depression: Selective serotonin reuptake inhibitors (SSRIs) and serotonin and norepinephrine reuptake inhibitors (SNRIs). Generally, medical treatment of depression is only advised for

moderate- to severe cases of depression (Park & Zarate, 2019). In Denmark, SSRIs are the dominant antidepressant drug class, considered first-line medical treatment and used by approximately 250.000 patients each year (Sundhedsdatastyrelsen, 2022). SNRIs are a newer type of antidepressant, in Denmark prescribed as a second-line treatment option (Eplov et al., 2005).

Both SSRIs and SNRIs act on neurotransmitters in the brain. Specifically, they target serotonin and norepinephrine (noradrenaline). Serotonin is an important neurotransmitter in the sympathetic nervous system, associated with general cognitive functioning as well as with emotion and motivation (Meneses & Liy-Salmeron, 2012). Norepinephrine is a neurotransmitter related to regulation of sleep patterns, attention, memory, and mood (Moret & Briley, 2011). The monoamine hypothesis of depression suggests that depleted levels of these neurotransmitters (including dopamine) are causally related to the propensity to develop depression (Köhler et al., 2016). SSRIs inhibit presynaptic reuptake of serotonin to increase the serotonin levels in the brain (Edinoff et al., 2021), while SNRIs inhibit reuptake of both serotonin and norepinephrine to increase levels of both neurotransmitters (Fanelli et al., 2021). Although being the main underlying mechanism of these medications, the monoamine hypothesis is subject to critique. For example, there is a lack of consistent evidence that serotonin and depression are even associated, not to mention causally related (Moncrieff et al., 2022). Also, it has been suggested that effects found across clinical studies might relate to placebo effects (Kirsch, 2014, 2019), and even publication bias (Mathew & Charney, 2009). As such, there is no consensus on whether SSRIs and SNRIs work in the way originally thought (ref.) Despite these concerns, SSRIs and SNRIs do seem to relieve depressive symptoms in a proportion of patients (Bschor & Kilarski, 2016; Cartwright et al., 2016), with some evidence suggesting that SNRIs are slightly more effective than SSRIs (Fanelli et al., 2021). This might be related to their broader targeting in the neurotransmitter systems (Tian et al., 2022). However, they are also related to worse side-, interaction-, and withdrawal effects (Cipriani et al., 2018). Thus, SNRIs are usually prescribed only when SSRIs do not show satisfactory results (Eplov et al., 2005).

Generally, antidepressant treatment is accompanied by unwanted side effects. Even SSRIs are problematic, despite being recommended as first-line treatment due to their relatively

advantageous ratio between desired effect and side effects. It is estimated that the majority of patients treated with antidepressants such as SSRIs will experience one or multiple side effects (Cartwright et al., 2016). Due to the functional similarity of SSRIs and SNRIs, their related side effects appear to be similar, but with more research available on SSRIs (Fava et al., 2015; Fava et al., 2018). Common side effects include weight gain, cognitive impairments, emotional blunting and gastrointestinal distress (Goodwin et al., 2017). Additionally, sexual dysfunction, bleeding, and hyponatremia have been found to be more prominent in antidepressants with high serotonin selectivity, which is slightly higher in SSRIs than SNRIs (Wang et al., 2018). For SNRI use, common side effects include increased levels of anxiety, sleep-related issues, and sexual dysfunction (although perhaps to a lesser extent than for SSRIs) (Santarsieri & Schwartz, 2015). Compared to SSRIs, SNRIs are more clearly associated with nausea, dry mouth, and elevated blood pressure (Wang et al., 2018). Other side effects observed in both drug groups include altered sleep patterns, seizures, and possibly increased suicidality (Edinoff et al., 2021). A complete list of side effects would be extensive, as there is great variability between specific drugs, and high variability at an individual level.

It has been estimated that in up to 43 % of cases, side effects lead to premature treatment discontinuation (Hung, 2014). This is an issue in regards to the healing prospect. Further, it can lead to undesired withdrawal effects (Fava et al., 2018; Stockmann et al., 2018). For SSRIs, common symptoms of withdrawal include dizziness, nausea, anxiety, headaches, fatigue, sweating, and sleep alterations such as insomnia and nightmares. For SNRIs, there is less research available on the consequences of withdrawal, but a central tendency appears to be significant neurological symptoms. These are often described as ‘brain zaps’, which are sensations of electric shocks inside the head. Other neurological symptoms described by patients are experiences such as ‘brain sloshing’ and ‘vision lagging behind eye movements’ (Stockmann et al., 2018). Generally, withdrawal symptoms are an important issue which needs to be considered accordingly when evaluating the effect of SSRIs and SNRIs (Fava et al., 2018)

Lastly, it is relevant to address that treatment with medical drugs is not always successful. Rather, it has been estimated that over a third of patients with depression do not respond in a

satisfactory manner to first-line medication (Ruberto et al., 2020). The relatively high rate of unresponsiveness might relate to drug tolerance (Fava, 2020). It has also been suggested that treatment success correlates with placebo effects, i.e. belief in and expectancies towards treatment (Kirsch, 2019; Rutherford et al., 2017). The impact of such psychological effects are increasingly being recognized, and underscores the importance of understanding patient attitudes, thoughts, and concerns about treatment.

Difficulties in assessing the effectiveness of pharmacological treatment approaches **(Eva)**

Reliable measures of drug effects upon depression are challenging to obtain for several reasons. Firstly, medicine is generally a complementary treatment to non-medical interventions such as psychotherapy (McAllister-Williams et al., 2020), or prescribed in combination with other medicines. Assessment of the causal effect of a drug can therefore be challenging to discern from other measures. Another challenge is that treatment of depression aims to alter mental states of the patient, which are inherently difficult to measure. It has been found that large-scale research often fails to properly consider patient experiences (Cartwright et al., 2016). Most of what is known about the effect of pharmacological interventions on depression derive from research into quantifiable measures from surveys and rating scales (Johnston, 2013). However, whilst such measures make it possible to quantify and compare mental states, they suffer from certain limitations. These include induction of priming effects, limiting patients in free expression, and regression to the mean (Uher, 2018). Therefore, it is also important to assess resources which allow patients to express themselves freely. However, research queries into such un-restricted resources are challenging due to issues related to time, costs, and scalability, as it usually involves human effort in collecting and interpreting verbal reports. Research of this kind, using e.g. interviews (Wiles et al., 2018), open-ended surveys (Cartwright et al., 2016), and posts from Internet forums are therefore generally low-scale and performed in a limited magnitude (Bumpus, 2020; Stockmann et al., 2018). Thus, NLP-methods are a relevant course of action to explore to approximate an approach into large-scale, systematic and reliable assessment of patient experiences with medical treatment of depression.

Previous research using NLP **(Eva)**

The toolbox developed in the field of natural language processing is increasingly being used in the sciences for diverse objectives. They can for instance be used to uncover latent content in large quantities of text, making it possible to assess large-scale linguistic patterns in a quantified manner. Here, we review previous research which has applied NLP-tools to text-based accounts related to depression.

As for the time being, most applications of NLP in research on depression are concerned with *detection* (e.g. ref). Much of this research focuses on free-text data from social media, and aims to develop models which based on written accounts can identify individuals with e.g. untreated depression, high-risk profiles, pre-depression markers, or in high risk of suicide (Nanomi Arachchige et al., 2021; Zhang et al., 2022; Tadesse et al., 2019). This is clearly important, as it might make it possible to provide help at an early stage. A similar approach is to investigate the aspect in which written accounts relating to depression differ from other content; for instance, a study from 2014 applied i.a. sentiment analysis and topic modelling to blog-style texts from online communities broadly relating to depression in order to assess whether these measures could be used to discriminate depression- from non-depression-related community posts (Nguyen et al., 2014). Another application which has received some attention is using NLP to extract salient information from electronic health records (EHRs), which are usually in free-text form. In relation to depression, NLP-methods applied to EHRs have shown some success, e.g. with good results in extracting relevant information about medical treatment (Vaci et al., 2020), classifying terms which can distinguish between cured depression cases and treatment-resistant depressions (Perlis et al., 2012), and accurate identification of information about patients' mood and affective states (Panaite et al., 2022). NLP-methods have also been applied to EHRs in order to evaluate drug response in patients, although not specifically in relation to depression (Bhatnagar et al., 2022). However, EHRs are a second-opinion resource, written by clinicians and expressing their professional objective rather than the patient perspective. Generally, it appears that research which applies NLP-methods with the aim of investigating the content of first-hand accounts from patients in relation to depression and treatment of depression is lacking. Detection-oriented approaches do in some cases relate to these objectives, but are not directly concerned with interpretation. One study has investigated drug reviews written by patients in

order to identify drugs with similar therapeutic value based on word representations, but again, this study is not specific to depression (Shiju & He, 2021).

