PhD Thesis

Using spontaneously generated online patient experiences to improve healthcare: A case study using Modafinil

Julia Walsh

A thesis submitted for the degree of Doctor of Philosophy in Health Sciences

University of Warwick, Warwick Medical School, Department of Health Sciences

July 2021

Table of contents

List of tables

List of figures

List of Appendices

Acknowledgements

My first thank you must be to my supervisor Professor Frances Griffiths. You took an enormous chance on me when I first came up with this hare-brained scheme, and have been endlessly supportive, encouraging and patient with me ever since. You have an amazing ability to come up with excellent one-line insights which have really helped me clarify what I was trying to do. Supervisions were always thought provoking and yet fun at the same time. I freely admit I was terrified of you to begin with, but I now consider myself very lucky to have had you as my supervisor. I just hope you do not have too many regrets! Thanks also to my other supervisor Jonathan Cave for his many words of wisdom.

Thank you also to Warwick Medical School for funding this PhD, and to all the people who have encouraged and helped me during this and the Masters. I particularly need to give a big thank you to Sam Plumb who went out of her way to come and find me during one of my MSc modules, and encouraged me to apply for the PhD in the first place. Thank you, Sam, that made such a difference to my mindset.

Thanks also to all the Farmhouse and F106 Crew. I started this journey with Ines and Tommer, and have greatly appreciated their friendship, encouragement, cups of tea, quizzing skills as well as the endless debates about everything. I owe a massive thank you to them and Julia G, Ash and Michelle for their brilliant proof reading. I read somewhere that only other PhD students can really understand the highs and lows of the experience, and I have to say that I probably agree with that. We were so lucky to have those Farmhouse years - you and the rest of the Farmhouse community made it a very special place.

Thanks also to my parents, who waited very patiently for far longer than they should have to see me finally have the sense to embark on this academic journey. Sadly, my father is no longer here to see me finish it, but I'm so glad that he knew that I was on this last stage. And lots of appreciation to all the other family and friends who have been there for me, I'm very lucky to be surrounded by such lovely people.

And last, but not least, a big thank you to my two daughters. You were both in your own ways also responsible for kick starting me into starting this adventure. Sarah - you may be the 'proper' doctor, but as you are now an Allen, it's obviously down to me to make sure that we finally have another Dr Walsh in the family. Kate - thank you for just being you. I'd never have done it without you. I love you both very much.

Declaration / Signed statement

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree.

[redacted]

Julia Walsh

5th July 2021

List of submitted papers.

Walsh J, Cave J, Griffiths F. Spontaneously Generated Online Patient Experience of Modafinil: A Qualitative and NLP Analysis. Frontiers Digital Health [Internet]. 2021 Feb;3. Available from: http://dx.doi.org/10.3389/fdgth.2021.598431

Accepted Conference presentations

Oral Presentation at HealTAC 2019 (Cardiff, UK): “Using spontaneously generated online patient experiences (SGOPE) to generate knowledge and understanding to improve healthcare: A case study using Modafinil”.

Abstract

Background

Acknowledged issues with the RCT focus of EBM and recognition of the value of patient input have created a need for new methods of knowledge generation that can give the depth of qualitative studies but on a much larger scale. Almost half of the global population uses social media regularly, with increasing numbers of people using online spaces as either a first- or second-line health information and exchange resource. Estimates suggest the volume of online health related data grew by 300% between 2017 and 2020. As a data source, this unstructured freeform textual data is a form of patient generated health data, containing a mass of patient centred, contextually grounded detail about the perceptions and health concerns of those who post online. Methods for analysing it are at an early stage of development, but it is seen as having potential to add to clinical understanding, either by augmenting existing knowledge, or in aiding understanding of real-world usage of healthcare interventions and services.

Objectives

To explore how large-scale analysis of SGOPE can help with understanding patient perspectives of their conditions, symptoms, and self-management behaviours, assess the effectiveness of interventions, contribute to the process of knowledge and evidence creation, and consequently help healthcare systems improve outcomes in the most efficient manner. A secondary aim is to contribute to the development of methods that can be generalised across other interventions or services.

Methods

Using Modafinil as a case study, a multistage approach was taken. First, an exploratory study, comparing both qualitative and basic NLP techniques was undertaken on a small sample of 260 posts to identify topics, evaluate effectiveness and identify perceived causal text. An umbrella scoping review was then undertaken exploring how and for what purposes SGOPE data is currently being used within healthcare research. Findings from both then guided the main study, which used a variety of unsupervised NLP tools to explore the main dataset of over 69k posts. Individual methods were compared against each other. Results from both studies were compared and for evaluation.

Results

In contrast to the existing inconclusive systematic review evidence for Modafinil for anything other than narcolepsy, both studies found that Modafinil is seen as by posters as effective in treating fatigue and cognition symptoms in a wide range of conditions. Both identified the topics mentioned in the data, although more work needs to be done to develop the NLP methods to achieve a greater depth of understanding. The first study identified eight themes within the posts: reason for taking, impact of symptoms, acquisition, dosage, side-effects, comparison with other interventions, effectiveness, and quality of life outcomes. Effectiveness of Modafinil was found to be 68% positive, 12% mixed and 18% negative. Expressions of causal belief were identified. In the main study, effectiveness was measured with sentiment analysis, with all methods showing strong positive sentiment. Topic modelling identified groups of themes. Linguistic techniques extracted phrases indicating causality. Various analysis methods were compared to develop a method that could be generalised across other health topics.

Abbreviations

ADE	Adverse Drug Event
ADR	Adverse Drug Reaction
CERQual	Confidence in the Evidence from Reviews of Qualitative research
CI	Confidence Interval
CQL	Corpus Query Language
DHSN	Digital Health Social Network
DL	Deep Learning
EBM	Evidence Based Medicine
EHR	Electronic Health Record
ePAT	Electronic Patient Authored Text
ICER	Incremental Cost Effectiveness Ratio
KDD	Knowledge Discovery in Databases
LDA	Latent Dirichlet Allocation
LIWC	Linguistic Inquiry and Word Count
LSA	Latent Semantic Analysis
ML	Machine Learning
MUPS	Medically Unexplained Physical Symptoms
NER	Named Entity Recognition
NER+L	Named Entity Recognition plus Linking
NIC	Net Ingredient Cost
NLP	Natural Language Processing
NMF	Non-negative Matrix Factorisation
NNH	Number Needed to Harm
NNT	Number Needed to Treat
OSN	Online Social Networks
PAT	Patient Authored Text

PCA	Principal Component Analysis
PCC	Patient Centred Care
PGHD	Patient Generated Health Data
POS	Parts of Speech
PROM	Patient Reported Outcome Measurement
PtEx	Patient Experience
QALY	Quality Adjusted Life Year
QDA	Qualitative Data Analysis
RCT	Randomised Controlled Trial
RR	Risk Ratio
RWD	Real World Data
RWE	Real World Evidence
SGOPE	Spontaneously Generated Online Patient Experience
SIDER	Side Effect Resource
SM	Social Media
SoMe	Social Media
SSRI	Selective Serotonin Re-uptake Inhibitor
TCA	Tricyclic Antidepressant
TNRD	Treatment Non-Responsive Depression
UGC	User Generated Content
UMLS	Unified Medical Language System

Chapter 1 Introduction

1.1 Outline

This chapter introduces the motivation and rationale for the thesis and gives a brief background to the component subject areas. It concludes by setting out the thesis aims, objectives and the specific research question that will be answered.

1.2 Motivation – and why Modafinil?

The rationale for this project is multifaceted. The original idea came from an undergraduate assignment I did looking at various methods of biomedical enhancement. One such method was cognitive enhancement, where people were actively experimenting and discussing various methods of improving their cognitive ability. The background research for that essay illustrated the depth and extent of health-related information that is shared by individuals on social media. Although Modafinil, a drug currently only indicated for narcolepsy in the UK [1,2] was just one of the many approaches that posters were exploring, it was particularly interesting to realise just how many people were exchanging their experiences of using Modafinil purely for therapeutic use to relieve symptoms of fatigue and sudden onset cognitive dysfunction rather than for any form of enhancement.

These symptoms are seen in a variety of conditions and clinical presentations and are both debilitating and distressing. Modafinil is an out of patent oral wakefulness-promoting drug, first developed in the late 1990s, known to be relatively safe, and with low abuse potential [3]. Despite the narrow licensing in the UK which restricts its wider use [1,2], its US FDA status enables clinicians to prescribe it 'off label' to improve cognition or fatigue symptoms in many other conditions. Around 90% of its prescribed US usage is 'off label' [4]. Modafinil has been considered a potential therapy for a range of conditions [5], including ADHD [6], multiple sclerosis [7,8], premature ejaculation [9], depression [10], Parkinson's disease [11], chemotherapy related fatigue [12,13], traumatic brain injury [14] and cocaine dependence [15]. Findings have been mixed, with systematic reviews generally inconclusive, showing either insufficient [14,16–18] or low quality evidence [18–20]. Previous studies have commented on the lack of research into either long term [21] or 'as required' use [22]. However, despite the lack of conclusive trial based evidence upon which clinical guidelines are generally based [23–25] there appears to be a substantial amount of online discussion suggesting that there are many people for whom it has made a significant difference to their symptoms and quality of life [26].

Since becoming aware of this apparent contradiction between the evidence that guides clinical practice and the patient experiences as reported on social media, there have also been a steady stream of high-profile examples where changes in healthcare services and delivery have been driven by groups of individuals gathering in online spaces to exchange experiences and information. These include the withdrawal after 48 deaths and 873 pregnancy losses [27] in 2020 of devices such as Essure, an easily inserted permanent, non-hormonal contraceptive, which received FDA approved in 2002. Retrospective analysis of accessible social media conversations over the period 2002-2016 showed that women had been raising concerns on social media and online communities almost immediately, with individual reports to doctors often being dismissed as 'one off's [28]. Similar widely reported examples where social media communities have formed to exchange health experiences that conflict with current practice and evidence include the use of vaginal mesh, sodium valproate and Primodos [29].

Medical knowledge is continually developing, the COVID pandemic has clearly highlighted how neither medicine, science or governments have all the answers to how the disease impacts the body, how best to treat it or how to contain an outbreak. Initial assumptions that it was a purely respiratory disease proved incorrect and epidemiological models were wildly inaccurate [30]. The emergence of Long Covid with its wide range of symptoms including fatigue and cognition issues, is a recent example of a condition that was mainly discovered by patients finding each other on social media and collectively exchanging knowledge with each other - knowledge that challenged the early clinical and scientific assumptions about Covid-19. Rather than being generated by health researchers, knowledge of this condition was generated by patients in their online communities and was subsequently picked up by formal health and policy channels [31]. The Facebook Long Covid group which aims to restrict members to those who have either had Covid or care for someone who has ongoing symptoms currently has over 41,000 users sharing experiences and information. Acknowledgement of its existence is demonstrated by recent figures from the ONS estimating that 1 million people in the UK are impacted by it [32].

Posts written by individuals on social media platforms are creating vast resources of spontaneously generated online patient experience (SGOPE) data in the form of unstructured text, which may be able to contribute to our understanding of healthcare.

In this thesis, using Modafinil as a case study, I want to try and explore the potential of incorporating this narrative data to extend clinical knowledge, highlight new areas of research, and get a much better understanding of how patients manage their conditions and the outcomes and health values that are important to them.

1.3 Health related use of internet

Social media has become a cornerstone of everyday life, transforming the way that individuals and organisations share ideas, beliefs, news, and information. Of a global population of 7.953 billion, a report in Oct 2019 showed 4.479 billion as internet users [33] of whom 3.725 billion were using social media at least once a month [34]. Twitter has 955 million registered users [35]. Figures from two of the most popular sites report that on an average day, social media users are reported to be making around 500 million tweets and 2.5 million Reddit comments [36]. Within the UK, ONS figures indicate that 87.9% of UK adults had recently used the Internet (ONS 2016) and there are 45 million active social media users, representing 67% of the population as of January 2019 [37]. Usage is also growing rapidly among older people, often seen as the age group that is least engaged with new technology with 82% of 65-69 year old US adults using the internet [38] .

This rapid growth of social media has led to fundamental changes in the way that people look for and share health related information. In the US 72% of adults use the internet for health purposes [39], either as a first or second-line health information source [40] or exchange resource [41,42]. Estimates suggest the volume of online health related data grew by 300% between 2017 to 2020 [43].

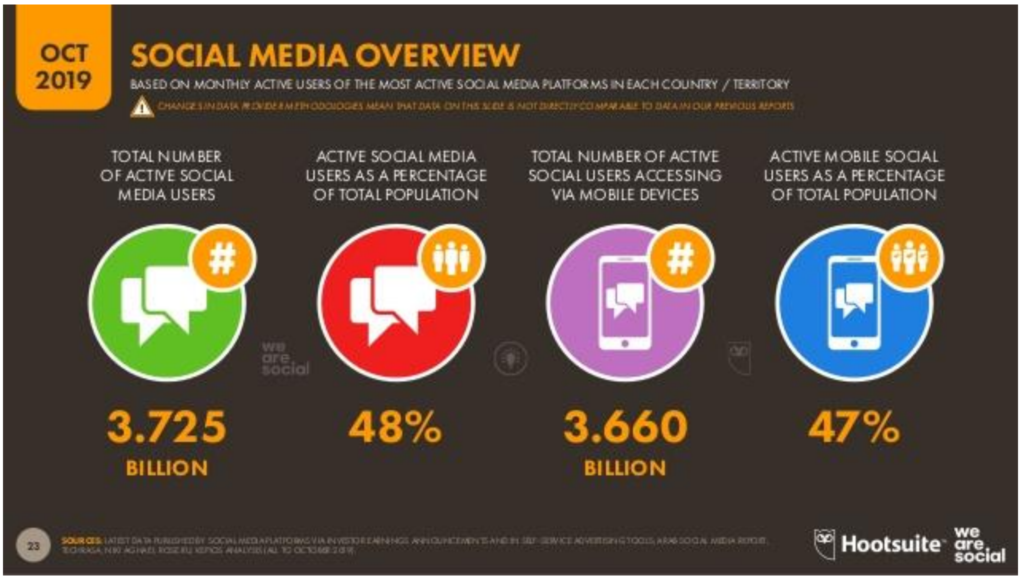


Figure 1-1: Global social media usage October 2019 [34]

In terms of where people are looking for health care information, Figure 1-2 shows the range of sources, other than clinicians, that were used in the UK in 2018, with online search engines being the most widely utilised. Between 23% to 75% of adults report using the internet to explore and appraise their symptoms [40].

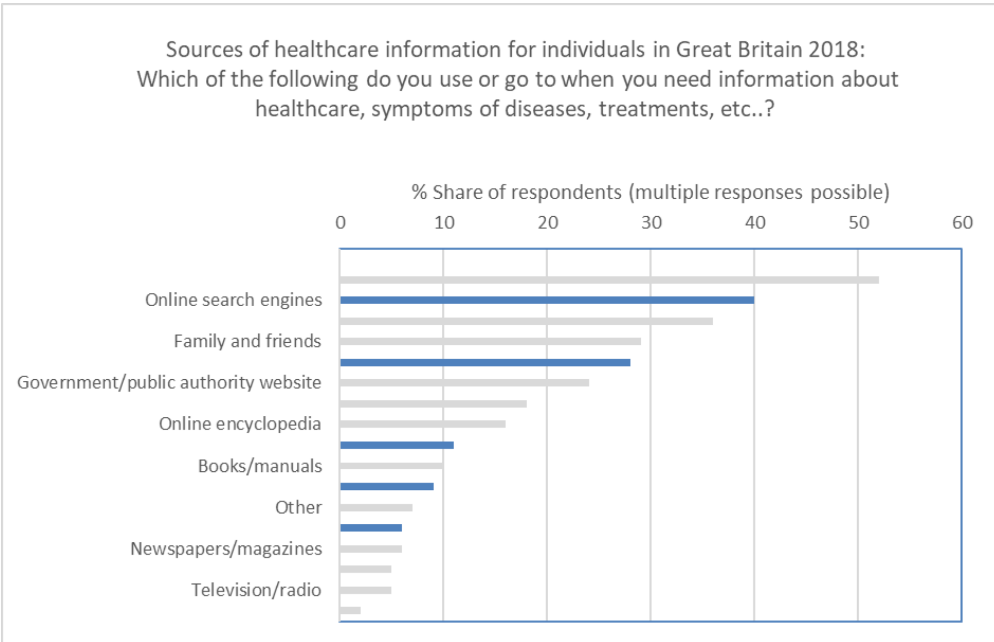


Figure 1-2: Sources of health care information UK 2018 (Ipsos Mori)

Restrictions and local lockdowns due to the global COVID-19 pandemic are likely to have led to an even greater use of health-related online use, as individuals may have avoided or been unable to access personal visits to clinicians [44].

1.4 What is SGOPE and why do people use it?

Spontaneously generated online patient experience (SGOPE) is defined as any form of textual content posted by an internet user on social media relating to a health experience. It is a component of the mass of patient generated health data (PGHD) that is becoming widely available from sources such as apps, wearable devices and home monitoring devices and being used to supplement existing clinical data [45] . The relative novelty and cross disciplinary nature of this data source has led to a variety of terms and acronyms being used to describe it (Table 1-1). For the purposes of this thesis, I will use SGOPE to refer to such data as this phrase most accurately describes the features that make it unique.

Table 1-1: Other acronyms for SGOPE

SGOPE	Spontaneously generated online patient experience
PGHD	Patient generated health data
ePAT	Electronic patient authored text
UGC	User generated content

Online health sites are a fast, convenient and free information source where patients or their carers can spend as long as they want, asking whatever questions they have with a degree of anonymity that encourages a form of disinhibition that may not be seen in a time restricted clinical encounter [46,47]. They can help individuals to both cope with and manage their conditions, overcome geographic boundaries or isolation, physical challenges and disease stigma [48]. Previous studies found that such online communities are often a result of patients and families seeking to add to the research about a condition [31,49], either to advance the current knowledge [50] or to benefit future patients [31,51–53]. For those with long term or rare conditions, online communities help provide a reference for normal illness experiences and trajectories, as well as building a knowledge base built on lived experience that can go beyond current medical knowledge [54]. Engaging in online patient communities has been shown to empower patients [55,56] and improve both self-management and self-efficacy [57]. Different types of users [58,59] and experiences have been identified [60].

Estimates suggest that around 90% of all the digital internet content which has greatest potential for creating new insights and knowledge exists in unstructured format [61]. However, the volume of data available and lack of established methodologies to analyse it inhibits its use [46,62–64]. Traditional qualitative methods, while capable of generating the depth of understanding are too time consuming to use on large quantities of data, while most conventional quantitative methods require the data to be both structured, and with values present for each field. This has created a need to find new methods of knowledge generation and evidence production to take account of the changes in technology, populations, and policies.

1.5 What is NLP?

Natural language processing (NLP) refers to the use of computational techniques and algorithms that aim to extract information, perform syntactic processing, capture meaning, assess sentiment and identify relationships from natural language free text [65]. The ultimate goal is to interpret the semantic meaning from large volumes of text [66]. Sitting across disciplines of data science, statistics and linguistics it is a rapidly developing area [67], and has emerged as an important tool for processing unstructured free text such as that found in SGOPE or EHRs [65]. Its development began with Alan Turing's 1950 paper describing a 'thinking machine' [68], designed to translate between languages. Methods were predominantly rules-based until the mid-1980s. By the late 1980's statistical approaches were being developed using decision trees or regression-based calculations. These are much easier to scale up, and use labelled or annotated data to make predictions about unseen data. One of the biggest issues is that they do not take any account of the usage and context of language and cannot make the 'common sense' interpretation of a human. Deep learning models first appeared in 2010, and the last few years have seen the widespread adoption of speech recognition systems such as Alexa and Siri.

NLP has been used on SGOPE data to explore health related social media usage [39,69–72], detecting drug or device related adverse events from user generated content [73,74], generating new understanding about treatment switching and adherence behaviour [75,76] and as a surveillance tool for infectious disease outbreaks [77,78] and suicide risk [79] although little work has been carried out into its use for assessing effectiveness [75].

The use of NLP methods offers an opportunity to scale up some of the advantages of qualitative methods, such as depth of understanding, to much larger samples [80,81] which could be seen as having greater legitimacy for inclusion in evidence bases.

1.6 Why does it matter / Relevance to current policy?

Making the most effective use of resources within healthcare is a significant ongoing issue. There is a growing recognition that patient input can help achieve this, particularly in terms of reflecting the outcomes that matter to patients [82,83]. The need to include the patient's perspective in health research is recognized by all the major health organizations and the last few years have seen a raft of initiatives to try to achieve this [82–86]. SGOPE is a data resource which could help address this need. It offers a wealth of narrative detail regarding the 'real world' patient experiences of interventions [70,71], conditions [72,87] and services [88] providing a unique window into how those writing on social media perceive the effectiveness of interventions and the health related outcomes that they value. The free form narrative text available gives a depth of response as to the experiences of either patients or their carers, including detail such items as the reasons for starting or stopping interventions, effects on their quality of life, whether tangible or intangible, comparisons with other interventions that they may have tried or moved onto, side effects, and how effective they find it in regard to the condition that is concerning them.

Exploring the potential of SGOPE will lead to a much broader perspective for both clinicians, researchers, and service providers. In addition to gaining insight into patient priorities, a better understanding of patient behaviours has many potential benefits for health systems, including making the most effective use of resources. Adherence or non-compliance with prescribed medication is recognised as a problem with significant economic costs [89]. Estimates of the annual costs just within the USA range from US$100 to U$290 billion [90]. Although an intervention may have been shown by high quality research to be effective, it only results in a good outcome if the patient adheres to the prescribed course of action or dosage, which they may not if it is having a negative impact on some area of their life [91]. Non-adherence to medication is rarely reported [92] although a WHO report found that adherence rates for long term therapies averaged 50% in developed countries and were even lower in developing countries [93]. SGOPE may be able to highlight and explain real-world non-compliance on a large scale.

Within any clinical encounter, time is limited, and no matter how comprehensive the notes taken are, they are just a snapshot of what was deemed clinically important, whereas SGOPE offers a deeper and potentially longitudinal view into the patient perspective of all aspects of their condition, treatments and management [94].

The natural, non-clinical language that the posters use can contain valuable information that may remain hidden from view in a clinical or research setting [95,96].

Trying to gain a better understanding of these experiences is an aspect that organizations such as the FDA [85] and The Royal Society of Medicine [97] have already demonstrated their commitment to by running a series of public meetings from which they aimed to elicit a patient based understanding of various conditions. The FDA Voice of the Patient initiative has been a five-year process, covering around twenty different disease areas, asking the following type of questions to patients and their carers.

Table 1-2: Typical questions asked of patients by FDA Voice of the Patient project

Typical questions asked of patients by FDA Voice of the Patient
Which three symptoms have the most impact on your life?
Are there specific activities that you cannot do because of your condition?
How has the condition changed over time?
What worries you most about the condition?
What are you currently doing to help treat it?
How well does the existing treatments treat the most significant symptoms?
What are the negative impacts of the treatments?
What would an ‘ideal’ treatment do for your condition

A superficial comparison of the reports produced by the FDA project [85,98] to posts from online sources read during my background research shows a remarkable similarity in the types of themes and concerns expressed, indicating the potential for SGOPE data to reliably express patient’s views on both a much larger scale than is possible through public meetings and at a significantly lower cost. Natural language processing and machine learning (ML) have emerged as an important tool for processing unstructured free text from patient experience feedback [65].

The project fits in with many of the current priorities within health care including:

- trying to reflect the outcomes that matter to patients [83]
- encouraging self-management of conditions [99,100]
- making the most use of the available resources - [101]
- stratified medicine (MRC 2017; ABPI 2014)
- accelerated access to effective treatments [102]
- repurposing [103]

- move to ‘value-based health care’ as a way of restructuring health systems - ICHOM [104]
- addressing some of the difficulties and concerns with the current research agenda [105]

1.7 Thesis aims, objectives and research question

The overall aim of this project is to explore the potential for SGOPE data to help with understanding patient perspectives of their conditions, symptoms, and self-management behaviours, assess the effectiveness of interventions, contribute to the process of knowledge and evidence creation, and consequently help healthcare systems improve outcomes in the most efficient manner.

As the project will focus on patient experiences of one drug from multiple social media sites, the main research question is:

Do spontaneously generated online patient experiences of Modafinil use have the potential to become part of the evidence as to its effectiveness for use in clinical practice?

Answering this question will help with the secondary project aim of contributing to the development of methods that can be generalised across other interventions or services.

Thesis Objectives

- Objective 1 (O1): To compare findings from a qualitative and basic NLP analysis of a sample dataset of publicly available SGOPE data relating to real-world experiences of the effectiveness of Modafinil.
- Objective 2 (O2): To summarise what is known about the tools, methods, and purposes of existing SGOPE based research using an umbrella scoping review.
- Objective 3 (O3): Based on the findings of the exploratory study and the umbrella scoping review, use NLP methods to explore a large-scale dataset of Modafinil experiences.
- Objective 4 (O4): To contribute to the development of methods that can be used for other forms of unstructured health data.

1.8 Thesis Structure

The thesis contains seven chapters. Figure 1-3 provides an overview of the structure, illustrating how each chapter addresses the objectives.

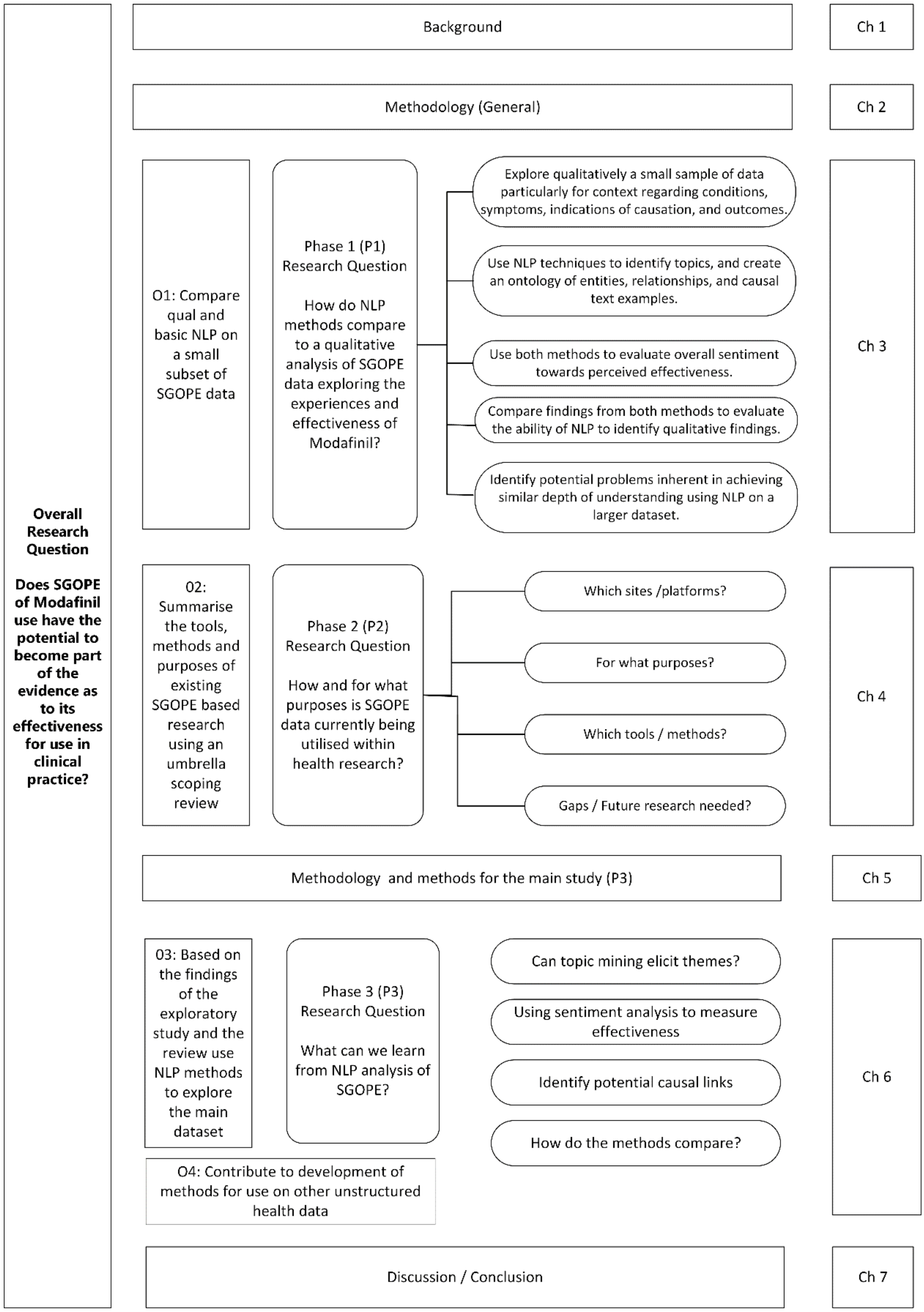


Figure 1-3: Thesis structure

Chapter 2 Methodology General

2.1 Outline

This chapter outlines the background to the methodological approach to the project. Recapping the thesis objectives, it includes a description of the overall study design and methods, background, and rationale for the design choices together with a brief discussion of the ethical issues involved in the use of SGOPE data.