Modelling approaches

To investigate patient perspectives, we leverage two methodological approaches based on the BERT-model architecture: Topic modelling supplemented with sentiment analysis.

BERT model architecture **(Liv)**

The BERT (Bidirectional Encoder Representations from Transformers) language representation model is a neural network-architecture developed in 2018 by a team from Google Research (Devlin et al., 2019). It has quickly become popular in the field of NLP, as it shows great performance, is generalizable to many different tasks without need for major task-specific modifications, and has low computational costs (Rogers et al., 2020). The BERT architecture is based on the transformer attention model developed by another Google Research team in 2017, Vaswani et al. (Vaswani et al., 2017). The transformer architecture revolutionised language modelling by utilising self-attention mechanisms exclusively to represent input- and output sequences. The self-attention mechanism computes attention scores for the interdependencies between input tokens, such as words in a sentence, making it possible to precisely encode token representations relative to relation to other tokens. Transformers utilise this to significantly reduce the amount of sequential processing compared to e.g. recurrent neural networks (Vaswani et al., 2017), resulting in high efficiency as well as a better ability to handle long-range interdependencies between tokens. In BERT, the attention architecture is leveraged to create an efficient language representation model. BERTs central advantage over other language representation models based on the transformer architecture (such as GPT) is that BERT is bidirectional, i.e. embeddings are computed in relation to dependencies in both directions; tokens (such as words) are thus represented not only according to what came before, but also according to what comes after, an important feature especially for sentence-level tasks (Devlin et al., 2019). As other language representation models, BERT is pre-trained on vast quantities of textual data to represent language as precisely as possible as vocabulary embeddings in an interconnected, multidimensional vector space. These embeddings underlie task-specific fine-tuning checkpoints

of the model. In fine-tuning for task-specific applications, the typical approach is to add connected layers on top of the pre-existing BERT encoding layers (Merchant et al., 2020). In the current analysis, we apply two fine-tuned BERT-based models: The BERTopic model designed for topic modelling and a fine-tuned checkpoint SieBERT of RoBERTa large for sentiment analysis.

BERTopic **(Eva)**

Topic modelling is a useful method when we want to understand what semantic structures are present in a set of documents, i.e. when we want to know what people talk about (Egger, 2022) Here, it is leveraged to explore which themes are commonly emerging amongst patients' reports of drug treatment effects. Topic modelling is especially suitable for large text corpora, where manual identification of semantic patterns would be demanding or entirely unfeasible. Before the advent of transformer-based models, topic model approaches typically applied a bag-of-words principle, why word co-occurrence relations were not considered in topic generation (Egger, 2022). This issue is overcome in the current model due to the BERT-foundation. BERTopic works through four primary steps (Grootendorst, 2022): First, by using a pre-trained embedding model to create embeddings of the given documents; second, by reducing the dimensionality of the embeddings; third, by clustering the embeddings into similar groups representing distinct topics; and fourth, by applying class-level TF-IDF to create topic representations for each cluster based on their difference from other clusters. A central asset of BERTopic is the split between document clustering and topic extraction. Because similar documents are grouped together based on their embeddings, and the TF-IDF procedure is applied at a cluster-level, the topics which are generated take into account the distinction between- and similarities within clusters (Egger & Yu, 2022). Overall, BERTopic shows good performance compared to other popular topic models such as LDA and Top2Vec, but it essential to note that evaluation of topic models is difficult as there are no simple accuracy metrics to compare; rather, evaluation of topic models are inherently dependent of human interpretation and expertise (Hannigan et al., 2019; Egger & Yu, 2022). In our case, BERTopic is a good choice because it allows us to assess diverse themes occurring across reviews and can provide novel insights into themes dependent on semantic context in each review (Egger & Yu, 2022).

Sentiment analysis **(Eva)**

Sentiment analysis is a computational method used to identify and quantify opinions in texts, usually as positive or negative attitudes (Habimana et al., 2019). In our case, we are interested in identification of sentiment in drug reviews according to their topic membership to facilitate a better understanding and deeper insight into patient attitudes about topics related to drug treatment. Here, binary sentiment analysis is applied for three reasons: 1) Binary sentiment is easy to interpret, 2) Binary sentiment classification models generally perform much better than more complex sentiment models (such as e.g. emotion detection models), and 3) Some topics are only represented in few reviews, why we do not have enough data to perform nuanced categorization.

Methods

Data **(Liv)**

The current study applies a dataset of drug reviews collected by Gräßer et al. (2018) from Drugs.com and Druglib.com retrieved from the Kaggle (UCI ML Drug Review Dataset). The dataset contains user reviews of medical drugs, and includes the name of the drug reviewed, the health condition the drug is used to treat, a satisfaction rating from 1-10 and anonymized user identification labels. We created a specific dataset containing only the entries relating to depression, specifically under the condition labels “depression” and “major depressive disorder”. Furthermore, we performed general data cleaning by excluding entries with NA’s and uninformative labels, excluding duplicates, and cleaning reviews for HTML-codes.

Data classification **(Eva)**

In the current study, we are interested in detecting patterns of semantic content and emotional sentiment in reviews in regards to two measures: Which drug class a drug belongs to, and which differing patterns occur for reviews that are rated positively or negatively within each drug class. The dataset consists of reviews of 102 different drugs, of which approximately half are identical due to the use of both generic (i.e. chemical classification) and brand names. Although the data set contained a variety of relevant drugs, we focus on the two most prominent classes in clinical use, as well as in the dataset, namely SSRIs and SNRIs. We labelled specific drug names as SSRI

and SNRI, respectively, and disregarded other entries relating to other drugs. For both classes, we further labelled whether each review had received a mostly negative satisfaction rating (from 1 to 5), or a mostly positive rating (from 6 to 10). Subsequently, the data included 5578 reviews in total, of which 3675 were for SSRIS and 1902 were for SNRIs. Amongst SSRI reviews, 2643 were mostly positive and 1032 mostly negative. Amongst SNRI reviews, 1289 were mostly positive and 614 were mostly negative.

Models **(Eva)**

To examine the patterns of sentiment and semantic content in reviews, we applied two different models based on the BERT-architecture. Firstly, to extract which themes occurred in each subset, we performed topic modelling using the model BERTopic (Grootendorst, 2022). Secondly, to extract measures of dominant attitudes within topics, we performed sentiment analysis using a fine-tuned instance of RoBERTa large, SieBERT, derived from Hugging Face (Hartmann et al., 2022).

BERTopic

Data wrangling **(Eva)**

Pilot analyses with BERTopic revealed that most generated topics centred around specific drug names. This is to be expected, as drug reviews refer to a specific drug often mentioned directly in the review. However, our interest lies in the semantic patterns which occur within drug classes rather than for individual drugs. Therefore, we substituted all mentions of drug names in the reviews with one generic word, “drug”. We chose this approach because it improves interpretability of the topics, aligns better with our goal of understanding the themes which occur in relation to the two different drug classes, and because substitution will generally not disrupt the semantic- and syntactic meaning in a sentence. An example of a review with substitution looks as the following: "I was on drug for two years. I started taking it after I had a panic attack due to extreme stress. It worked well for me. I tried drug for a few days [...]. drug was amazing.” Furthermore, we encountered a similar problem with frequent mentions of specific drug dosages. Again, as we are interested in class-level drugs, topics which refer to dosages of specific drugs

are not informative. Thus, we substituted all mentions of dosages (such as 50 mg) into a single label ("dosis").

BERTopic specifications **(Eva)**

For the analysis, we passed all depression-related reviews to the model to create a global topic model. Then, the topic representations for the defined categories were calculated, i.e. 1) for each drug class, and 2) for each drug class and satisfaction rating group. The model were specified as follows:

Embedding model. We used "all-MiniLM-L6-v2" from Hugging Face as the embedding model, an efficient sentence-transformer model designed to capture semantic information when encoding sentences and short paragraphs and specifically intended for clustering (Reimers & Gurevych, 2019).

Dimensionality reduction. We apply the UMAP algorithm (Uniform Manifold Approximation and Projection) which reduces the dimensions of any high-dimensional embedding model whilst still preserving central features from the original high-dimensional space (McInnes et al., 2020). UMAP decreases the distance between connected samples of the manifold (i.e. the embedding space), which is leveraged when clustering. We specified the number of neighbouring samples used for manifold approximation to 10, a relatively low value which enables a sufficiently local view for the lower sample size categories. As UMAP is stochastic, we specify a seed (15) for reproducibility.

Clustering. The clustering algorithm applied is HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise), which clusters high-density representations together whilst considering varying density structures, and allows for modelling low-density representations as noise (McInnes et al., 2017). HDBSCAN considers the distances between samples as derived by the UMAP algorithm and regards areas of proximal samples as high-density. We specified the minimum cluster size and the minimum number of samples to 5, meaning that to be considered a cluster, at least ten documents must be proximal in the embedding space, and smaller groupings are considered noise (and are thus not considered in the topic extraction).