Thesis Objectives

1. To compare findings from a qualitative and basic NLP analysis of a sample dataset of publicly available SGOPE data relating to real-world experiences of the effectiveness of Modafinil.
2. To summarise what is known about the tools, methods, and purposes of existing SGOPE based research using an umbrella scoping review.
3. Based on the findings of the exploratory study and the review, use NLP methods to explore a large-scale dataset of Modafinil experiences.
4. To contribute to the development of methods that can be used for other forms of unstructured health data.

2.2 How is clinical knowledge and evidence currently generated?

Widening what we know about the effectiveness of any healthcare intervention requires understanding the way that knowledge is currently generated within healthcare. For the last few decades, the overriding culture in health provision has been that of evidence-based medicine (EBM). Defined as being a systematic approach to clinical problem solving reached by the integration of best research evidence with real-world clinical expertise and patient values [106], it is intended to ensure that clinical guidelines and decisions are made on the most up-to-date, solid, reliable, scientific evidence. While theoretically EBM is designed to reflect the best combination of all forms of knowledge, in reality the pyramid shaped hierarchy of evidence quality ensures that it is findings from RCTs and subsequent systematic reviews, rather than any other form of knowledge, that tend to dominate and be reflected in the guidelines [23–25]. However there are ongoing methodological and philosophical debates as to the epistemic primacy of RCTs and whether or not other methods of assessing and evaluating the efficacy of interventions need to be included in the evidence generation process [107–110]. Although RCT's are undoubtedly the best way for generating certain types of knowledge, healthcare evidence is not a finished product. In addition to the apparent potential contradiction between the existing RCT evidence and the patient perspective of the effectiveness of Modafinil that is the subject of this project, there are many other research questions that could be answered by analysis of SGOPE.

Based on small, tightly defined samples of the population, RCTs aim to reduce the presence of any confounding variables that may impact on the measurement of the results. However, the strictly controlled selection of patients onto trials, means that efficacy results from a trial setting may not be reflected when the same intervention is used in a real-world setting [111] especially where increasing numbers of patients are presenting with complex multimorbidities or dispositions [111]. A report on the Salford Lung Study highlighted how the specific inclusion and exclusion criteria required for a patient's entry into a RCT limited the proportion of possible patients with asthma and COPD that could be included to just 3% and 7% respectively [112].

Results from clinical trials are also dependent on the ability to accurately measure the outcome or end point that is targeted. In simplistic terms, effectiveness is indicated by the presence of a beneficial causal effect of the intervention on the target. This is much easier to achieve if the outcome is one that can be recorded objectively such as cholesterol levels in blood, or tumour size but is much more difficult if the effect size is harder to measure or subjective, as is the case with self-reported fatigue or cognition problems.

On a practical level, it is recognised that the envisaged 'quality' standard of RCTs can be also be flawed by a variety of factors such as conflicts of interest from commercial or academic funders [113], systematic biases [114,115], poor design, lack of generalizability to wider populations [23], methodological discrepancies in the literature [116], the replacement of one set of vested interests with another [117,118], and a lack of accounting for the patient's perspective [105,119].

One of the limitations of the focus on RCTs of EBM is that it is a 'population' driven approach, based on establishing what intervention is most likely to work across a population. At a very simplistic level, this takes a somewhat naive rationalistic view of healthcare assuming that there can be a 'one size fits all' approach to treatment. In reality, patients all have different sets of characteristics, properties or dispositions that can interact to either cause disease or affect their response to treatments.

This population approach is also in conflict with the move to patient centred care that current policy aims towards. Arguments continue both as to how patient centred EBM is, and how evidence based patient centred care should be [120]. Taking a genuinely patient centred approach to practice cannot happen without changes in the underlying ontology [121].

Causal dispositionalism is an alternative theory to the non-reductionist approach to causation, which may be relevant to this type of data. This takes a more nuanced view of how the characteristics or dispositions of both the intervention and the individual combine to affect the effectiveness [122]. Rather than taking a statistically based population level view, marginal cases and outliers are used as a starting point for further investigation of potential predicates [123]. It suggests that population level health research should be only part of the evidence generation process, and that it is listening to the patient narrative that can be the key to understanding their individual health [124]. One of the strengths of narrative data such as SGOPE is that it can enable both the author and the reader to make sense of the interplay of actions and contexts in the text in a way that conveys perceived causality [125].

2.3 Identifying perceived causation

Causation is central to healthcare, both in understanding the onset of disease or symptoms and the effectiveness of any interventions or management strategy used to treat them [111]. Showing causation in healthcare from non RCT data is problematic. There is the structural or cultural problem in that it is seen as something that can only be shown in empirical settings such as an RCT where all confounding factors are removed and the Humean principle of same cause same effect can be repeatedly shown [110,124]. Traditionally qualitative or any other non-trial based research has avoided describing any findings as 'causal' [126]. Indeed, the current author guidelines for journals in the JAMA network instruct that 'any form of causal language including use of terms such as effect and efficacy should only be used in RCTs.' They add that 'in any other type of study whatsoever results should be confined to using terms such as association or correlation' [127]. The mantra that 'correlation does not equal causation' is justifiably used, but that leaves the question of how it is possible to determine causation. Williamson argues that causation can be shown by identifying or understanding the underlying mechanism between a correlated cause and effect [128]. Traditionally, findings from health based qualitative studies have been seen as anecdotal, unrepresentative and not generalizable across populations [105]. But is there a point at which if enough people are independently reporting a health effect that it can become population level evidence [124]?

Causation can be defined as a reaction between two events, a cause event, and its consequence. The cause must precede the consequence and is counterfactual in the fact that the consequence would not have occurred without the cause. While this sounds quite logical and straightforward, causation theories are not necessarily definitive explanations of how events occur but rather represent how humans make sense of and understand the world [129]. Language used to describe cause and effect can be crucial to understanding the semantic meaning of a text but is not always easy to identify. One method is using transition words, where both a reason and consequence are provided, or that identify a sequence of events.

Table 2-1: Examples of text indicating sequential events

Firstly, to begin with, next, then following this, at this time, now, at this point, previously, before this, after, afterwards, subsequently, finally, at last, simultaneously, meanwhile

2.4 How can exploring SGOPE add to knowledge?

We already know that people react differently to drugs, and no matter how convincing results are obtained at a population level, there will be individuals for whom the effect will differ. While there are many clearly defined and understood intervention mechanisms within healthcare, the relationship between treatment and outcomes is complex and not always understood. Methods that can help identify those patients who will not respond well to a treatment, no matter how significant its efficacy has been shown to be at a population level, will give benefit both to the patient, in avoiding a treatment that will at best be ineffective, and at worst cause harm, and to the healthcare system in the avoidance of wasting resources on an intervention that will not be the most appropriate for the patient. Antidepressants are a very common example of this. Frequently prescribed for symptoms of fatigue or cognitive dysfunction they are known to be ineffective in around 30% of cases [130–132]. But at the moment we do not know which 30% of patients they will benefit or possibly harm until the patient has taken them for the required number of weeks before any effect is expected to become noticeable. Conversely, identifying those subgroups that are most or least likely to benefit from an intervention or service will benefit both patient and providers.

The need for a plurality of evidence generating methods has already been recognised [30,121,133]. SGOPE is one type of data that falls under the umbrella terms of real-world data (RWD) and real-world evidence (RWE). RWD includes healthcare data generated from anywhere other than through conventional RCTs, while RWE is defined as evidence derived from the aggregation and analysis of RWD [85] and is argued to have significant advantages that can be used to supplement or augment RCT findings, including the ability to identify 'clinical gaps' [134], indicating effectiveness of an intervention in the real world, on much larger populations, and much faster than can be achieved within the artificial and highly constrained confines of a RCT [135,136]. Combining data sources such as SGOPE with new methods of analysing unstructured data will enable the development of new and different approaches to knowledge and evidence generation.

Despite the known limitations, high quality trials that show conclusive findings, whether positive or negative, are an important component of medical knowledge. However, the position is less clear when the 'evidence' for an intervention comprises multiple inconclusive or low-quality studies.

Patient narrative is already recognised as a tool that can help patients, clinicians and researchers [125,137]. Containing a mix of both objective and subjective views SGOPE data gives a unique perspective into the way that patients perceive, manage and react to their conditions, how such conditions impact on their life, their treatments, or any other aspects of their health [138]. Generated by the posters themselves, it comprises a mass of contextually grounded questions, experiences, ideas, and suggestions raised by the people posting. As such, it is far more likely to relate to and reflect the aspects of their individual experiences than those questions that may be formulated by researchers. It has potential to add to clinical understanding, either by adding to knowledge where existing evidence is inconclusive [139], or in aiding understanding of real-world usage [140], although the methods for analysing it are still at a early stage of development [79,141–146]. One of the values of SGOPE data is that it can give an extended insight into the ideas, concerns, and expectations (ICE) of the posters. ICE is now seen as a valuable part of the clinical encounter but there are limits as to how much these can be explored or recorded in a typical encounter.

If we are to move towards a truly patient centred healthcare model, we do need to genuinely listen to what patients are saying. They may not always be right in their assumptions or understanding of how interventions work, but their experiences combine objective and subjective observations and generally reflect their normative state.

Analysis of SGOPE also offers an opportunity to address another ongoing debate as to the real-world value of much of the clinical research that is currently undertaken. In Table 2-2 I show the criteria that Ioannidis [117] believes should be fulfilled to ensure the clinical usefulness of research, the amount of clinical research he estimates that achieves this, compared with my estimate of the potential for SGOPE based studies to contribute to or generate research knowledge that address all eight criteria.

Table 2-2: Percentages of clinical research fulfilling 'clinically useful' criteria (adapted from (Ioannidis, 2016)

Criteria	Description	Studies published in major medical journals	All clinical research	Potential of SGOPE studies
Problem base	Will it solve a real problem?	Variable	<50%	Y – High incidence of fatigue/unexplained cognitive issues
Context placement	Has the need for new studies been demonstrated?	Variable	<20%	Y – Apparent disparity between patient narratives and published evidence
Information gain	Will it produce real knowledge, either supporting or refuting the hypothesis?	Between 50-80%	<1%	? – Aim of this study
Pragmatism	Does the research reflect 'real life'?	Rare	<1%	Y – Posts describe use in real life
Patient centred	Is research done to look at what really matters to patients, or is it led by funder or scientist interests?	<20%	<1%	Y – Data spontaneously generated by patients and for patients
Value	Will the research give value for money?	N/K rarely assessed	N/K rarely assessed	Y – Significantly cheaper than RCT
Feasible	Can it be done?	>99%	50-80%	? – Aim of this study
Transparent:	Are all aspects verifiable and unbiased?			? – potential for bias, hard to quantify, take steps to minimise
Data sharing		<20%	<20%	Y – Using publicly available data
Trial registration		>99%	<50%	N/A
Other study registration		<20%	<20%	?

2.5 What is different about SGOPE as a data source?

Although SGOPE is an extremely data rich source it presents many challenges for health researchers. This type of user generated content from multiple sources, in a variety of formats is not only unstructured data, in that it does not fit into conventional rows and tables but contains unstructured content; user generated free flowing text that does not have to follow any kinds of rules in either content or language. As a source, it contains both quantitative and qualitative data. The size and complexity of this type of data mean that it falls into a 'methodological gap' between that of traditional qualitative research (Figure 2-1), and that of 'big data' [147]. The volume of qualitative data that is available is far greater than is usually obtained by means of interviews or focus groups while the data content is determined by the posters rather than being led in any way by a researcher.

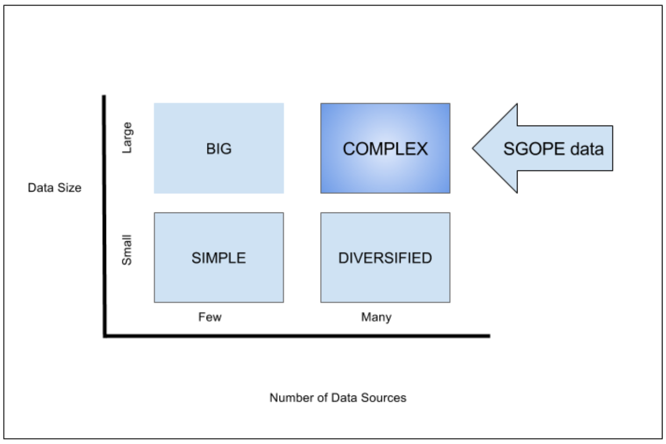


Figure 2-1: Complexity of SGOPE data

Despite the methodological complexities of analysing large volumes of unstructured natural language text, there has been increased interest from both commercial and academic researchers into methods of generating knowledge from it and new methods are developing rapidly [148–150]. Although analysis of the unstructured data may highlight patterns or clusters leading to new ideas, the real potential for creating new insights and knowledge will be in using natural language processing techniques to analyse the unstructured textual content [151] to assist in gaining a much deeper understanding of how patients are managing their conditions in the real world. Looking for knowledge from within the narrative moves the classification of this type of data away from being machine focused 'big data' looking for correlations, to people based 'small data' that can help identify inferred causality [152].

2.6 Strengths and limitations of SGOPE as a data source

As a data resource, SGOPE has a variety of strengths and limitations which are summarised in the SWOT analysis below (Figure 2-2).