Additional parameters. Additionally, we specify that the model should calculate probabilities for each topic; that the model should derive up to the top 10 words relevant in a topic; that the minimum topic size should be 10 documents; and that the generated number of topics should not exceed 15.

Vectorization. For vectorization, we apply the CountVectorizer algorithm from the sci-kit library (Pedregosa et al., 2011), which transforms the text inputs into vectors according to their frequency. Together with the c-TF-IDF calculation, these vectors define which topics are generated. We specify the n-gram range to (1, 2), allowing topic representations to consist of either one or two related words. We specify exclusion of stop-words, i.e. common words which do not carry relevant information. We specify a minimum frequency (how few appearances of a word is allowed) of 5, to allow most words but reduce the variability introduced by very rare words. To enable later-stage fine-tuning of these parameters, we apply the CountVectorizer after model training and update the topic representation calculations.

Sentiment analysis (Liv)

For sentiment analysis, we utilised the fine-tuned checkpoint SieBERT of the RoBERTa-large, a self-supervised language model based on the BERT architecture (Liu et al., 2019). SieBERT is trained on 15 diverse datasets, including reviews, and is highly generalizable across different text sources with an average accuracy of 93.2 % in predicting sentiment correctly. The model assigns a label for the most probable sentiment in each review. To obtain comparable estimates between topics, we calculated the percentages of sentiment labels (positive/negative) detected in each topic relative to the number of reviews assigned to that topic. The model was applied individually for each subset of reviews that were assigned to topics of relevance by the BERTopic analysis, resulting in percentage estimates for eleven topics (topic 1-6, topic 8-10, topic 12, and topic 14).

Results

Topic modelling **(Liv)**

For all data, we obtained fifteen topics after excluding topic -1, which generally in BERTopic analysis is uninformative. The topics are labelled according to their two most central words and numbered by size. Some topics appear to revolve around similar themes, which overall are the following: Sex (topic 1, topic 2), withdrawal (topic 3, topic 14), positive effects (topic 4, topic 5), anxiety (topic 5, topic 6, topic 9), insurance and drug brand (topic 8), and weight (topic 10). Topic 7, 11, and 13 appears uninterpretable, consisting of words such as "drug", "depression", "dosis", "effects", and "antidepressant". Topic 12 might concern sleep, but is difficult to interpret as well. In the appendix, table 2 and 3 includes measures of topic frequency (i.e. how many reviews belong to a certain topic), relative frequency (frequency relative to number of reviews in each category), and word representations for all generated topics.

Below, a visualisation of hierarchical clustering shows the relations between topics in terms of cosine distance. For example, this shows that topics related to withdrawal effects (topic 2 and 14) are closely related, whilst topic 3 regarding positive effects appears to relate globally to all other topics.

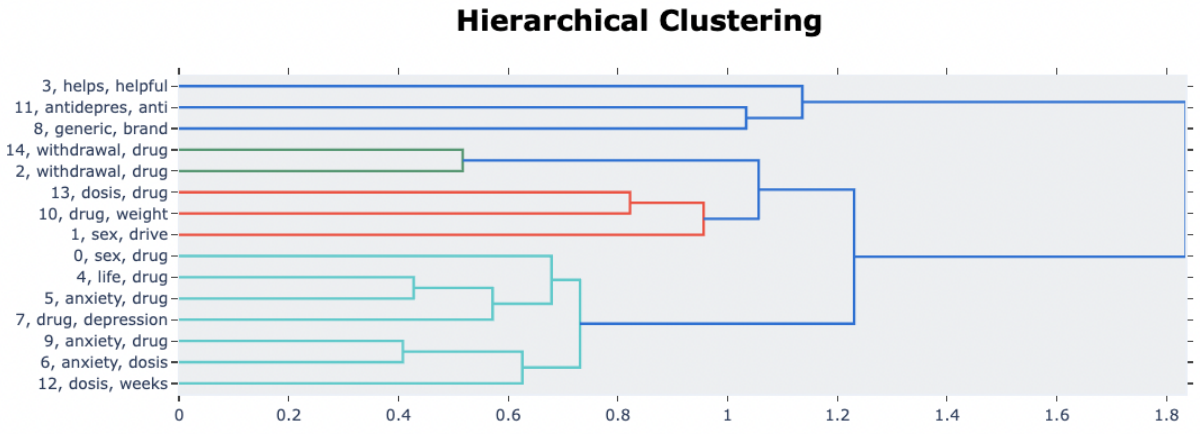

Figure 1: Intertopic relations as given by topic embedding cosine similarity.

Further, the top words occurring for each topic across all categories are extracted (*see Figure 2*). These can provide additional information about the topic profile. For instance, it shows how topic 8 with the top words generic and brand also involves the word "insurance", indicating that it relates to economic matters.

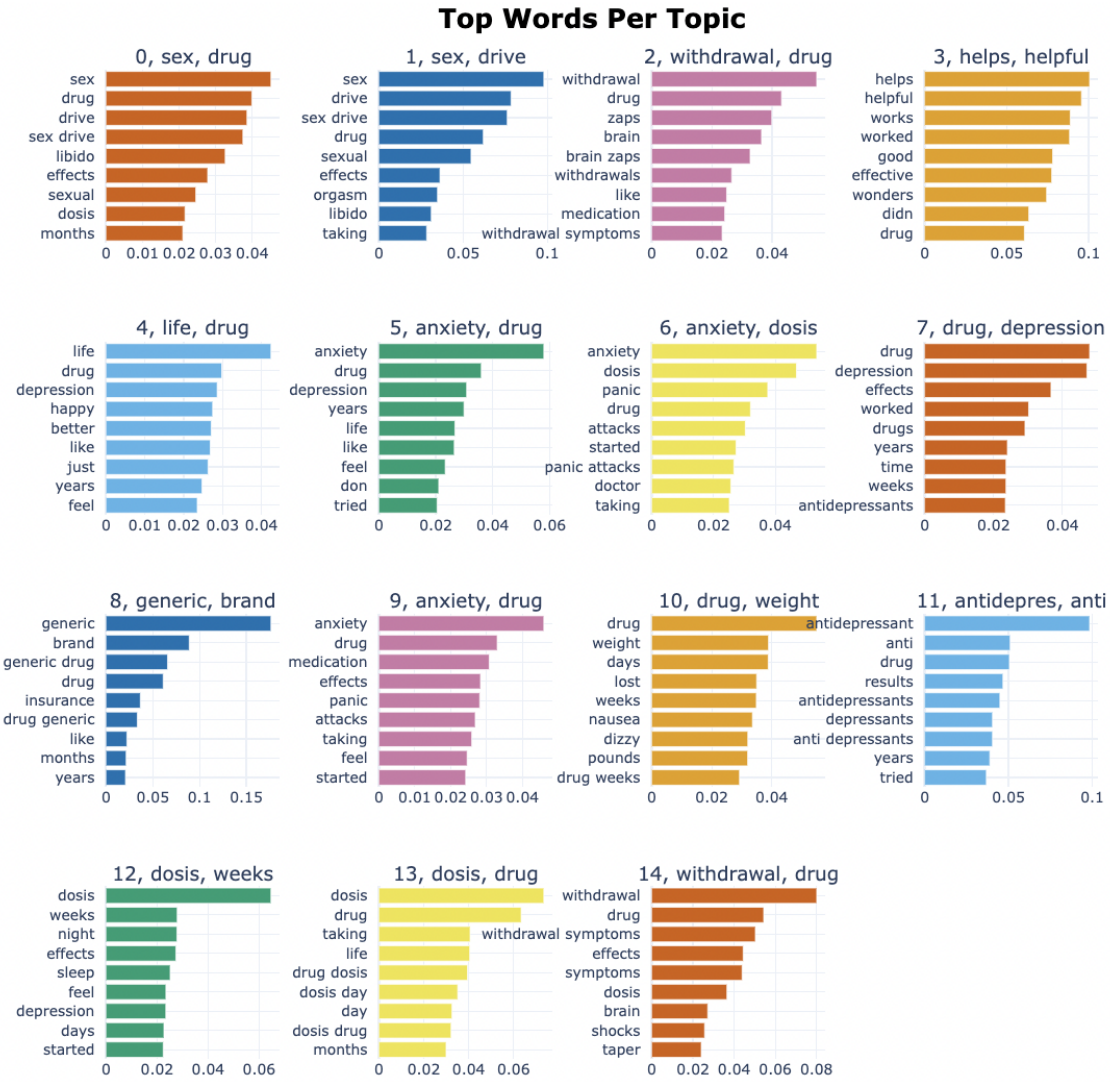

Figure 2: Top nine words extracted for each topic.