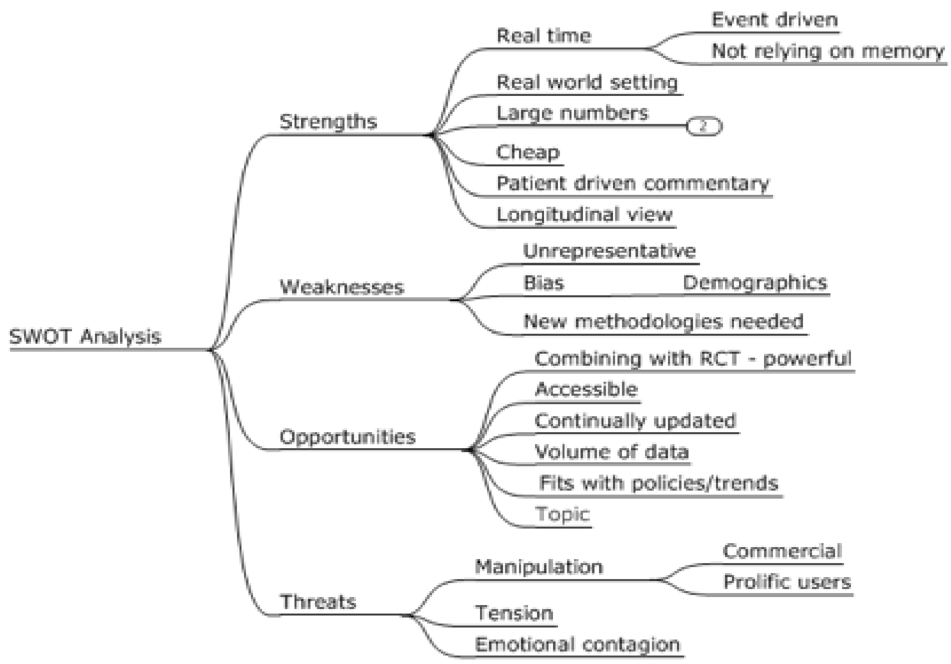


Figure 2-2: SWOT analysis of SGOPE as a data source

Utilising SGOPE data opens up unique opportunities for a more patient centric form of EBM, offering a new window into patient experiences [66]. It has many potential advantages over more traditional data sources such as surveys or interviews including ease of access [47] to large quantities of real time naturally occurring data, with a wider geographical spread that may encompass many traditionally harder to reach groups[96,153]. Posts are generated and can be collated in real time which is particularly useful when looking at rapidly evolving events [154]. The potential benefits of social media as a research resource for healthcare include reducing research costs [155], improving patient empowerment [156], engagement [157] and health communication [158].

Traditional qualitative analysis methods allow the researcher to guide the direction of the data collection and to probe further into areas that are perceived by the research team to be interesting. This is not possible when retrospectively analysing SGOPE data, as the researcher is not present when the data is generated. However, this can also be a strength as the posters are interacting with each other, and therefore guide the conversations in the directions that they feel are important to themselves.

Its global nature means that it can be difficult to identify either the geographic location or demographics of respondents [47]. Data sparsity can be an issue. Trending topics or media coverage can result in a tendency for some elements to be over-represented [159,160], while rarer or less trending phenomena are harder to identify and analyse [161].

Traditionally patient experience, particularly that gathered from social media, has been viewed as low-quality form of evidence, with questions asked as to its validity and reliability. The reliability and veracity of online content, or any type of self-generated digital patient experience feedback, is impossible to guarantee. However even within RCTs, when trying to measure effects on factors such as cognition and fatigue as against distinct biomarkers that can easily and empirically be measured, the study relies on the individual perspective, belief and self-reporting of any change brought about by the intervention. Although the identity of the trial participant may be easy to identify there is no way to guarantee that their response to the study organizers is any more accurate than those found in an online community. In fact, there is a possibility that a real-world response given to an interested group in an anonymous setting may be more representative of their experience, than one given in the artificial and sometimes unnatural confines of a study setting.

2.7 Overall Approach and Study Design

As this type of analysis does not yet have established methodologies [46,62–64], the study was designed using a pragmatic, multistage evaluation approach [162,163] to ensure the findings are as robust as possible (Figure 2-3). Combining a variety of methods, it focuses on patient experiences of one drug across multiple data sources as a case study to explore the potential of SGOPE.

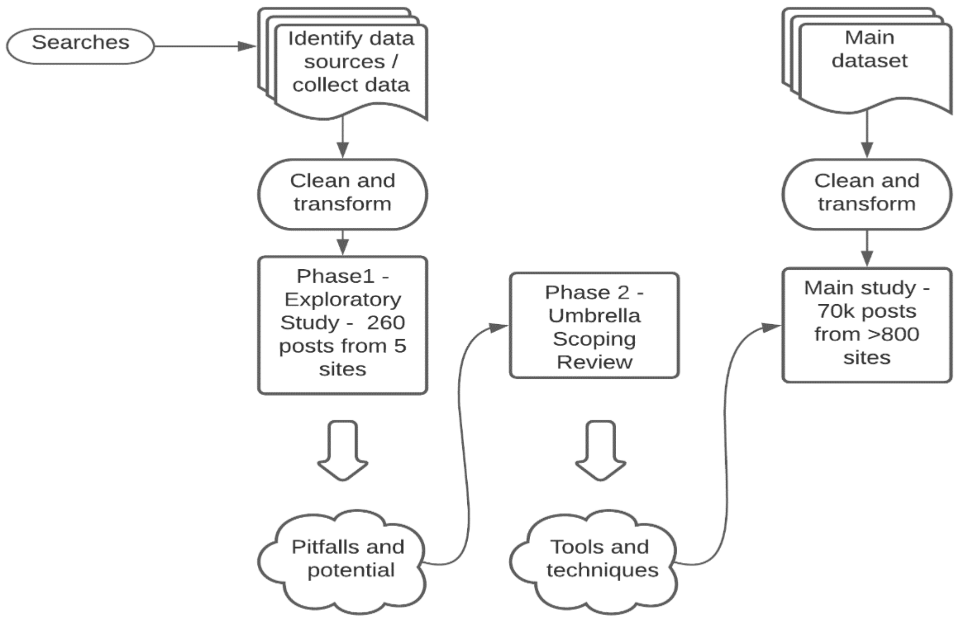


Figure 2-3: Overall study design

Choosing a case study approach allows a deeper exploration of a single intervention [164], which in this case will be an in-depth and multi-faceted analysis of the posters experiences of Modafinil in a real-world setting [165].

The study is also designed to include data from a wide range of SM sources. Existing studies have generally focused on analysing data from a single social media platform [166], however this can increase the potential for bias, particularly in terms of demographics [161,167]. The diversity of social media spaces, while adding to the technical challenges, also adds to the opportunities that it offers. The enormous volumes of data on the more general social networks such as Twitter, Reddit or Facebook can give almost a population view of a health issue, while the smaller condition specific spaces have a more targeted population, but may offer more in terms of focused data [66]. Each site comprises posts from a 'community' of people who feel comfortable there, potentially leading to an element of emotional contagion between the posters [168,169]. This clustering of individuals can also lead to a confirmation bias as consensus has been shown to have a positive impact on the perceived effectiveness of treatment [161]. Both studies use data from multiple sites to mitigate these biases and to maximise the range of voices to be heard.

The staged approach of this project is aimed to develop a clearly explained set of methods that compares the two methodologies, and that can generate findings that can contribute to both the evidence base and improved practice. The methods are also aimed to be generalizable to a wider range of interventions or evaluations, or other forms of unstructured text data.

Taking a multistage approach allowed me to explore qualitatively a small sample of the data, gaining a deeper understanding of what it contained. Performing basic NLP on the same dataset allowed me to compare the approaches, as well as identifying potential pitfalls in a full NLP analysis of the main dataset. Undertaking the review after the exploratory study ensured that the findings were as current as possible. Findings from both the exploratory study and the review were then used to guide the methodology of the main study. Developing the code for the main study analysis on the exploratory dataset enabled a further comparison of the findings as shown in Figure 2-4.

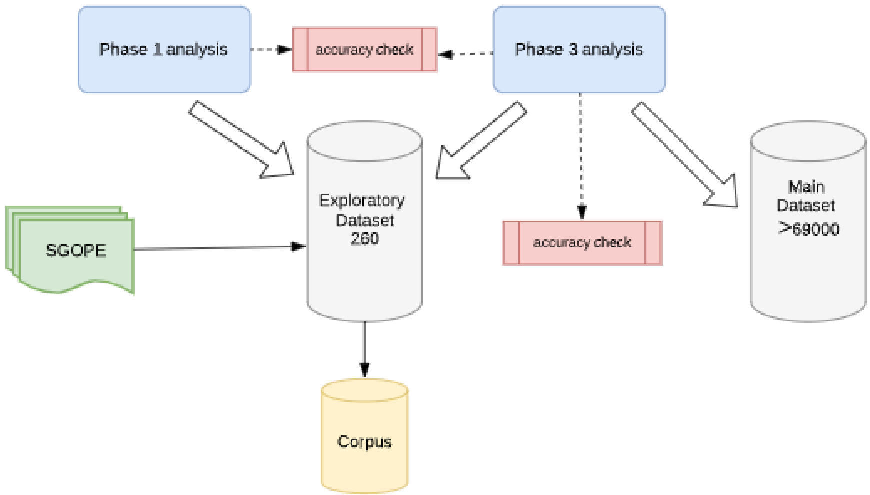


Figure 2-4: Built in validity checking

2.8 Phase 1 Exploratory study (P1)

The overall project is looking at the potential of NLP to explore SGOPE for large scale analysis. This exploratory phase combines a qualitative inductive thematic analysis [170] and basic NLP analysis of a small sample of 260 posts spontaneously generated online patient experience data (SGOPE) as in Figure 2-5. Findings from both methods are compared, and potential problems in a full NLP analysis identified.

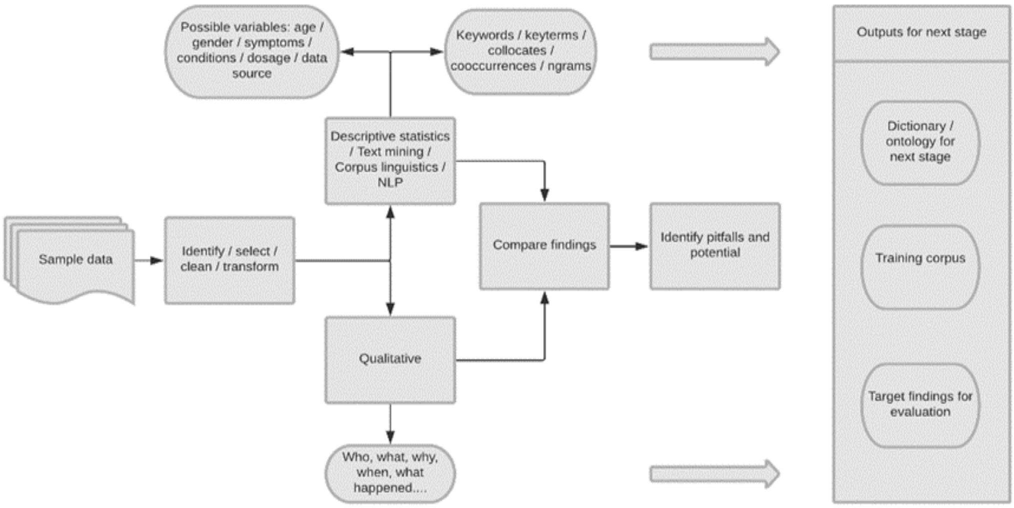


Figure 2-5: P1 Exploratory study design

.

Qualitative studies are widely accepted as a research method capable of achieving an in depth understanding of the data [171] although the small sample sizes mean that the findings are not seen as generalisable to wider populations so they are often seen as having limited value in the EBM hierarchy [172]. Combining a thematic qualitative study with a basic NLP analysis on the same dataset allowed a comparison between the two methods.

As the sample size of the exploratory P1 dataset is too small for a full NLP approach, I used frequency-based corpus linguistics term extraction techniques to emulate theme extraction. Corpus linguistics and NLP come from two distinct disciplines, NLP has its roots in mathematics and statistics, whereas linguistics explores the complexities and usage of languages. However, they also share some similarities in approach. Both start by breaking text down into the smallest possible unit, they can both use part-of-speech (POS) tags to identify grammatical constructs, and both can use stemming and or lemmatisation to simplify the text into a smaller number of word roots. There are no standard methods or procedures in a corpus linguistic methodology [47], but the well-established corpus techniques of keywords and keyterm extraction, together with collocation and concordance can be combined to quantitatively identify the key themes before exploring the linguistic features of the themes in a more qualitative manner [47].

Keyword and keyterm extraction is a method of identifying the most relevant or specific words and phrases within a corpus. While frequency analysis simply identifies the most frequently used words and phrases within a text, keyword and keyterm identification methods extend this by first calculating the frequency in the focus text and then comparing that to the frequency of the same word or phrase in a general language reference corpus. This helps identify the words and terms that are most specific to the corpus of data and which can be used to generate the themes and topics that are relevant to it. Keyword methods are useful for identifying unknown or rare entities. They can also identify many of the spelling variations of a term that may be within the text and can be used to generate domain or corpus specific dictionaries. Collocation refers to the association between words based on patterns of co-occurrence. It uses a word association measure to evaluate how often two or more words occur within close proximity of each other and whether this is significant rather than occurring by chance [47].

In addition to manually evaluating the poster's expression of the effectiveness of Modafinil, it was possible to use lexicon-based sentiment analysis NLP techniques to calculate effectiveness. Sentiment analysis is a well-known and widely used technique within NLP that analyses text for positive, neutral, or negative sentiment or emotion aiming to extract an understanding of the meaning, mood, context, and intent. It has already been shown to be capable of reasonable agreement with online comments including a Likert scale [173].

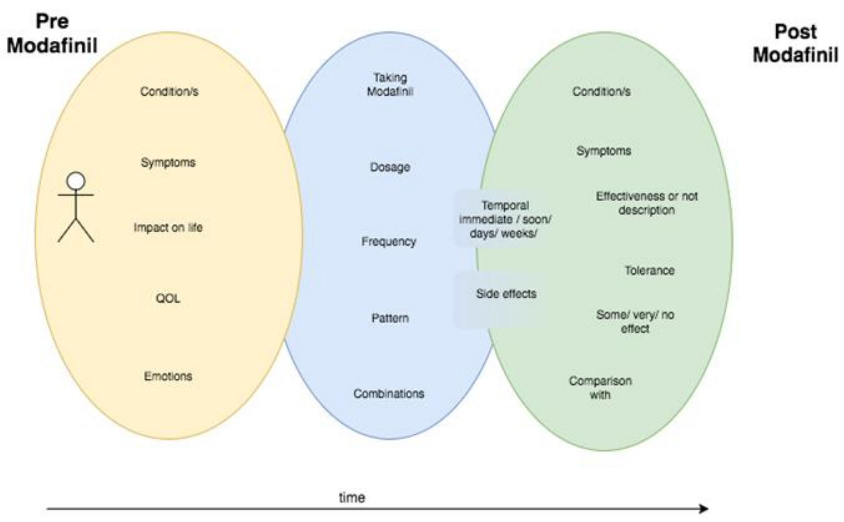


Figure 2-6: Information identified in the pre /post split

2.9 Phase 2 – Umbrella scoping review (P2)

The second stage is an umbrella scoping review of published reviews that looks at how and why SGOPE data is currently being used within health research. An umbrella review is a form of knowledge synthesis which by summarising existing review papers, aims to describe the subject area, what is currently known about it and identify the gaps in knowledge. I chose this method for two main reasons. Firstly, comparing existing reviews gives a wide overview of the subject area, highlighting existing evidence and illustrating how researchers across the various disciplines are exploring the topic. By avoiding the repetition of searches, screening of individual papers, and the resynthesizing of existing studies it provides an overall picture of the current state of the art that can be used as a broad base to build from [174]. Secondly, as the literature base is so varied, a scoping review enables the inclusion of other relevant 'grey' review literature that would not otherwise be included within a systematic review. NLP research sits across various disciplines and some of the most relevant and current research is not published in the mainstream health based journals [175]. Although not always subject to the peer review process of the more traditional journals, these other sources are an important source of information for a review on such a rapidly evolving subject area. Widening the scope adds both to the depth and breadth of the literature as well as reducing the potential for any possible publication bias. Undertaking this review after the first phase will enable the review to focus on the areas of interest and potential difficulties highlighted from the exploratory study.

2.10 Phase 3 – Main study (P3)

The final decisions for the methods for the main study were not made until both the exploratory study and state of the art review had been completed. This enabled me to incorporate the findings from both into the eventual study design. The methodology, rationale and methods used for this phase are explained in Chapter 5.

2.11 Ethical issues when using SGOPE data

The ethical issues surrounding the use of SGOPE data for research purposes are complex and continue to evolve [176–178]. The basic principles of research ethics are based on the four basic tenets intended to protect the fundamental human rights of dignity, autonomy, justice, maximising benefits, and minimising harm. These have been codified in several documents such as The Declaration of Helsinki [179].

The general principle of informed consent has always been a crucial component in traditional research studies, but internet-based research has added new levels of complexity to the ethical implications of using online material as a data source. Obtaining informed consent for the use of this type of data is recognised as being problematic [178,180].

There has also been significant recent debate around expectations of privacy for people posting on social media [180,181]. Making a clear distinction between public and private spaces online can be difficult [182,183]. Concerns exist that individuals could be identified from the posts they make, and that they may consequently suffer harm from some unforeseen use of the data. Potential harms range from unwanted commercial marketing use to profiling that could negatively impact on the poster's life such as future insurance or career choices [184]. Despite these concerns, some studies looking at user attitudes found that social media users were generally positive towards their posts being used for research provided that they were protected from harm and that the research had potential public benefit [180] rather than for private gain [178]. There are also many examples of social media communities that have been deliberately formed in open online spaces to enable individuals to come together to form a voice that can be heard by health systems [29,185].

It is impossible to know either the motivation, or expectation of privacy of each poster in publishing their content, but it would seem reasonable to assume that posters writing on sites that are password protected or restricted to members may have greater expectations that their privacy will be protected [47]. Some sites such as Twitter or reddit make it clear that any content posted on the respective platforms can and may be read or accessed by anyone. They both provide access to their archives enabling anyone to download the posts for research [69,186,187]. SGOPE can be classified as publicly available data [188] but as it was originally collated by the online sites and contains detail of individuals it does not fit the narrower definition of open data which can be freely used, re-used and distributed by anyone [189].

At the time of the design of this study there was a lack of clear guidance from UK Research Councils or other organisations [182,190] as to how best to use it whilst minimising the risk of any harms to the poster. The prevailing view at the time of data collection was that posts made on totally public forums, without any form of password type restrictions were generally deemed to be in the public domain. There was an assumption that if posters were placing posts in such a public place, implied consent had been given [184,190]. Many studies were deemed not to need ethical approval on the basis that the data was publicly posted and publicly visible [178].

In this study I tried to be guided by the core ethical principle of avoiding harm while respecting privacy. I therefore aimed to use sources of the type often described as being in the public domain, i.e., that required no form of log in to view posts. Using sites where no registration or login was required, and where contributors were informed that any posts would be publicly viewable was designed to minimise the risk of harm from a breach of privacy. The methods were also designed to minimise the potential for any form of harm. No IP address or other geographical data was collected, all forms of usernames were removed, and the dates of the post reduced to a year value to minimise any risk of reidentification [183,191]. There was no contact with any of the individual posters, so respondent validation was not possible.

All the sites included in the exploratory study invited posters to submit experience reports for publication on the respective platform. Content from drugs.com [192] and WebMD [193] carried clear messages to posters that posts were publicly viewable and could be read, collected, and used by others. Use of this type of data is covered under the doctrine of fair use [194,195], but I successfully arranged a data sharing agreement with AskAPatient. The drugs.com dataset is freely available to download [196]. Erowid position themselves as working with academics and medical experts and state that they generally agree to research use, although I received no response from repeated requests. ModUp was similarly contacted but the site no longer exists.

2.12 Ethics approval

At the time of the exploratory study, I was informed by the BSREC committee that there was no need for ethical approval. By the time of the second study this had changed and so BSREC approval was successfully sought for the main study (Appendix A)

Chapter 3 P1 Exploratory Study

3.1 Outline

This chapter comprises the methods and results of the exploratory study. Beginning with the aims and objectives, it explains the data collection, methods of cleaning, transforming and analysis, together with the results from both approaches.

3.2 Aims and objectives of exploratory study

This exploratory study aimed to compare findings from both a qualitative and a basic NLP analysis of a sample dataset of publicly available SGOPE data relating to poster's real-world experiences of the effectiveness of Modafinil. The results were then used to inform the design and operationalisation of the main study. The objectives included:

- Qualitatively exploring a small sample of data particularly for context regarding conditions, symptoms, indications of causation, and outcomes.
- Using NLP and corpus linguistics to identify topics and create an ontology of entities, relationships, and causal text.
- Evaluate overall sentiment towards perceived effectiveness using both methods.
- Compare findings from both approaches to evaluate the ability of NLP methods to identify qualitative findings.

3.3 Methods

The study was conducted by following the study design as explained in the previous methodology chapter and as per figure 3-1.

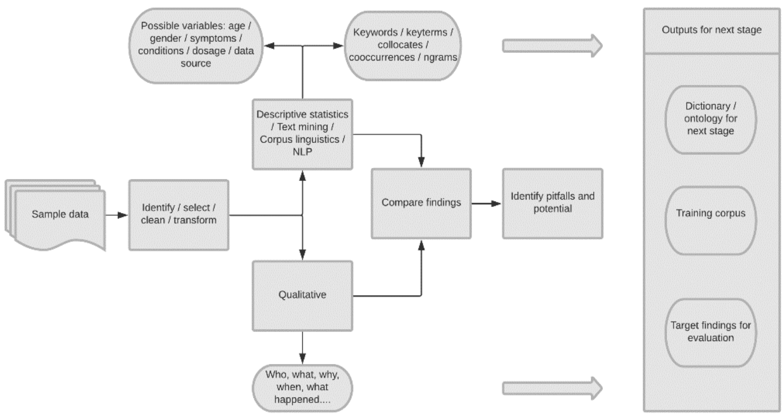


Figure 3-1: Study design for exploratory study

3.3.1 Data selection and preparation

In January 2017, I used various Google searches to identify websites containing publicly available experience of Modafinil use that included a variety of poster types, post lengths and record structure. Publicly available data is defined as where the data was available to view by anyone without any form of login, password or registration being required [184]. Using this type of data and removing all forms of poster identification meant that ethical approval was not required at that time. From the returned results I selected sites containing single comment 'User review' posts enabling comparison of findings across the different data sources, as well as by condition, symptom or another variable. The final selection included five different data sources. AskAPatient, Drugs.com, and WebMD each comprised short condition-based experiences, while posts from Erowid and ModUp were much longer personal experiences with greater detail of symptoms, side-effects, and self-experimentation. Online spaces can be transient and unfortunately the ModUp site no longer exists online, but all the others are still visible. After selecting the sites, individual posts were selected by entering one or more of the widely known name variants (Modafinil, Provigil, Armodafinil, or Nuvigil) as search terms into the identified sites search engine. To ensure maximum validity all relevant data from each site was collected, with no date restrictions. I then used random number generation to reduce the sample size to 260 posts from 5 sites. This volume of data was likely to be sufficient to reach data saturation for the qualitative analysis and be sufficient for linguistic analysis.

Each site had its own data structure with a variety of fields (Appendix B). Age and gender self-definition were optional on each of the sites. To prepare the data, the data from the individual sources had to be 'cleaned' and standardised as far as possible. This was done in a sequence of steps that involved: -

- Standardising field names across sources.
- Translating/encoding coded values: e.g., M/F or male/female
- Standardising numerical ratings scores for experience of Modafinil. Erowid and ModUp had no numerical rating; AskAPatient had a rating from 1-5 and drugs.com from 1-10 for effectiveness of Modafinil, and WebMD had ratings for effectiveness, ease of use and satisfaction, each from 1-5. For the latter, the average of the three scores was calculated. All ratings were standardised to a value between 1 and 10.
- Ages and duration of taking, where given as an identifiable field, were grouped into ranges, and standardised across the sources.
- Posting date simplified to PostYear.

All poster identification was removed, and a unique code allocated to each post. The cleaned data structure is included as Appendix C. Descriptive statistics were generated by first calculating post lengths, before coding and quantifying the included gender, age groups, duration of taking Modafinil, and numeric ratings.

3.3.2 Qualitative analysis

For the qualitative component I took an inductive thematic analysis approach [170] using MaxQDA software [197] using an iterative process of code identification and review as I progressed through the data.

3.3.3 Manual evaluation of effectiveness

Following a preliminary familiarisation reading of the posts I initially categorized text within each post into one of four broad categories, PreModafinil, Acquisition, Dosage and PostModafinil. These categories align with the base state, action and consequence sequence required to indicate a possible perceived causal effect [129,198] (Table 3-1). I then compared the coded sections of each post across the sequence categories to identify the poster's own view of whether taking Modafinil was linked to a causal belief and identifiable outcome.

Table 3-1: Using categories to identify causal text and perceived effectiveness

Sequence	Post categories	Text describing
Base state	PreModafinil	Symptoms + context
Action	Acquisition / Dosage	Took/ did/ prescribed
Consequence	PostModafinil	Effect on symptoms / side effects / context /QOL

Each post was then classified for perceived effectiveness (positive, mixed, negative, neutral, unclear (Table 3-2). Originally intended as a basic three category classification scheme of positive, neutral, and negative, it became clear that further categorisation was needed to cover the range of views expressed. The posts were assessed in isolation; balancing the positive and negative aspects of language used, reported benefits and side effects, and reference to the continued use or cessation.

In order to validate the classification, a second reviewer independently assessed around 20% of the posts (50/260). The initial agreement rate was 68%. All relevant posts were discussed, and an agreed category assigned (Full list in Appendix E). Following this I then classified the remaining 210 posts. For posts which had associated numerical ratings I categorised ratings of 0-3 as negative, 4-7 as mixed and 8-10 as positive. Using chi squared test I compared the manually assessed classification with the poster's own numeric rating. Further validation compared the assessed sentiment with the numerical ratings as entered by the posters.

Table 3-2: Examples of sentiment grading

Grade	Explanation	Example
Positive	Overall positive. Ranged from overwhelmingly positive to indicating that benefits outweigh the disadvantages	"significantly improved my quality-of-life with the only side effect being minor occasional headache" (2400)
Mixed	Both positive and negative effects were reported; unclear as to which sentiment prevailed	"It is a lifeline to me, but the side effects are many and do suck. Though i can honestly say, i don't find it addictive." (1117)
Negative	Predominantly negative, usually regarding side effects.	"… feel really strange shaking overall out of sorts I am not falling asleep at work but feel so weird I'm wired that I need to find something else" (1123)
Neutral	No response or side effects noticed.	"...didn't notice any effect whatsoever, not even a side effect. "(330)
Unclear	Unclear	"After a week taking this drug, I get severe headaches. Is this a side effect?" (542)

3.3.4 NLP

The narrative fields were extracted to create a corpus. Due to the small size of this exploratory dataset, I used a corpus linguistics tool, SketchEngine [199] for the structural analysis of the text. Typical NLP projects return best results from very large datasets, while corpus linguistics can be used on smaller data sets of the size also amenable to qualitative analysis. Corpus linguistics and NLP share some similar analysis techniques [200].

Pre-processing for both NLP and corpus linguistics begins by dividing the text into tokens representing the smallest possible linguistic unit. Each token was assigned a part-of-speech (POS) tag from the English TreeTagger POS tagset with SketchEngine modifications [201]. The full list of these tags is included in Appendix E.

I used case independent word frequency and term extraction to identify the keywords and topics specific to the corpus. Similar to the TF-IDF technique in NLP, term extraction is a measure of the originality of a word within a corpus [202]. It identifies the terms most specific to the text by calculating their frequency in the text compared to the frequency of the same term in the reference corpus. For the reference corpus I used the English Web corpus 2013 (enTenTen13) [203], a corpus of 19 billion words collected from online sources. I extracted the top 500 specific keywords and terms. The top 100 of each indicated the most prevalent topics. The least frequent were used to identify instances of spelling variations or non-words; these were added to the domain specific dictionary intended for use in the next stage of the project.

3.3.4.1 Entity Identification

To identify relevant entities, I used the following POS tokens tagged as nouns:

- Drug Names - both name variations of Modafinil and other drugs; those taken previously, concurrently, or subsequently in addition to some that may have no relevance to the post.
- Condition Names - identifiable condition names were categorised from term extraction analysis. Sleep related disorders were classified in line with the ICSD3 classification systems [204]. (Appendix D)
- Symptoms - symptoms of interest in this study related to fatigue or cognitive issues. Initial dictionary entries were created from common synonyms, with further additions identified from the previous analysis.
- Action - the action of taking Modafinil has two main components: amount and frequency. Terms and phrases to identify both were found within the posts and included in the dictionary.
- Side Effects - term extraction was particularly useful in identifying side effects that the poster described, as patients often use a wide range of terms to describe them that may not map easily to recognisable medical terms.

3.3.4.2 Relationship Identification

Identifying the relationships between the entities is central to understanding the semantic meaning of the text. After using stemming and lemmatization to assign inflected words to the same term, reducing the number of inflectional forms of a word and reducing variants to a common base [205,206] I used three methods to do this:

1. POS tagging of verbs occurring between entities to indicate simple relationships.
2. Collocation analysis [207] to reveal patterns and meanings that may not be apparent from frequency lists or manual reading of the texts;
3. Co-occurrence analysis: this assumes that if two entities co-exist within so many words that there is an underlying relationship between them. Unlike collocations, the relevant words need not be adjacent to each other, but occur within the same unit of text. Co-occurrences can highlight relationships indicating a causal link such as a side effect, outcome event or demonstrate a negated drug event - one which denies a causal relationship between the drug and the event.