Category-wise representation **(Eva)**

In order to assess how the identified topics are represented in each drug class (*Figure 3*) and for different satisfaction ratings *(Figure 4)*, we address their relative frequency distributions (measured in percentage as topic frequency relative to the number of reviews per category.)

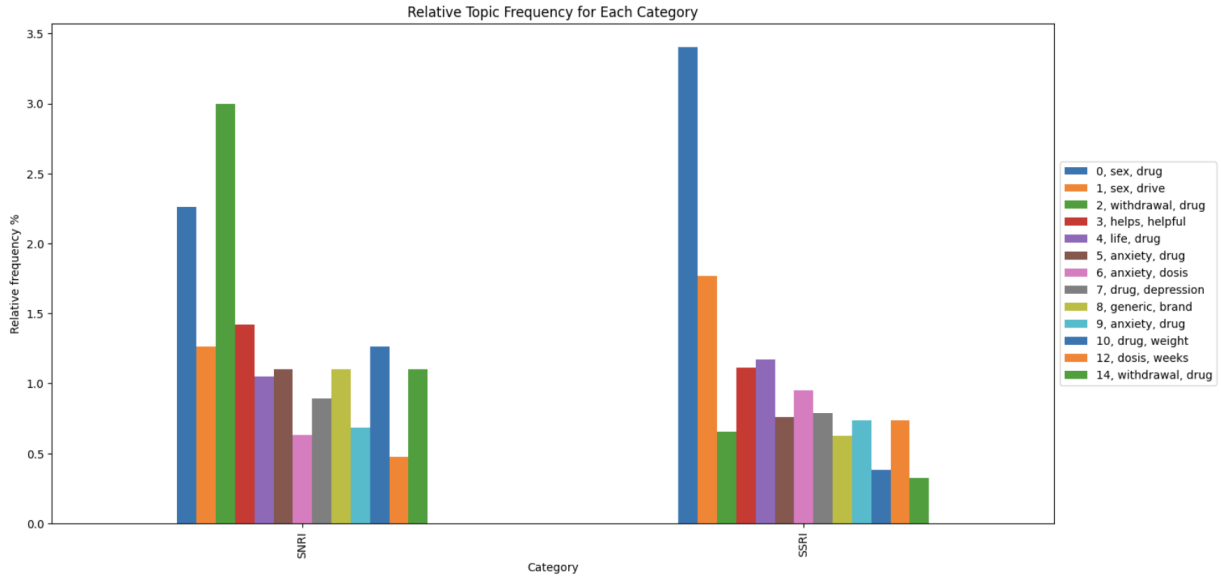

Figure 3: Topics represented by relative frequency for each drug class (SNRI / SSRI).

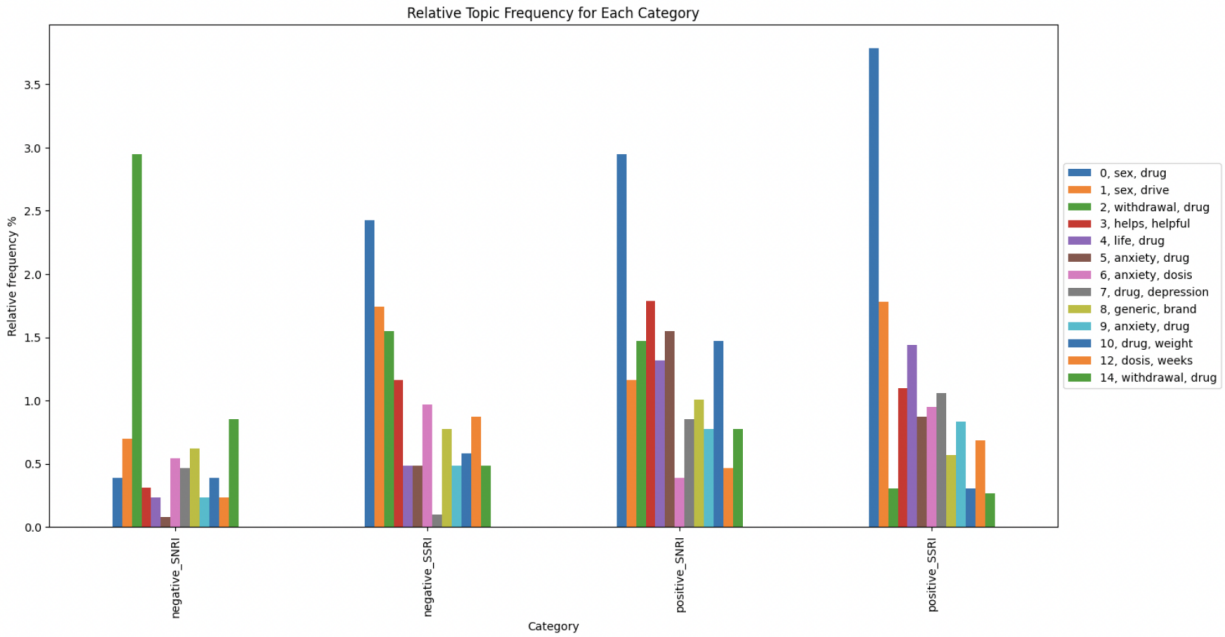

Figure 4: Topics represented by relative frequency for each drug class and rating level (positive / negative).

Sentiment analysis **(Eva)**

To obtain comparable measures of sentiment profiles for the different topics, the percentage of reviews assigned to specific topics is shown in table 1.

Topic nr.	Topic label and interpretation	Positive (%)	Negative (%)
0	Sex, drug (sex)	70.8	29.2
1	Sex, drive (sex)	49.4	50.6
2	Withdrawal, drug (withdrawal)	30.9	69.1
3	Help, helpful (positive effects)	73.5	26.5
4	Life, drug (positive effects)	88.9	11.1
5	Anxiety, drug (anxiety)	83.7	16.3
6	Anxiety, dosis (anxiety)	63.8	36.2
8	Generic, brand (insurance and brand)	45.5	54.5
9	Anxiety, drug (anxiety)	77.5	22.5
10	Drug, weight (weight)	57.9	42.1
12	Dosis, weeks (sleep)	72.2	27.8
14	Withdrawal, drug (withdrawal)	45.5	54.5

Table 1: Percentages of positive and negative sentiment labels assigned to reviews relating to specific topics.

Discussion

(Liv)

The first main question posed in this paper regards what semantic themes and attitudes occur in drug reviews for SSRIs and SNRIs used as depression treatment, and whether these differ according to patients' satisfaction rates. We applied transformer-based topic modelling and sentiment analysis to explore this question. The first thing to note is that no topics are particularly frequent. This might indicate that the reviews are highly variable with no major themes, or indicate methodological issues that we will consider subsequently. However, we derive eleven out of fifteen generated topics which are of clear interpretive value. Of these, some

topics are similar, leaving six clearly distinct themes: Sex (topic 1, topic 2), withdrawal (topic 3, topic 14), positive effects (topic 4, topic 5), anxiety (topic 5, topic 6, topic 9), insurance and drug brand (topic 8) and weight (topic 10). The topics relating to sex, anxiety, weight, and sleep overlaps with well-known side effects to SSRI- and SNRI use. The topics on withdrawal effects are clearly related to symptoms related to discontinuation or termination, with "withdrawal" and "symptoms" repeatedly represented. Further, words relating to the specific symptom "brain zaps" are preeminent. The topic on positive effects appears to concern successful treatment, with words such as "work", "good", "effective", "wonders", "happy" and "better". Lastly, one topic seems to concern matters of insurance and drug brands, which might suggest that the economic aspect of pharmacological treatment is something that concerns patients.

To explore the differences in how the generated topics are distributed for the two drug classes, as well as for their corresponding satisfaction levels, we assessed a measure of relative frequency of topic representation within each category. A central note for all interpretations of topics according to satisfaction rating is that even though some topics occur more frequently for positively- or negatively rated reviews, this does not suggest anything decisive on whether reference to that topic is positive or negative. Rather, it makes it possible to provide an indication to whether the topic is related to the overall satisfaction level with the drug. Conversely, the sentiment analysis does contribute with a direct measure of attitude within reviews attributed specific topics; however, not according to category. Here, we highlight a few topics. The topics of highest frequency overall are those relating to sex drive (topic 1 and 2). These topics appear more prevalent for SSRIs than for SNRIs, which could indicate that patients experience more alterations of their sex life when taking SSRIs than SNRIs. Further, the topics are more prevalent in positive- than negative reviews for both drug categories, possibly indicating a positive relationship between mentioning sex and drug treatment in the majority of cases where sex is mentioned. This notion is somewhat supported by the sentiment scores for these topics, where 70.8 % of reviews in topic 0 are positive, whilst it is 50/50 for topic 1. Thus, there appears to be a notable prevalence of negative sentiment in reviews related to sex, as well. Overall, the relatively high prevalence of this topic indicates that sex related effects are important for patients.

The two topics related to withdrawal effects (topic 2 and 14) are clearly most prominent amongst SNRI reviews, in both cases three- to four times that of SSRI reviews. Symptoms related to discontinuation, especially when treatment is ended abruptly, are as previously presented well documented for both drug classes. To our knowledge, there is no evidence that SNRIs should cause worse discontinuation effects than SSRIs, but this matter is not sufficiently described in the literature (e.g. Fava et al., 2018). It is however documented that the neurological phenomenon of “brain zaps" is significantly more prevalent for SNRI withdrawal (Stockmann et al., 2018). This is consistent with the higher prevalence of topic 2 in SNRIs, where the word representation shows that “brain zaps” is a prominent word in this topic. Discontinuation is generally reported as an uncomfortable experience (Hosenbocus & Chahal, 2011), and the relatively high percentage of negative sentiment for especially topic 2 (69.1 %) seems to suggest that especially the withdrawal symptom “brain zaps” is related to an overall negative attitude in the reviews. Further supporting this notion, the withdrawal-related topics are approximately twice as prevalent in negative reviews for both drug categories, indicating that withdrawal symptoms might influence the overall satisfaction with the treatment.

Lastly, the topics relating to anxiety (topic 5, 6 and 9) are worth mentioning. Although all concerns feelings of anxiety, they appear to be differently embedded. For instance, topic 6 seems to relate especially to feelings of panic and panic attacks, and is equally distributed between positive- and negative SSRI reviews, whilst more common in negative- than positive SNRI reviews. Topic 5 and 9 seems to relate more generally to anxiety, and are both significantly more prevalent in positive- than negative reviews for both drug categories. Together, this could indicate that whilst some patients experience increased levels of anxiety and panic attacks in response to treatment, others might experience a decrease in their feelings of anxiety. However, sentiment analysis indicates that reviews related to anxiety are generally positive, although least so for the panic-related topic 6 (63.8 % positive) compared to topic 5 and 9 (83.7 % and 77.5 %). This might indicate that the relation between drug use and anxiety is mostly positive, which is sensible considering that antidepressants are used for anxiety diagnoses, as well (Hoffman & Mathew, 2008). Relatedly, it is important to note that we do not have any information about additional anxiety diagnoses, why we are not able to address whether patients refer to a diagnosed condition or generally anxious feelings.

The second question posed in this paper concerned whether NLP, and particularly topic modelling, can provide relevant, informative indicators about how patients experience drug treatment. As presented in the previous paragraphs, several trends which appear in touch with the literature were indeed identified in our analysis, although not of particularly high detection frequency. However, our results might suggest that topic modelling can be used to identify theoretically sound patterns in large-scale, free-text data. When applied in an unsupervised manner as in the current study, the approach appears mainly useful for providing a rough overview of inherent themes in the reviews, but also to be able to provide novel insight into topics that might not be expected. It is for instance interesting to note that the topics which appear to relate to side effects are generally concerning themes that often matter a lot to people, such as sex and weight, rather than concrete side effects such as dry mouth or increased blood pressure. However, the difficulties in assessing the context in which topics occur poses some constraints regarding implementation of the method. For instance, we cannot assess whether topics such as those relating to side effects occur because patients experience side effects, or because they mention that they do not (this an important issue in BERT-models generally, as they have been shown to struggle with representing negations correctly) (Ettinger, 2020). It should be clear that the method presented is mostly useful to guide and propose subjects relevant for further research efforts in a time-efficient manner. It is also relevant to address that the model might not capture topics that could be important, e.g. if the topic is more local than model specification allows for, or if a topic which appears coherent for a human is not embedded as coherent for the model. A general issue with this approach is thus that transparency is compromised, and that interpretation still remains a matter of qualitative human assessment.

Methodological constraints and prospects for future research **(Eva)**

There are a number of constraints related to the methodology applied. Firstly, there are central limitations to the specific dataset. Depression is a complex disorder, and treatment efficiency is generally influenced by many different factors that are difficult to consider and control for. This difficulty arises for example due to a lack of additional information in the data, such as demographics, diagnostic history, and additional medical- or non-medical treatment.

For the models applied, there are important technical- as well as general limitations. As mentioned, when using BERTopic, human judgement has a central influence upon both parameter tuning and interpretation of final results. Specification of parameters such as cluster size, number of desired topics, and minimum topic size are manually tweaked depending on subjective assessment of the model output. As our sample sizes for the categories of interest are not very big seen from an NLP-perspective (down to 614 entries in negative SNRI reviews), many of these parameters had to be specified to a rather local view in order to obtain results. Thus, some of the detected topics are small and quite specific to a few reviews, and it introduces a risk of generating topics which should have been considered noise (Egger & Yu, 2022). Relatedly, as we have manually specified a maximum number of topics for the model to generate, rather than allowing automatic topic reduction using HDBSCAN, we might introduce some forced topic merging of unrelated topics (ref). However, as we find several topics relating to similar themes, this does not appear to be an issue in the current analysis. Additionally, the clustering algorithm applied (UMAP) is stochastic, introducing a level of randomness into topic generation. The influence of this randomness increases when smaller clusters are included, like in the current analysis (UMAP 0.5 Documentation, n.d.). Using a different seed might therefore yield quite different results. Finally, as mentioned, interpretation of the final results depends on human judgement as well. This is true both in terms of deciphering interpretative value of the derived topics, and in terms of requiring theoretical domain knowledge to understand and apply the relevant context for the topics (Egger and Yu, 2022; Nikolenko et al., 2017). Here, none of the writers are psychiatric professionals, so interpretation of the topics are likely flawed due to a lack of expertise.

Moreover, category-wise topic representation in BERTopic is a novel feature which still lacks some functionality. For instance, only a single visualisation is available for class-level information, and there is no built-in interpretable relative frequency measure available to compare topic representation across categories. Moreover, it is not transparent how the category-level topics are generated. As the technique generates topics for the full data before assigning defined categories into these topics, it is not clear whether important category-specific topics are even captured.

Lastly, the most important limitation is that we have implemented direct alterations in the natural language data by substituting drug names and dosage specifications with “drug” and “dosis”. This is normally discouraged, both because it changes the original structure of the language and thus affects the extraction of latent semantic information, and because it reduces the generalizability of the approach. However, as previously stated, meaningful topics could not be extracted without these alterations. This is potentially a shortcoming of applying an transformer-based approach to perform topic modelling in certain cases. As the attention mechanisms attract the topic generation towards the central object of the reviews, topics will intuitively center around drug names, although these are not informative for the current intent. This also explains why we see the substitution words “drug” and “dosis” in many of the generated topics. Thus, it is relevant to consider whether our case could have benefited from applying a different approach than BERTopic, such as a more conventional approach based on a bag-of-words term frequency, or an approach where non-informative words could be pre-specified.

Future studies could focus on improving the current analysis by taking the above-presented limitations into account. Furthermore, we compare two drug categories that are quite similar in how they function; future research could assess how which topics arise and how they distribute over more distinct treatment options, whether that are for different drug categories, or entirely alternative treatments such as psychedelics or therapy. Such an approach might further contribute with a better foundation to validate the applicability of topic modelling for addressing central themes in patient experience of treatment, as topics of a more distinct nature would likely occur. Future studies should also consider how comparison of results are conducted, and how to control for confounding variables. From the current analysis, there are many possible research paths that could be taken to improve or elaborate upon the approach.

Conclusion

The study at hand aimed to explore two primary questions: Which semantic themes and corresponding sentiment occur in drug reviews for SSRIs and SNRIs, and how these measures differ according to overall satisfaction rate, and whether NLP-methods and specifically topic

modelling can be used to derive informative indicators about how different types of medical drugs used to treat depression are experienced by patients. Overall, we find a number of topics which appear relevant to the experience of the drug treatments investigated. We find that topics are distributed differently for the SSRIs and SNRIs and for different ratings, and that these distributions correspond well to the literature. Further, we find that sentiment differs in an apparently meaningful way for many of the generated topics. In conclusion, we find that NLP proves promising as a method to explore which semantic patterns are prevalent in free-text sources relating to drug treatment, and might be useful for pointing towards important aspects of patient experience of drug effects that should be addressed.