To identify possible causal text, I split the corpus into two sub corpora based on the text categories PreModafinil and PostModafinil (see Qualitative analysis above) and used n-gram analysis on each, looking for phrases between 3 and 5 words long that occurred at least five times in the corpora. Where a ngram was ambiguous I examined the collocation and co-occurrence analysis to assist categorisation.

3.3.4.3 Sentiment Analysis

To evaluate sentiment I used the Python 'TextBlob' package [208] to calculate the polarity of each post as a value between -1 (negative) and +1 (positive). Pre-processing included converting text to lower case, removing punctuation, and removal of the default stop words.

3.3.5 Comparing the methods

Comparison of the methods was done on the assumption that the in-depth qualitative analysis would give the most comprehensive understanding of the post content. I performed a manual mapping of each of the 100 most frequent key words and terms from the computational corpus analysis to the themes that emerged from the qualitative analysis. Where a word/term was ambiguous or related to negation, time, or scale they were placed in a separate group.

The comparison of the NLP sentiment analysis to the qualitative categorisation of positive, mixed, neutral, or negative used two comparison scales. The first classifying a 'mixed' result as being in the range +-0.01 and the second widening the 'mixed' range to +-0.05. In both cases a polarity value of 0 was mapped as neutral. I compared the NLP sentiment analysis with the qualitative analysis results for perceived effectiveness of Modafinil as follows: comparison of totals for each type of perceived effectiveness/sentiment; comparison of analysis of individual posts. The accuracy of the post level comparison was then assessed using a confusion matrix.

To assess the methods of identifying examples of causal text I used the subdivision of the corpus into Pre and Post Modafinil codes, mapping each of the 3–5-word length ngrams to the themes and code from the qualitative analysis. Where an ngram could apply to more than a single theme, I again used collocation and co-occurrence techniques to map it to the theme or group for which it was most prevalent.

3.4 Results

The dataset included posts with a total length of 72,427 words (average 279; minimum 15; maximum 2384). Posts from AskAPatient (30-417 words), Drugs.com (15 -204), WebMD (29-358), Erowid (44-2384) and ModUp (125-1030). Posts covered a date range from 2002 to Jan 2017.

Of the posters, 158/260 (61%) identified their gender and 156/260 (60%) included their age, either as an integer or as being within a range. From the two sites with 100% gender identification, there were 65% female posters on AskAPatient and 22% on Erowid. The defined age-groups ranged from under 18 to over 75, with the largest age-group being 45-54 years.

The quantifiable length of time that posters stated they had been taking Modafinil was included in 184/260 (70%) of posts. Of these 34 (18.5%) had taken it for 7 days or less, 31 (17%) 8-31 days, 61 (33.1%) for between 2 and 12 months and 58 (31.5%) for longer than 1 year.

3.4.1 Qualitative analysis

The thematic coding returned 342 codes, which were grouped into eight main themes. The eight main themes are described below.

3.4.1.1 Reason for taking Modafinil

All posts were concerned with finding a solution for symptoms of fatigue, sleep and or cognitive dysfunction. Although Modafinil is only indicated for a single condition within the UK, thirty-three different health conditions were mentioned within this small sample of 260 posts. The most frequent were central disorders of hypersomnolence (mentioned in 26% of posts), depression (22%), sleep related breathing disorders (16%), general fatigue (9%), CFS/ME (7.5%), ADHD/ADD (6%), and MS (6%). Other conditions included cancer, traumatic brain injury, diabetes, epilepsy, fibromyalgia, autoimmune conditions, pain, IBS, hepatitis C, or post stroke fatigue. Multi-morbidity was a regular feature. While many posts referred to a single diagnosed condition, 23% referred to two concurrent conditions, 3% to three and 1.5% to 4.

3.4.1.2 Impact of symptoms

Almost all posts contained detail of how these fatigue or cognitive symptoms affect their lives, emotionally, socially, and practically. Responses to their conditions included fear, desperation, hopelessness, resignation, embarrassment, and guilt:

Life was miserable. I was being treated for depression and had even considered suicide. There was no way out of this rut. [422]

I had resigned myself to life handicapped with fatigue, and I felt really hopeless about it [321]

Frustration was a common theme, often at their own inability to engage with 'normal' life:

I couldn't stand being this form of myself any longer — it's not me [424]

Symptoms were described as having considerable impact on family and social relationships, putting a strain on marriages, partnerships and affecting parenting:

My husband gets sick and tired of me being tired all the time and particularly hates it when I have to have a nap [503]

Before Nuvigal I couldn't keep my eyes open and live my normal life with 3 boys! Now, after Nuvigal I can actually play with my kids and be a normal mother. [2348]

The loss, or anticipated loss of a job featured in 47 (18%) of the posts. In addition, 18 (7%) posters detailed their fear of driving, either because they had experienced falling asleep at the wheel or were concerned that they would.

3.4.1.3 Effectiveness of Modafinil

Posts were classified as follows: 68% positive, 18% mixed, and 12% negative; 4 posts were neutral (see Table 3-3). A total of 181 posts had the potential to include a numeric rating of the effectiveness of Modafinil of which 178 posters completed the rating. The average value (after standardisation) was 7.5/10. I found no significant difference between the posters numeric rating and our assessment (χ^2 3.3419, p = 0.3).

There was considerable variation in the proportion of posters reporting positive effect of Modafinil across the different sites: positive values ranged from 46%-100%, mixed from 0-27%, and negative from 0-25% (Table 3-3).

Table 3-3: Manual assessment of perceived effectiveness across data sources (%age)

	AAP (n=79)	DCUR (n=53)	Erowid (n=41)	ModUp (n=38)	WebMd (n=49)	All sources (n=260)
Positive	46%	72%	78%	100%	61%	68%
Mixed	27%	15%	12%	0%	27%	18%
Negative	25%	13%	7%	0%	10%	12%
No effect	1%	0%	2%	0%	0%	1%

3.4.1.4 Impact of effectiveness on QOL

A recurring topic among those finding Modafinil effective, was how it allowed them to return to what they felt was their personal ’normal’ state rather than enhancing their abilities in any way.

> *This stuff is pretty amazing, i can actually have a normal day rather than fighting just to get through one. It's not what i feel but what i don't feel which is the constant fatigue, without that life has returned to 'normal'. [1388]*

3.4.1.5 Dosage

Of the 141 (55%) posts including text relating to Modafinil dosage the reported dosage taken ranged from 25g to 1200 mg per day in one extreme case. Although clinical guidelines usually suggest 200-400 mg daily [209], there are indications that a lower dose was found to be more effective for some posters, with 17 reporting taking 100 mg/day. Tolerance was described as an issue for some, with 51 (20%) posters commenting on an apparent reduced effectiveness after weeks or months of regular daily use. Some posters reported that stopping taking Modafinil for a few days before resuming a daily dose appeared to restore its effectiveness:

> *After a week or so, effects not as strong and can make you feel paradoxically very tired. Take 2-3 days off, and it will resume working. [2344]*

Whereas others felt it was better to take it only when they felt that they would most benefit from it:

> *I did notice however that I have to take breaks from it for it to remain effective. I now only take it if I have a full day planned and have to go out, otherwise I stay at home and take a nap. [502]*

The posts also illustrated how users have experimented to find a dosage pattern that they find effective (Table 3-4). Almost half the posts contained text detailing the variations in frequency they had tried and those they found most effective. Comments also included the cause /effect results of experimentation of increasing or lowering the dose, taking before or after meals, with or without alcohol and how that impacted on the side effects and effectiveness:

I found if i took 50mg every couple of days, and then 100mg on busy days, it kept the headaches / migraines at bay. [1117]

Table 3-4: Qualitative Analysis: Dosage frequency

	Documents	Percentage	Percentage (valid)
Daily	53	20.38	44.92
As required	30	11.54	25.42
Twice daily	21	8.08	17.8
M-F	13	5	11.02
Unclear (variable)	10	3.85	8.47
Every other day	2	0.77	1.69
Posts with code(s)	118	45.38	100
Unspecified	142	54.62	-
Total	260	100	

3.4.1.6 Side effects

Of the 260 posts, 128 (49%) specifically mentioned one or more side effects they considered related to the use of Modafinil. 34 posts (13%) stated that they did not suffer any side effects at all, while the remaining 98 (38%) did not mention any specific side effect. Across the sample the most commonly reported side effects were headaches (57), mental health/mood related (43), changes to appetite (30), gastric (18), urinary (16), oral (16), skin (15), cardiovascular (11), feeling jittery (10), and difficulty sleeping (10). Other side effects including muscular effects, vision effects, motor function, weight gain, tinnitus, shortness of breath, magnified pain, neuropathy, lupus flare up, swollen tongue, weight loss and increased libido were mentioned by less than 10 posters. The impact of side effects varied, 12 posts described them as minimal, while 13 felt they were temporary, passing within a few days. 9 posters stated that they had stopped taking Modafinil; 8 due to side effects and 1 because of an interaction with an MAOI antidepressant.

3.4.1.7 Acquisition of Modafinil

Detail of how the poster found out about or acquired Modafinil was present in 136/260 (52%) posts, with 82 (31%) stating they were prescribed Modafinil by a clinician, while 54 (21%) discovered it through either their own research or via word of mouth. Difficulties in obtaining it, either within the NHS where its use is restricted to narcolepsy, or in the US where insurance companies often will not cover the cost despite clinicians prescribing it, were mentioned by 37/177 (21%) of those finding Modafinil beneficial. Self-purchasing from online sources was reported by 35 (13%) of posters:

"Now because they say Modafinil is not a bi-polar medicine they refuse to pay for it. I will not be able to afford the $650 a month. Without it I wake with nightmares. It's very sad insurance says they know better than a group of doctors and 10 years of success using a prescription" [2098]

3.4.1.8 Other Interventions

Almost all posts included details of previously prescribed or tried interventions including self-help or lifestyle changes, and any interventions taken in combination with Modafinil. Posts often include comparative descriptors both of effect and/or side effects of the alternative intervention or combination.

I find modafinil it more effective than caffeine although the initial effects seemed to wear off after about 8 hours or so. There are definitely less side effects than with other prescription stimulants such as phentermine or ritalin. [2016]

3.4.1.9 Causality

Perceptions of the poster's experience both pre and post Modafinil were manually identified in 209/260 (80%) posts. Of these, 258 (99%) contained text relating to the effect of taking Modafinil. Identification of causal text was helped by the reported rapid onset of any effect, with many posters who believed it to have an effect, either positive or negative, noticing changes within an hour of taking it.

3.4.2 Comparing qualitative and corpus results

Of the 100 highest frequency keywords 88 mapped directly to qualitative themes, 7 related to negation or scale and 5 could not be classified. Of the 100 highest frequency key terms, 84 mapped directly to the qualitative themes, 7 referred to negation and temporal aspects, and 9 could not be classified (Table 3-5).

Table 3-5: 100 highest frequency keywords and keyterms by topic

Theme	Keyword	Keyterm
Drug	modafinil; provigil; nuvigil; armodafinil; modalert; nootropic; modafanil; modafinal; modvigil; nuvigal; modavigil; moda;	200mg provigil;

Theme	Keyword	Keyterm
Condition	narcolepsy; hypersomnia; apnea; idiopathic; fatigue; fibromyalgia; cfs; insomnia;	sleep apnea; daytime sleepiness; sleep cycle; chronic fatigue; excessive sleepiness; sleep disorder; excessive daytime sleepiness; extreme fatigue; severe sleep apnea; obstructive sleep apnea;shift work; nerve entrapment;
Symptom	sleepiness; sleepy; tiredness; drowsiness; asleep; fatigue; drowsy; procrastination; sleep; lethargy; nap; spaciness; procrastinate; doze; irritable; tired; exhaustion	head fog; daytime sleepiness; sleep cycle; chronic fatigue; excessive sleepiness; term memory; day time; short term memory; anxious state; afternoon fatigue; constant fatigue; brain fog;
Acquisition	reddit; mymodafinil;	prescription drug; sleep study;
Dosage	mg; dose; tolerance; pill;	200mg dose; 200mg pill; full dose; first dose; second dose; empty stomach; 100mg dose; second pill; 200mg provigil; 1st week; drink plenty; daily dose;
Side effect	jittery; jitter; headache; hallucination; irritability; impulsiveness; irritable; itchiness; appetite; nausea; grouchy; clench; bpm;	side effect; dry mouth; smelly urine; heart rate; mild anxiety; slight headache; unpleasant side; bad side; heart beat; negative side; anxious state; jittery feeling
Other drug/ intervention	stimulant; piracetam; ritalin; cpap; caffeine; amphetamine; ephedrine; adderall; adrafinil; phenylephrine; bupropion; ssri; med; pseudoephedrine; caffiene; methylphenidate; fluoxetine; cocaine;	Taking bupropion
Effect	wakefulness; awake; alertness; euphoria; enhancer; psychoactive; nighter; schoolwork; talkative; lifesaver; palpitation; impulsiveness; amped; chatty;	term memory; cognitive enhancer; normal sleep; productive day; mental acuity; normal sleep schedule; mental clarity; positive impact; short term memory;
Outcome		new person; normal sleep schedule; miracle drug; wonder drug;
Negation	didn't; wasn't; hasn't; couldn't; wouldn't; hadn't	i didn’t; i wasn’t;
Temporal		1st week; first dose; second dose; hour period; entire day;

Theme	Keyword	Keyterm
Scale	hyper	
Ungrouped	comedown; sleepless; midterm, cephalon; had	sleep deprivation; side note; trouble sleeping; placebo effect; college student; enhancing drug; study aid; year old male; work day;

3.4.3 Sentiment Analysis

The NLP TextBlob package returns sentiment polarity as a value between -1 (negative) and +1 (positive). Of the 260 posts 188 (72%) indicated positive sentiment, 10 (4%) neutral and 62 (24%) negative. The range of polarity values of posts was from -0.26 to 0.4. Table 3-6 and Table 3-7 show the results of comparing the classification of each method for each post. Matching was accurate in 64% of posts. If we allow for one category difference matching was accurate in 85% of posts. The full list of sentiment agreement is in Appendix F.