Bibliography:

Bhatnagar, R., Sardar, S., Beheshti, M., & Podichetty, J. T. (2022). How can natural language processing help model informed drug development?: A review. *JAMIA Open*, *5*(2), ooac043. https://doi.org/10.1093/jamiaopen/ooac043

Bschor, T., & Kilarski, L. L. (2016). Are antidepressants effective? A debate on their efficacy for the treatment of major depression in adults. *Expert Review of Neurotherapeutics*, *16*(4), 367–374. https://doi.org/10.1586/14737175.2016.1155985

Bumpus, J. A. (2020). Low-Dose Tramadol as an Off-Label Antidepressant: A Data Mining Analysis from the Patients' Perspective. *ACS Pharmacology & Translational Science*, *3*(6), 1293–1303. https://doi.org/10.1021/acsptsci.0c00132

Cartwright, C., Gibson, K., Read, J., Cowan, O., & Dehar, T. (2016). Long-term antidepressant use: Patient perspectives of benefits and adverse effects. *Patient Preference and Adherence*, *10*, 1401–1407. https://doi.org/10.2147/PPA.S110632

Cipriani, A., Furukawa, T. A., Salanti, G., Chaimani, A., Atkinson, L. Z., Ogawa, Y., Leucht, S., Ruhe, H. G., Turner, E. H., Higgins, J. P. T., Egger, M., Takeshima, N., Hayasaka, Y., Imai, H., Shinohara, K., Tajika, A., Ioannidis, J. P. A., & Geddes, J. R. (2018). Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: A systematic review and network meta-analysis. *Lancet (London, England)*, *391*(10128), 1357–1366. https://doi.org/10.1016/S0140-6736(17)32802-7

Depression. (n.d.). Retrieved 6 January 2023, from https://www.who.int/news-room/fact-sheets/detail/depression

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding* (arXiv:1810.04805). arXiv. http://arxiv.org/abs/1810.04805

Edinoff, A. N., Akuly, H. A., Hanna, T. A., Ochoa, C. O., Patti, S. J., Ghaffar, Y. A., Kaye, A. D., Viswanath, O., Urits, I., Boyer, A. G., Cornett, E. M., & Kaye, A. M. (2021). Selective Serotonin Reuptake Inhibitors and Adverse Effects: A Narrative Review. *Neurology International*, *13*(3), Article 3. https://doi.org/10.3390/neurolint13030038

Egger, R. (2022). Topic Modelling. In R. Egger (Ed.), *Applied Data Science in Tourism: Interdisciplinary Approaches, Methodologies, and Applications* (pp. 375–403). Springer International Publishing. https://doi.org/10.1007/978-3-030-88389-8_18

Egger, R., & Yu, J. (2022). A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology*, *7*, 886498. https://doi.org/10.3389/fsoc.2022.886498

Eplov, Jørgensen, Hougaard, Kessing, Larsen, Stage, Videbech. (2005). *Behandling af depressionssygdomme. Kommentering af: Behandling af depressionssjukdomar. En systematisk litteraturoversikt.* København: Sundhedsstyrelsen, Center for Evaluering og Medicinsk Teknologivurdering.

Ettinger, A. (2020). *What BERT is not: Lessons from a new suite of psycholinguistic diagnostics for language models* (arXiv:1907.13528). arXiv. https://doi.org/10.48550/arXiv.1907.13528

Fanelli, D., Weller, G., & Liu, H. (2021). New Serotonin-Norepinephrine Reuptake Inhibitors and Their Anesthetic and Analgesic Considerations. *Neurology International*, *13*(4), 497–509. https://doi.org/10.3390/neurolint13040049

Fava, G. A., Benasi, G., Lucente, M., Offidani, E., Cosci, F., & Guidi, J. (2018). Withdrawal Symptoms after Serotonin-Noradrenaline Reuptake Inhibitor Discontinuation: Systematic Review. *Psychotherapy and Psychosomatics*, *87*(4), 195–203. https://doi.org/10.1159/000491524

Fava, G. A., Gatti, A., Belaise, C., Guidi, J., & Offidani, E. (2015). Withdrawal Symptoms after Selective Serotonin Reuptake Inhibitor Discontinuation: A Systematic Review. *Psychotherapy and Psychosomatics*, *84*(2), 72–81. https://doi.org/10.1159/000370338

Goodwin, G. M., Price, J., De Bodinat, C., & Laredo, J. (2017). Emotional blunting with antidepressant treatments: A survey among depressed patients. *Journal of Affective Disorders*, *221*, 31–35. https://doi.org/10.1016/j.jad.2017.05.048

Gräßer, F., Kallumadi, S., Malberg, H., & Zaunseder, S. (2018). Aspect-Based Sentiment Analysis of Drug Reviews Applying Cross-Domain and Cross-Data Learning. *Proceedings of the 2018 International Conference on Digital Health*, 121–125. https://doi.org/10.1145/3194658.3194677

Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*

(arXiv:2203.05794). arXiv. https://doi.org/10.48550/arXiv.2203.05794
Habimana, O., Li, Y., Li, R., Gu, X., & Yu, G. (2019). Sentiment analysis using deep learning approaches: An overview. *Science China Information Sciences*, *63*(1), 111102. https://doi.org/10.1007/s11432-018-9941-6
Hannigan, T. R., Haans, R. F. J., Vakili, K., Tchalian, H., Glaser, V. L., Wang, M. S., Kaplan, S., & Jennings, P. D. (2019). Topic Modeling in Management Research: Rendering New Theory from Textual Data. *Academy of Management Annals*, *13*(2), 586–632. https://doi.org/10.5465/annals.2017.0099
Hartmann, J., Heitmann, M., Siebert, C., & Schamp, C. (2022). More than a Feeling: Accuracy and Application of Sentiment Analysis. *International Journal of Research in Marketing*. https://doi.org/10.1016/j.ijresmar.2022.05.005
Hoffman, E. J., & Mathew, S. J. (2008). Anxiety disorders: A comprehensive review of pharmacotherapies. *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine*, *75*(3), 248–262. https://doi.org/10.1002/msj.20041
Hosenbocus, S., & Chahal, R. (2011). SSRIs and SNRIs: A review of the Discontinuation Syndrome in Children and Adolescents. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, *20*(1), 60–67.
Hung, C.-I. (2014). Factors predicting adherence to antidepressant treatment. *Current Opinion in Psychiatry*, *27*(5), 344–349. https://doi.org/10.1097/YCO.0000000000000086
Johnston, B. (2013). *The Role of Patient Experience and its Influence on Adherence to Antidepressant Treatment*. Journal of Psychosocial Nursing and Mental Health Services. https://journals.healio.com/doi/10.3928/02793695-20130930-04
Kirsch, I. (2014). Antidepressants and the Placebo Effect. *Zeitschrift Für Psychologie*, *222*(3), 128–134. https://doi.org/10.1027/2151-2604/a000176
Kirsch, I. (2019). Placebo Effect in the Treatment of Depression and Anxiety. *Frontiers in Psychiatry*, *10*. https://www.frontiersin.org/articles/10.3389/fpsyt.2019.00407
Köhler, S., Cierpinsky, K., Kronenberg, G., & Adli, M. (2016). The serotonergic system in the neurobiology of depression: Relevance for novel antidepressants. *Journal of Psychopharmacology (Oxford, England)*, *30*(1), 13–22. https://doi.org/10.1177/0269881115609072
Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). *RoBERTa: A Robustly Optimized BERT Pretraining Approach* (arXiv:1907.11692). arXiv. https://doi.org/10.48550/arXiv.1907.11692
Mathew, S. J., & Charney, D. S. (2009). Publication Bias and the Efficacy of Antidepressants. *American Journal of Psychiatry*, *166*(2), 140–145. https://doi.org/10.1176/appi.ajp.2008.08071102
May antidepressant drugs worsen the conditions they are supposed to treat? The clinical foundations of the oppositional model of tolerance—Giovanni A. Fava, 2020. (n.d.). Retrieved 6 January 2023, from https://journals.sagepub.com/doi/full/10.1177/2045125320970325
McAllister-Williams, R. H., Arango, C., Blier, P., Demyttenaere, K., Falkai, P., Gorwood, P., Hopwood, M., Javed, A., Kasper, S., Malhi, G. S., Soares, J. C., Vieta, E., Young, A. H., Papadopoulos, A., & Rush, A. J. (2020). The identification, assessment and management of difficult-to-treat depression: An international consensus statement. *Journal of Affective Disorders*, *267*, 264–282. https://doi.org/10.1016/j.jad.2020.02.023
McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *Journal of Open Source Software*, *2*(11), 205. https://doi.org/10.21105/joss.00205
McInnes, L., Healy, J., & Melville, J. (2020). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction* (arXiv:1802.03426). arXiv. https://doi.org/10.48550/arXiv.1802.03426
Merchant, A., Rahimtoroghi, E., Pavlick, E., & Tenney, I. (2020). *What Happens To BERT Embeddings During Fine-tuning?* (arXiv:2004.14448). arXiv. http://arxiv.org/abs/2004.14448
Moncrieff, J., Cooper, R. E., Stockmann, T., Amendola, S., Hengartner, M. P., & Horowitz, M. A. (2022). The serotonin theory of depression: A systematic umbrella review of the evidence. *Molecular Psychiatry*, 1–14. https://doi.org/10.1038/s41380-022-01661-0
Moret, C., & Briley, M. (2011). The importance of norepinephrine in depression. *Neuropsychiatric Disease and Treatment*, *7*(Suppl 1), 9–13. https://doi.org/10.2147/NDT.S19619
Nanomi Arachchige, I. A., Sandanapitchai, P., & Weerasinghe, R. (2021). Investigating Machine Learning & Natural Language Processing Techniques Applied for Predicting Depression Disorder from Online Support Forums: A Systematic Literature Review. *Information*, *12*(11), Article 11. https://doi.org/10.3390/info12110444
Nguyen, T., Phung, D., Dao, B., Venkatesh, S., & Berk, M. (2014). Affective and Content Analysis of Online