Table 3-6: Sentiment analysis confusion matrix [+-0.01]

	NLP							
Qual	>=0.01	+=-0.01-<0.01	0	<=-0.01	Total			
Positive	145	2	4	23	174			Agreed evaluation
Mixed	28	3	3	13	47			1 category different
Neutral	0	0	1	3	4			2 categories different
Negative	11	4	2	18	35			Completely opposite evaluation
Total	184	9	10	57	260			
Accuracy	0.642							

Table 3-7:Sentiment analysis confusion matrix [+-0.05]

	NLP							
Qual	>=0.05	+=-0.05-<0.05	0	<=-0.05	Total			
Positive	130	26	4	14	174			Agreed evaluation
Mixed	24	9	3	11	47			1 category different
Neutral	0	0	1	3	4			2 categories different
Negative	9	11	2	13	35			Completely opposite evaluation
Total	163	46	10	41	260			
Accuracy	0.588							

Table 3-8: Example posts with conflicting sentiment analysis results

Manual grade negative – NLP grade positive	Manual grade positive – NLP grade negative
*First day was great (started **ar** 150 dose) then falling asleep during day. Increased to 250, didn't fall asleep during day but very nervous and couldn't sleep at night. Going to breakup dosage to see if that helps SideEffects Itching, **cant** sleep at night [1146]*	*I have sleep apnea and also Multiple Sclerosis. I use the 100mg tab, but not on a daily basis. I use this when I feel tired from my MS. I recommend this for MS patients that tend to have no energy when they wake up in the mornings. It doesn't seem to affect my symptoms of MS; i.e., the tingling in my feet or legs. It just gives me energy to get **throught** the day, when I need to do what I need to do. I see my Neurologists for the medication prescription. [5037]*

The 3-5-word ngram analysis on both the pre-Modafanil (35) and post-Modafanil (106) text generated ngrams classified into the 8 themes and 6 categories reported in Table 3-9.

Table 3-9: PreModafinil and PostModafinil 3-5-word ngrams grouped by theme

Theme	PreModafinil	PostModafinil
Reasons	i was diagnosed with; obstructive sleep apnea; chronic fatigue syndrome; i have been; sleep apnea and; i suffer from; at the age of;	
Symptoms	i have been; that i was; i found myself; i have to; i suffer from; for the last; i was a; at the age of; and i was; i used to; i was still; i wake up; i had to	
Other interventions	i started taking;	i was on;

Theme	PreModafinil	PostModafinil
Acquisition	i went to; i was prescribed; i decided to	
Dosage		early in the morning; I don't take it; i have been taking; i started taking; i take it; i don't take; i have found that; i took it; i have to; in the morning; on days that i; if i don't; when i don't; to take it
Side effects		with no side effects; i didn't notice; and i was; don't have; the next day; i don't feel; the first time i took; the side effects;
Effectiveness		i am able to; I don't feel; to be able to; get out of bed; i didn't notice; first time i took; i began to; and i was; don't have; the next day; go to sleep; i feel like i; i have found that; i was able to; i don't think; i have not; i felt like i; i used to; if i don't; that i could;
Outcome		to go to; to be able to; was able to; that i could; i felt like i;
Temporal	all the time; during the day; through the day; for the last; in the morning;	a few days; first time i took; during the day; for a few days; in the morning and; through the day; the first time
Sequential	for the last; i used to; i was still;	at the same time; as soon as i; for the first time; for a few days; the next day: i have found that; i used to; on days that i; i had to; if i don't; if i need to; the first time i took; the first time;
Negation	i don't; i didn't;	don't have; i can not; i didn't feel; i didn't have; i didn't notice; i don't feel; i am not; i did not; i don't have; i don't know; i have not; i do not; i was not; it does not; n't be able to;
Confirmation		i was able to; i felt like i; i want to; i used to; i was on; i had to; it was a;
Ungrouped	a lot of	a lot of; hours of sleep; i need to; i have been; that i could; that i had; that i have; that i was; the rest of the; to be a; to take a
Causal ngram		i began to; i have found that; on days that i; if i don't; when i don't;

As with the keywords and keyterms many of these ngrams correlated with and mapped onto the themes that emerged from the qualitative analysis. Others related specifically to temporal, sequential, negation or confirmation text that could be used to identify phrases inferring causality. The frequently occurring ngram *"I have found that'* seen in 9 posts was used to describe ways of taking the drug to maximise the effectiveness. Examples of generic ngrams and the context in which they were used are given in Table 3-10.

Table 3-10: Example ngrams in context

ngram - I have been (PreModafinil) - categorised as 'Reason for taking'
I have been battling MS Fatigue to the point of almost thinking of quitting my job, but desperately need the money.
Depression **I have been** working for many years with one combination after another of medications for bi polar disorder.
Before Nuvigil **I have been** suffering for the past 3 years or so with marked fatigue.
For the last few years **I have been** taking medicines to calm me down and ease my stress levels.
With that said, recently **I have been** back and forth to the doctor for four months now.
I have been fighting this for as long as my memory will take me.
I consider my own experiences to be significant in that **I have been** on the SSRI cipralex (celexa, escitalopram) since age 19, having experienced bouts of diagnosed major depression in my late teens.
I have been taking Provigil for about 9 months now after my sleep disorder kept me awake for 8 days even after being on a 6 mg dose of lorazepam to sleep at night for several years.
In addition to Provigil **I have been** on Effexor XR at 150mg/day for my mild depression.
I have a severe lack of motivation and **I have been** diagnosed with ADHD.
I have been diagnosed with and suffering from Idiopathic Hypersomnia for the last six years.
I have been through 2 sleep studies and I wasn't quite a match for the CPAP machine but according to my doctor at the Mayo Clinic, there is obviously something wrong with how tired I am and how easily I can fall asleep.
ngram - I used to (PreModafinil) - categorised as 'Symptoms'
I used to fight sleep all day at work, it got to the point where I was staying home because I just couldn't stay awake.
After the buzz wore off, my life became normal, which was a great improvement over the constant feelings of lethargy and helplessness **I used to** feel.
As someone who works online, as a writer and retailer, **I used to** find myself researching an article one minute, and somehow snapping out of a haze a few hours later.

ngram - I have been (PreModafinil) - categorised as 'Reason for taking'
I used to drink coffee for this kind of thing, but tolerance builds up quickly and by the end of exams I'd be drinking a few cups a day and it made me feel no good.
I used to be a PhD student who was heavily dependent on Adderall for a cognitive and motivational boost.
ngram - I have found that - (PostModafinil) - categorised as 'Dosage'
I have found that if I don't take it on the weekends that it works better.
Overall, I have found that effects of Modafinil, for me at least, are extremely subtle and almost unnoticeable until I start to think back and examine the things that I have done on a given day.
I have found that a very effective remedy is to take a couple of co-codamol tablets, which each contain 8mg of codeine.
I have found that my own lack of worry or guilt in situations like these prevents people from becoming suspicious - nobody batted an eyelid.
The one concern I have is that I have found that cutting the dose (as I did once for several days when I didn't place the online order in time) seems to have a dramatic negative effect.
I have found that Modafinil gives me a very clear mind for problem solving.
I have found that if I skip the workout, I don't have as much energy throughout the day.
I have found that if I eat a lighter lunch, the dip is not as bad.
I have found that overall mod just works best on its own.
ngram - the first time (PostModafinil) -categorised as 'Effectiveness', 'Temporal' & 'Sequential'
The first day on Nuvigil I felt like I had never felt before: My mind felt awake for the first time in what seems like forever.
I feel my age for the first time ever!
I have done some research and found that taking a "drug holiday" or going a day or two out of the week without it will help it stay just as potent as the first time I used it.
Like many others, the first time I took it was great!
The first time around, I nearly drove my family crazy with my talking and myself crazy trying to keep my rapid thoughts to myself.
The first time I took it, I did not have the headaches, smelly urine, post nasal drip, fuzzy vision, or muscle aches.
All I can say is that I'm now a 4.0 college student and I feel like I'm actually awake for the first time in my life.
The first time I took this (prescribed for obstructive sleep apnea) I thought "wow, this is the answer!
At first I was sceptical that it had been the Modafinil that had caused the happiness because I tend to go through short bursts of depression and happiness and I assumed

ngram - I have been (PreModafinil) - categorised as 'Reason for taking'
that I had just been on a good day the first time I took it but looking back I haven't had any significantly bad days while I was on Modafinil.
I am dramatically more productive at work and for the first time in my life I feel capable of planning for the future.
I will never forget how deep my mind sank that week, it was the first time I'd ever felt truly depressed - not even extended family deaths or the comedown from 220mg of pure MDMA was as bad as how I was feeling that morning.
When I took it for an exam the first time it was amazing, I took 400mg at around 8 and stayed up the entire night studying with no problems.
I was captivated by his work, which really was excellent (we both received 1sts for our efforts), but for the first time that day I was ever so slightly distracted while I was reading.
The first time, I seriously wondered about the efficacy of the things...
I think it important to note that I have never taken it every day and usually never take more than 200 mg. 600 mgs, which I took for the first time today, really has me jacked.
The first time I tried Modafinil the effect was immediate.
Right from the first time I took Modafinil not only did I feel better, but I could tell how much more I was worried about my work, the amount of detail I would put into my projects even shocked myself.
The first time I took modafinil I understood what all the hype was about.

I was able to match ngrams to the expression of causal analysis identified by the qualitative analysis (Table 3-11).

Table 3-11: Examples of Causation Reason and Consequence

Document	Causation: Reason	Causation: Consequence
1063	***I used to*** fight sleep all day at work,	Once I started Nuvigil I have not had this problem at all, ***I feel like I*** have more energy, and my mental alertness has improved 100%.
1090	I've tried 44 anti depressants and only got about 25 percent relief	Nuvigil obliterated my depression and eased my anxiety by about 60 percent
1065	Prior to provigil, I would regularly fall asleep at work or in meetings. I was afraid to drive alone for more than an hour for fear that I would fall asleep driving. (I had many close calls!).	This drug has made a significant improvement in my quality of life.

1065	I ***had felt*** (just weeks earlier) that I could not go on any further.	***I can keep*** my job and haven't wrecked any cars.
1136	constant feelings of lethargy and helplessness ***I used to feel***.	This drug has returned my pre-MS life to me. ***I can fully f***unction on the job
1207	I forgot to take it one day	and could not stay awake and could not stop eating, just like prior to starting Provigil.

3.5 Discussion

Within this exploratory study of the unstructured narrative post content, both methods successfully demonstrated how the majority of posters with a wide range of conditions found Modafinil effective in reducing fatigue or cognitive symptoms.

In performing the human based qualitative study first, those findings acted as an informal benchmark for the automated NLP study. The eight themes generated reflected the main aspects of patient experiences of an intervention. The study also explored the detailed context that was often included within the poster's evaluation, including the reasons for starting or stopping using it, comparisons with other medications that they may have tried or moved onto, side effects and tangible or intangible effects on their quality of life.

The sample size was too small to realistically expect good results from the NLP analysis, but by using the corpus linguistics tool which used some methods found in a full NLP approach I could demonstrate how an NLP methodology could be used on a much larger scale to both extract topics/themes, expressions of perceived causality and evaluate effectiveness from unstructured text.

As with a recent paper comparing grounded theory with topic modelling on survey data [210], the NLP based methods successfully identified many of the qualitative findings, demonstrating how this form of data has the potential to identify effectiveness and the topics discussed within the posts. In terms of sentiment analysis, the results highlight some of the current issues with NLP methods. Although both methods show a majority of posters finding it effective for them, the confusion matrices (Tables 6 &7) highlighted some of the issues with applying generic sentiment analysis tools to health-related data. Rule based methods that determine sentiment are based on a lexicon of prelabelled words and the accuracy of the results is heavily dependent on the data that the model was trained on and the words that are considered important to that model. The majority of the existing generic

NLP sentiment analysis tools were trained on either film, restaurant or Amazon product reviews as these represent some of the largest shared annotated sentiment resources [145]. Looking at some of the posts with opposing categorisations (Table 3-8), demonstrates how many of the concepts that posters describe in their evaluations include stopwords or words that may not be evaluated as expressing sentiment. Improved accuracy will require the development or use of a domain specific model.

3.5.1 Identifying causal inference

Identifying causal text requires showing temporality; the effect occurring after the cause. Dividing the corpus into pre and post intervention by tagging the tense of tokens facilitated this classification, while ngrams and other POS tags helped me to identify sequential events.

One of the issues of identifying causality in any kind of non RCT study has always been in differentiating between correlation and causation [211]. Identified patterns and correlations can indicate that 'something is happening' but not necessarily explain 'why' [212,213] as it does not differentiate between the causes of patterns, whether they are true, coincidental or as a result of bias. Increasing the volume and range of data may achieve a higher degree of precision and external validity [214] and while summarising and visualisation may be useful in analysing SGOPE datasets, correlation is not the same as causation and on its own it is unlikely to be robust enough to add to an evidence base.

The Bradford Hill criteria of strength, consistency, specificity, temporality, biological gradient, plausibility, coherence, experiment and analogy are the standard method for establishing whether or not epidemiological associations are causal [215] and have also been applied to user generated data [216]. In this study, strength is demonstrated by how almost all posters reported an effect, either positive, negative, or mixed. Consistency of findings across populations is shown by using multiple data sources and including patients with a wide range of conditions. The reported rapid onset of effect shows specificity and a biological gradient, with the cause/effect sequencing showing temporality.

The purpose of this research is not to provide a statistical proof of effectiveness across the whole patient population, but to generate a better understanding of the patient experience of using Modafinil, by exploring individual patient's perspective of whether it is effective for them.

However, no matter how accurately causal text is identified, the possibility of a placebo effect, recognised as a powerful factor in a patient's assessment of effectiveness both in and out of trials [217–219] means that it is impossible to tell how much of the sentiment towards effects, either positive or negative, is due to such an effect rather than the Modafinil itself.

3.5.2 Strengths & Limitations

Using content purely from the public domain is both a strength and a limitation. Although the easiest to access, it may not contain the richest patient experience data, which may be posted on sites requiring a 'login'. However, using public domain data enables future replication. The study validity is increased by using a diverse range of data sources. Using multiple sites can mitigate emotional contagion and increase the representativeness of the populations while the scale of the data being analysed should negate the problems of an individual post being incorrectly classified or missed. Although there will always be an element of the unknown about the motivations and authenticity of such posts, analysing them on a large scale rather than just a small subsection, can negate the impact of those individuals or organisations who might try to create an inaccurate impression, while techniques are continually being developed to identify spam or non-genuine posts.