Depression Communities. *IEEE Transactions on Affective Computing*, *5*(3), 217–226. https://doi.org/10.1109/TAFFC.2014.2315623

Nikolenko, S. I., Koltcov, S., & Koltsova, O. (2017). Topic modelling for qualitative studies. *Journal of Information Science*, *43*(1), 88–102. https://doi.org/10.1177/0165551515617393

Panaite, V., Devendorf, A. R., Finch, D., Bouayad, L., Luther, S. L., & Schultz, S. K. (2022). The Value of Extracting Clinician-Recorded Affect for Advancing Clinical Research on Depression: Proof-of-Concept Study Applying Natural Language Processing to Electronic Health Records. *JMIR Formative Research*, *6*(5), e34436. https://doi.org/10.2196/34436

Park, L. T., & Zarate, C. A. (2019). Depression in the Primary Care Setting. *New England Journal of Medicine*, *380*(6), 559–568. https://doi.org/10.1056/NEJMcp1712493

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*(85), 2825–2830.

Perlis, R. H., Iosifescu, D. V., Castro, V. M., Murphy, S. N., Gainer, V. S., Minnier, J., Cai, T., Goryachev, S., Zeng, Q., Gallagher, P. J., Fava, M., Weilburg, J. B., Churchill, S. E., Kohane, I. S., & Smoller, J. W. (2012). Using electronic medical records to enable large-scale studies in psychiatry: Treatment resistant depression as a model. *Psychological Medicine*, *42*(1), 41–50. https://doi.org/10.1017/S0033291711000997

Reimers, N., & Gurevych, I. (2019). *Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks* (arXiv:1908.10084). arXiv. https://doi.org/10.48550/arXiv.1908.10084

Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). A Primer in BERTology: What We Know About How BERT Works. *Transactions of the Association for Computational Linguistics*, *8*, 842–866. https://doi.org/10.1162/tacl_a_00349

Ruberto, V. L., Jha, M. K., & Murrough, J. W. (2020). Pharmacological Treatments for Patients with Treatment-Resistant Depression. *Pharmaceuticals*, *13*(6), Article 6. https://doi.org/10.3390/ph13060116

Rutherford, B. R., Wall, M. M., Brown, P. J., Choo, T.-H., Wager, T. D., Peterson, B. S., Chung, S., Kirsch, I., & Roose, S. P. (2017). Patient Expectancy as a Mediator of Placebo Effects in Antidepressant Clinical Trials. *American Journal of Psychiatry*, *174*(2), 135–142. https://doi.org/10.1176/appi.ajp.2016.16020225

Santarsieri, D., & Schwartz, T. L. (2015). Antidepressant efficacy and side-effect burden: A quick guide for clinicians. *Drugs in Context*, *4*, 212290. https://doi.org/10.7573/dic.212290

Sartorius, N. (2001). The economic and social burden of depression. *The Journal of Clinical Psychiatry*, *62 Suppl 15*, 8–11.

Shiju, A., & He, Z. (2021). *Classifying Drug Ratings Using User Reviews with Transformer-Based Language Models* (p. 2021.04.15.21255573). medRxiv. https://doi.org/10.1101/2021.04.15.21255573

Stockmann, T., Odegbaro, D., Timimi, S., & Moncrieff, J. (2018). SSRI and SNRI withdrawal symptoms reported on an internet forum. *International Journal of Risk & Safety in Medicine*, *29*(3–4), 175–180. https://doi.org/10.3233/JRS-180018

Sundhedsdatastyrelsen—Statistikker. (n.d.). Retrieved 6 January 2023, from https://medstat.dk/da

Tadesse, M. M., Lin, H., Xu, B., & Yang, L. (2019). Detection of Depression-Related Posts in Reddit Social Media Forum. *IEEE Access*, *7*, 44883–44893. https://doi.org/10.1109/ACCESS.2019.2909180

Tian, H., Hu, Z., Xu, J., & Wang, C. (2022). The molecular pathophysiology of depression and the new therapeutics. *MedComm*, *3*(3), e156. https://doi.org/10.1002/mco2.156

UCI ML Drug Review dataset. (n.d.). Retrieved 6 January 2023, from https://www.kaggle.com/datasets/jessicali9530/kuc-hackathon-winter-2018

Uher, J. (2018). Quantitative Data From Rating Scales: An Epistemological and Methodological Enquiry. *Frontiers in Psychology*, *9*. https://www.frontiersin.org/articles/10.3389/fpsyg.2018.02599

UMAP Reproducibility—Umap 0.5 documentation. (n.d.). Retrieved 6 January 2023, from https://umap-learn.readthedocs.io/en/latest/reproducibility.html

Vaci, N., Liu, Q., Kormilitzin, A., De Crescenzo, F., Kurtulmus, A., Harvey, J., O'Dell, B., Innocent, S., Tomlinson, A., Cipriani, A., & Nevado-Holgado, A. (2020). Natural language processing for structuring clinical text data on depression using UK-CRIS. *Evidence Based Mental Health*, *23*(1), 21–26. https://doi.org/10.1136/ebmental-2019-300134

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need* (arXiv:1706.03762). arXiv. http://arxiv.org/abs/1706.03762

Wang, S.-M., Han, C., Bahk, W.-M., Lee, S.-J., Patkar, A. A., Masand, P. S., & Pae, C.-U. (2018). Addressing the Side Effects of Contemporary Antidepressant Drugs: A Comprehensive Review. *Chonnam Medical Journal*, *54*(2), 101. https://doi.org/10.4068/cmj.2018.54.2.101

Wiles, N., Taylor, A., Turner, N., Barnes, M., Campbell, J., Lewis, G., Morrison, J., Peters, T. J., Thomas, L., Turner, K., & Kessler, D. (2018). Management of treatment-resistant depression in primary care: A mixed-methods study. *The British Journal of General Practice*, *68*(675), e673–e681. https://doi.org/10.3399/bjgp18X699053

Zhang, T., Schoene, A. M., Ji, S., & Ananiadou, S. (2022). Natural language processing applied to mental illness detection: A narrative review. *Npj Digital Medicine*, *5*(1), Article 1. https://doi.org/10.1038/s41746-022-00589-7

Appendix

Link to Github: Including data, data cleaning scripts, topic modelling analysis and sentiment analysis

https://github.com/evasahlholdt/NLP_exam_2023_Eva-Liv

Additional results

Table 2. Topic modelling results, distributed in SSRIs and SNRIs.

Topic nr.	Topic label	Representative words	Category	Frequency	Relative frequency %
0	Sex, drug	Sex, drug, drive, sex drive, libido	SSRI	125	3.4
0	Sex, drug	Sex, drug, drive, libido, sex drive	SNRI	43	2.26
1	Sex, drive	Sex, drive, sex drive, drug, sexual	SSRI	65	1.77
1	Sex, drive	Sex, sexual, drug, drive, sex drive	SNRI	24	1.26
2	Withdrawal, drug	Withdrawal, zaps, drug, brain, withdrawals	SSRI	24	0.65
2	Withdrawal, drug	Withdrawal, drug, zaps, brain, brain zaps	SNRI	57	3.0
3	Helps, helpful	Helps, works, helpful, effective, worked	SSRI	41	1.12
3	Helps, helpful	Helps, works, helpful, wonders, good	SNRI	27	1.42
4	Life, drug	Life, happy, better, depression, drug	SSRI	43	1.17
4	Life, drug	Life, drug, just, depression, year	SNRI	20	1.05
5	Anxiety, drug	Anxiety, drug, depression, years, life	SSRI	28	0.76
5	Anxiety, drug	Anxiety, drug, years, depression, life	SNRI	21	1.1
6	Anxiety, dosis	Anxiety, dosis, panic, attacks, drug	SSRI	35	0.95
6	Anxiety, dosis	Anxiety, dosis, drug, panic, attacks	SNRI	12	0.63
7	Drug, depression	Depression, drug, effects, drugs, worked	SSRI	29	0.79
7	Drug, depression	Drug, depression, effects, worked, antidepressants	SNRI	17	0.89
8	Generic, brand	Generic, brand, generic drug, drug, drug generic	SSRI	23	0.63

8	Generic, brand	Generic, brand, generic drug, drug, insurance	SNRI	21	1.1
9	Anxiety, drug	Anxiety, effects, drug, medication, panic	SSRI	27	0.73
9	Anxiety, drug	Anxiety, drug, medication, like, panic	SNRI	13	0.68
10	Drug, weight	Drug, weeks, weight, days, hair	SSRI	14	0.38
10	Drug, weight	Drug, nausea, days, weight, lost	SNRI	24	1.26
11	Antidepres, anti	Antidepressant, drug, anti, antidepressant, results	SSRI	17	0.46
11	Antidepres, anti	Antidepressant, anti, drug, results, antidepressants	SNRI	19	1.0
12	Dosis, weeks	Dosis, effects, night, weeks, feel	SSRI	27	0.73
12	Dosis, weeks	Dosis, sleep, weeks, night, started	SNRI	9	0.47
13	Dosis, drug	Dosis, drug, life, taking, drug dosis	SSRI	27	0.73
13	Dosis, drug	Dosis, drug, dosis day, snappy, drug dosis	SNRI	8	0.42
14	Withdrawal, drug	Withdrawal, drug, effects, withdrawal symptoms, dosis	SSRI	12	0.33
14	Withdrawal, drug	Withdrawal, withdrawal symptoms, symptoms, drug, effects	SNRI	21	1.1

Table 3. Topic modelling results, distributed in SSRIs and SNRIs and positive/negative ratings.

Topic	Words	Freque...	Class	Label	Relative Frequency
-1	drug, dosis, depression, day, taking	507	negative_SNRI	-1, drug, dosis	39.3328161365
-1	drug, dosis, depression, feel, effects	1060	positive_SNRI	-1, drug, dosis	82.2342901474
-1	drug, dosis, depression, feel, anxiety	2240	positive_SSRI	-1, drug, dosis	84.7521755581
-1	drug, dosis, depression, taking, feel	898	negative_SSRI	-1, drug, dosis	87.015503876
0	sex, drive, drug, sex drive, years	5	negative_SNRI	0, sex, drug	0.387897595
0	sex, drive, drug, sex drive, effects	25	negative_SSRI	0, sex, drug	2.4224806202
0	sex, libido, drug, drive, sex drive	38	positive_SNRI	0, sex, drug	2.9480217223
0	sex, drug, drive, sex drive, libido	100	positive_SSRI	0, sex, drug	3.783579266
1	sex, sexual, drug, drive, sex drive	9	negative_SNRI	1, sex, drive	0.6982156711
1	sex, drug, sex drive, drive, sexual	15	positive_SNRI	1, sex, drive	1.1636927851
1	sex, drug, drive, sex drive, sexual	18	negative_SSRI	1, sex, drive	1.7441860465
1	sex, drive, sex drive, drug, sexual	47	positive_SSRI	1, sex, drive	1.778282255
2	withdrawal, drug, zaps, medication, symptoms	8	positive_SSRI	2, withdrawal, drug	0.3026863413
2	withdrawal, drug, brain, zaps, brain zaps	19	positive_SNRI	2, withdrawal, drug	1.4740108611
2	zaps, brain, withdrawal, brain zaps, drug	16	negative_SSRI	2, withdrawal, drug	1.5503875969
2	withdrawal, drug, zaps, brain, brain zaps	38	negative_SNRI	2, withdrawal, drug	2.9480217223
3	weeks stopped, worked, stopped, begin, stopped working	4	negative_SNRI	3, helps, helpful	0.310318076
3	works, helps, helpful, effective, worked	29	positive_SSRI	3, helps, helpful	1.0972379871
3	didn, work, help, 2014, teens	12	negative_SSRI	3, helps, helpful	1.1627906977
3	helps, helpful, wonders, works, worked	23	positive_SNRI	3, helps, helpful	1.7843289372
4	couldn, got switched, drug, heart, got	3	negative_SNRI	4, life, drug	0.232738557
4	just, depressed, people, lack, drug	5	negative_SSRI	4, life, drug	0.484496124
4	life, just, drug, went, depression	17	positive_SNRI	4, life, drug	1.3188518231
4	life, happy, better, depression, drug	38	positive_SSRI	4, life, drug	1.4377601211
5	neck, kids, bad, play, moderate depression	1	negative_SNRI	5, anxiety, drug	0.077579519
5	anxiety, drug, depression, right, day	5	negative_SSRI	5, anxiety, drug	0.484496124
5	anxiety, drug, years, depression, like	23	positive_SSRI	5, anxiety, drug	0.8702232312
5	anxiety, drug, depression, years, like	20	positive_SNRI	5, anxiety, drug	1.5515903801
6	panic, dosis, anxiety, drug, attacks	5	positive_SNRI	6, anxiety, dosis	0.387897595
6	anxiety, dosis, drug, horrible, days	7	negative_SNRI	6, anxiety, dosis	0.543056633
6	anxiety, dosis, panic, attacks, drug	25	positive_SSRI	6, anxiety, dosis	0.9458948165
6	anxiety, dosis, panic, attacks, drug	10	negative_SSRI	6, anxiety, dosis	0.9689922481
7	wife, lol, money, month later, push	1	negative_SSRI	7, drug, depression	0.0968992248
7	drug, effects, antidepressants, noticeable, depression	6	negative_SNRI	7, drug, depression	0.465477114
7	drug, depression, worked, effects, drugs	11	positive_SNRI	7, drug, depression	0.8533747091
7	depression, drug, effects, drugs, worked	28	positive_SSRI	7, drug, depression	1.0594021945
8	generic, brand, generic drug, drug, insurance	15	positive_SSRI	8, generic, brand	0.5675368899
8	generic, brand, generic drug, drug, insurance	8	negative_SNRI	8, generic, brand	0.6206361521
8	generic, brand, generic drug, drug, drug generic	8	negative_SSRI	8, generic, brand	0.7751937984
8	generic, brand, generic drug, drug, insurance	13	positive_SNRI	8, generic, brand	1.0085337471
9	panic, drug, anxiety, drugs, having	3	negative_SNRI	9, anxiety, drug	0.232738557
9	anxiety, attacks, taking, panic, drug	5	negative_SSRI	9, anxiety, drug	0.484496124
9	anxiety, drug, medication, feel, taking	10	positive_SNRI	9, anxiety, drug	0.7757951901
10	drug, pounds, weight, feel, weeks	8	positive_SSRI	10, drug, weight	0.3026863413
10	nausea, feeling, haven, drug, 10	5	negative_SNRI	10, drug, weight	0.387897595
10	hair, weeks, drug, started, days	6	negative_SSRI	10, drug, weight	0.5813953488
10	drug, lost, days, weight, dizzy	19	positive_SNRI	10, drug, weight	1.4740108611
11	quit cold, just wanted, prescribe, hated, week felt	1	negative_SNRI	11, antidepres, anti	0.077579519
11	antidepressant, antidepressant ve, depression anxiety, successful, anxiety	4	negative_SSRI	11, antidepres, anti	0.3875968992
11	antidepressant, anti, drug, results, depressants	13	positive_SSRI	11, antidepres, anti	0.4918653046
11	antidepressant, anti, drug, results, best	18	positive_SNRI	11, antidepres, anti	1.3964313421
12	dosis, sleep, dose, eat, started	3	negative_SNRI	12, dosis, weeks	0.232738557
12	dosis, sleep, weeks, dosis weeks, night	6	positive_SNRI	12, dosis, weeks	0.465477114
12	dosis, night, effects, depression, feel	18	positive_SSRI	12, dosis, weeks	0.6810442679
12	dosis, effects, ve, weeks, just	9	negative_SSRI	12, dosis, weeks	0.8720930233
13	snappy, really help, didn really, dosis, didn	3	negative_SNRI	13, dosis, drug	0.232738557
13	dosis, dosis day, drug, days, lot	5	positive_SNRI	13, dosis, drug	0.387897595
13	dosis, drug dosis, drug, cold, taking	5	negative_SSRI	13, dosis, drug	0.484496124
13	dosis, drug, life, taking, friend	22	positive_SSRI	13, dosis, drug	0.8323874385
14	withdrawal, drug, shocks, effects, taper	7	positive_SSRI	14, withdrawal, drug	0.2648505486
14	withdrawal, drug, effects, dosis, symptoms	5	negative_SSRI	14, withdrawal, drug	0.484496124
14	withdrawal, drug, withdrawal symptoms, symptoms, effective	10	positive_SNRI	14, withdrawal, drug	0.7757951901
14	withdrawal, withdrawal symptoms, symptoms, effects, drug	11	negative_SNRI	14, withdrawal, drug	0.8533747091