As the content is generated entirely by the poster, SGOPE relies on the poster's self-description of their condition, which may include self-diagnosis, rather than that of a clinician. Reporting of symptoms and outcomes may not be as accurate or complete as it could be although this limitation can apply to any form of self-reported data, whether in a trial, clinical encounter or online. Self-reported data, especially on hard to measure factors such as fatigue and cognition is subjective, but generally reflects the normative value of the patient.

The natural, non-clinical language used within unstructured text can contain valuable information that may remain unexplored in a clinical or research setting [95], but it can also contain many spelling or grammatical errors as well as slang terms or colloquialisms that are problematic even for domain specific NLP methods created for electronic health records (EHRs) [220].

3.6 Conclusion

The findings from this exploratory study show that user generated patient experience in the form of SGOPE is a rich resource for evaluating real world effectiveness, understanding patient perspectives, and identifying research gaps. Although the two methods are very different, both methods successfully identified the entities and topics contained in the posts. In contrast to current evidence, posters with a wide range of other conditions found

Modafinil effective. Perceived causality and effectiveness were identified by both methods demonstrating the potential to augment existing knowledge. Although much work is needed to refine the techniques and address the challenges identified, my comparison suggests NLP can be used to look beyond the literal meaning of the words, gaining an understanding of how posters assess the effectiveness of a healthcare intervention and the outcomes that they value, on a much greater scale than is possible from qualitative studies.

Chapter 4 Umbrella Scoping Review (P2)

4.1 Chapter Outline

With the benefit of having performed the exploratory study, this chapter reviews the current state of the art of health research that utilises SGOPE data to inform the design of the main study.

4.2 Background

As the number of individuals using the internet for health-related purposes continues to rise, there has been a corresponding increased interest in exploring this online user generated content as a data source for health research [146]. Although the use of health-related social media as a data source is a relatively new subject area, it is being actively researched across many other disciplines, including computer science, sociology, philosophy and business. The volume of published literature is growing rapidly and includes both academic and grey sources, but as yet there is little literature bringing together the developments in the area [35,140].

4.3 Aims & Objectives

The main research question for this review is "How and for what purposes is SGOPE data currently being utilised within health research?" The sub-questions are listed below (Table 4-1).

Table 4-1 Research sub questions

RQ1	Which sites / platforms are being used as data sources?
RQ2	What purposes is SGOPE data used for?
RQ3	What tools and methods are being used in the studies?
RQ4	What are the knowledge gaps and areas of future research needed?

4.4 Methods

4.4.1 Study design, reason & justification

The study takes the form of an umbrella scoping review. Intended as a state-of-the-art review it looks at the relevant research, debates and contributions in a field, summarizing trends, agreements and debates to provide a synthesis of current thinking [221]. This type of review was chosen over a scoping review, as although they both take a broad overview, bringing together evidence from a wide range of disparate sources, a scoping review is often used to identify gaps where further knowledge is needed [222] whereas the aim of this review is to assess the methodological landscape in order to utilize the most appropriate tools and techniques in Phase 3.

As with a scoping review there is no formal critical appraisal of the literature [223]. The umbrella review process was conducted following the methodology and included elements as suggested by the Joanna Briggs Institute [174].

4.4.2 Search strategy

I searched the following databases: Medline, Embase, PubMed, PsychInfo, Web of Science, ACM and IEEE Xplore, as well as Google Scholar, Twitter, Google and other text or opinion literature. Additional literature, both published and 'grey' was added from reference lists and an existing bibliography. Searches were conducted on January 21st, 2021 with an alert set up to notify me of any more recent publications.

4.4.3 Search terms

Searching across disciplines, especially in an area where the terminology varies rapidly means that it is difficult to use tightly defined search terms. Many keywords relating to SM have yet to be indexed so it is not possible to rely purely on MeSH terms for searching [146,224]. The original search terms were based on keywords that were clustered around the main areas of setting, analysis, content, usage, and methods (Appendix G). Keywords were combined in various options and the searches conducted as an iterative process, repeated as further search terms were identified to optimise the efficiency and targeting of the process. Wildcards (*) were used where possible to pick up multiple word endings or ambiguities over hyphen usage.

Papers were included if they were any type of review: systematic, scoping, literature, or general that included the use of SGOPE data for health research. Non review papers, those not referring to SGOPE usage, not health related, entirely mathematical, or statistical, not in English or published before 2015 were excluded. Table 4-2 defines the context of the review.

Table 4-2: PICO / PICoT concept

Population	SGOPE across all forms of social media, general, health specific, blogs etc.
Phenomena of interest	Any analysis of SGOPE for health research purposes
Context	Looking at types of use and general methods
Outcome	Any type of health experience
Type of study	Review paper

4.4.4 Study Selection

After duplicate removal, I carried out the initial screening following a three phased approach. Using the Rayyan screening tool [225] to assess the reviews, papers were initially classified as include, unsure or exclude based on the title or headline. Second level screening involved reading the abstract or first paragraph of those not initially excluded. Full texts were retrieved of all relevant papers. Although critical appraisal is not required for a scoping review, the reviews were informally assessed using the questions from the Confidence in the Evidence from Reviews of Qualitative research (CERQual) [226] appraisal tool (Table 4-3) to ensure their suitability for inclusion.

Table 4-3: Components of CERQual appraisal tool (GRADE CERQual, 2017)

Methodological limitations	Are the methods suitable for this project?
Relevance	Do the findings relate to the research question?
Coherence	How well does the data relate to the finding?
Adequacy	Richness & quantity of data supporting the finding

The final selection of included papers was then collated into a marked list on the Web of Science database for basic bibliometric analysis.

4.4.5 Data Extraction/ Analysis

The data extracted from each review included the title, author(s), date of publication, journal, keywords, review type, objectives, research questions (where stated), number and type of included studies, settings and population studied, data sources, date ranges of included studies, key findings, future research needed if identified, and strengths and limitations if included. I also classified the papers as focusing on either the use of SGOPE data, the methods used to analyse it, or both if this was the case. This enabled me to analyse the reviews in line with the research questions. Frequency analysis was performed on the author generated words for each included paper, and a word cloud generated.

4.5 Results

Of the 1759 records initially identified from the searches, 58 were included in the final review. Details of the included papers are summarised in Table 5. The PRISMA flow diagram (Figure 4-2) shows the number of papers included and excluded at each stage of the process.

4.5.1 General characteristics

The 58 included reviews covered the period 2015 to 2020 with the reported number of associated papers published increasing each year, especially since 2017 [67,69,159,227–229]. This is illustrated in the breakdown of included review papers by publication year (Figure 4-1).

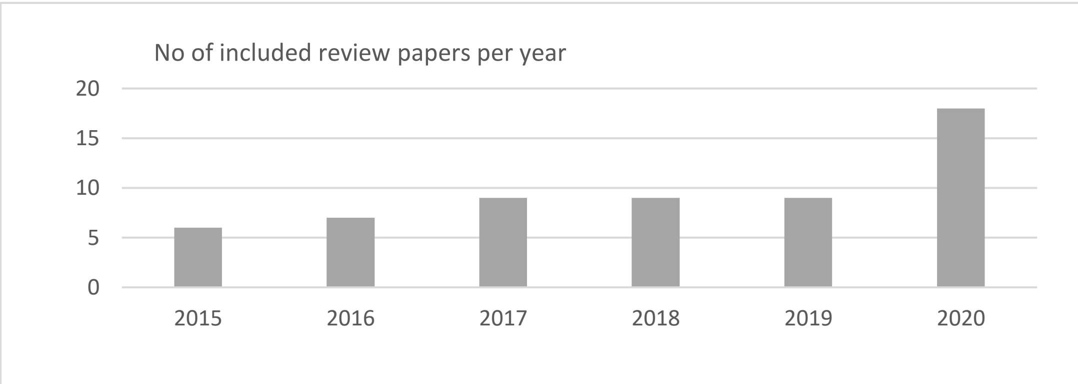


Figure 4-1: No of included review papers by year

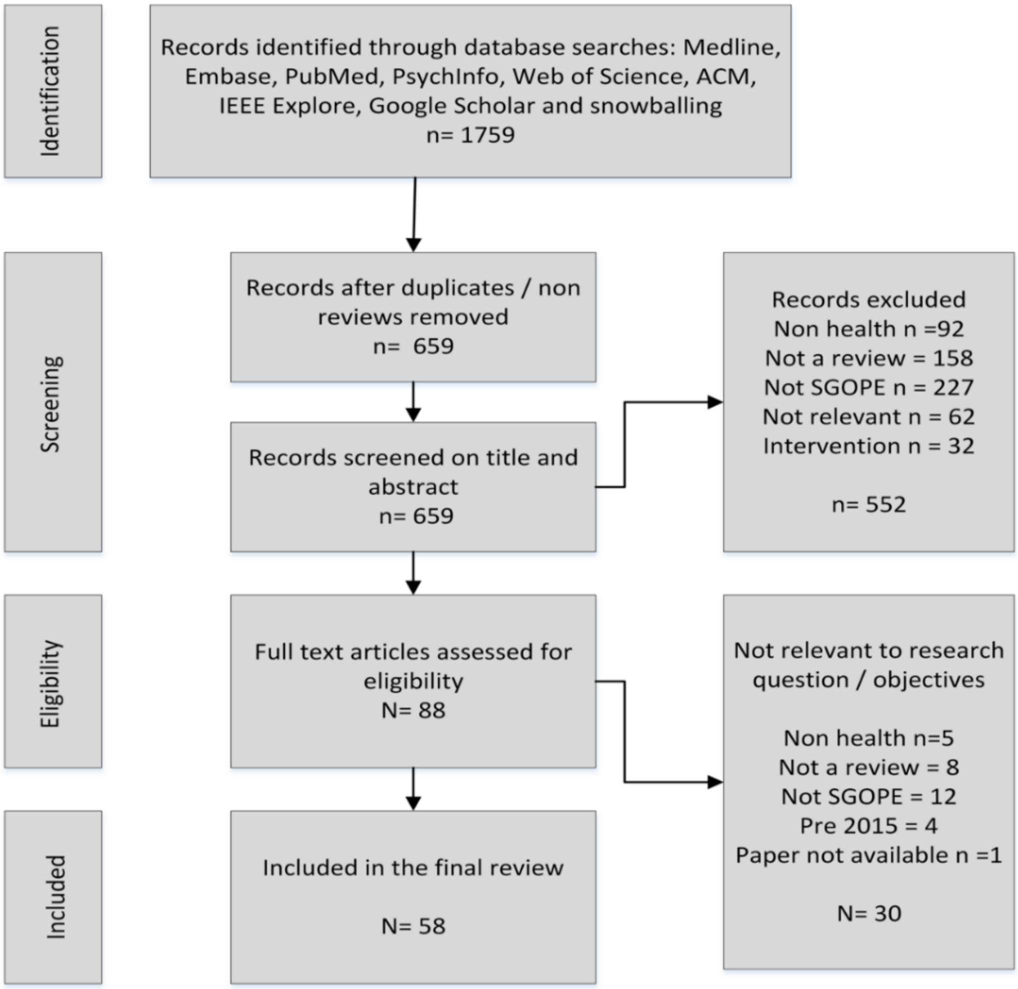


Figure 4-2 Prisma flow diagram

The included studies came from wide range of journals, although only six journals provided more than one review. Journal of Internet Mediated Research (JIMR) contributed 11/58 (19%), Yearbook of Medical Informatics 7/58 (12%), and IEEE 3/58 (5%), while PloSone, Journal of Biomedical Informatics and the International Journal of Qualitative Methods each provided 2/58 (3%). The other 31 papers originated from individual journals from a wide range of research areas (Appendix H).

The interdisciplinary nature of the topic is reflected in the included tree map of the research areas that the papers are from (Figure 4-3). One of the larger reviews analysed the discipline of each of the first authors of the 414 papers; finding that in 90 papers (22%) they were either from a non-health or unspecified background [67].

Figure 4-3: Research areas of included reviews

The general characteristics of the 58 papers including the aims, health condition of interest, data sources, review type and number of included papers in each study are shown in Table 4-5. In terms of the review types, almost half of the reviews, 27/58, were described as systematic, 12/58 scoping, 15/58 general, with one each of narrative, critical, survey and bibliometric.

The number of included papers in each review ranged from 5 [230,231] to 3419 [229], with an average of 118.

In line with the inclusion criteria the selected reviews cover two main areas; how spontaneously generated data is used within health research and the methods and tools that are used to analyse it. Just over half (34/58) of the reviews covered both questions while 13/58 were primarily focused on uses and 11/58 mainly focused on the methods used (Table 4-4).

Table 4-4: Categorisation of the main purpose of the review

Uses	(22,29,36,41,46,56,57,61,65,66,69,70,74)
Methods	(20,28,30–34,43,50,58)
Both	(7,19,23–27,35,37–40,42,44,45,47–49,51–53,55,59,60,62–64,67,68,71–73,75,76)

The word cloud of the individual author generated keywords illustrates the range and frequency of the intended purposes of the included reviews (Figure 4-4).

Figure 4-4: Word cloud of author generated keywords

4.5.2 RQ1: Which sites and platforms are being used as data sources?

Twitter has been by far the most utilised data source, although a wide variety of general social networks and disease specific communities have also been used. Six (10%) of the reviews looked entirely at Twitter based studies [35,69,187,232–234]. A further eighteen (31%) reviews included a wider range of sites but reported that Twitter was the most frequently used source [74,77,78,96,145,153,160,167,175,227,228,235–240]. Both general health sites and disease specific communities covering a wide range of conditions were also widely accessed [54,141,142,144,145,153,167,175,237,241–243]. One review focused entirely on looking at the potential of blogs as a qualitative data source [244]. Six reviews (10%) used both SGOPE and electronic health records (EHRs) [66,143,230,245–247] while Abbe [141] used a combination of online posts, EHR's, biomedical literature and qualitative studies.

Four reviews highlighted that although there were exceptions, most of the individual papers within their reviews used data from a single source [74,160,175,248].

4.5.3 RQ2: What purposes is SGOPE data used for

The use cases for SGOPE data to date extend from improving public health at a population level to fine grained understanding of patient perspectives. The aims, outcomes and key findings from the included reviews are summarised in Table 4-6.

In terms of the specific health topic of interest, 22/58 papers included any health condition [35,66,67,137,140,143,145,146,167,187,224,228–231,233,236,240,244,246,249,250]. 12 focused on mental health conditions [79,141,175,237–239,242,243,247,251–253], 9 on adverse drug reactions (ADRs) [74,139,142,234,241,245,248,254,255], 4 on infectious diseases [77,78,160,235], two each on chronic disease [54,256], substance misuse [96,232], public health [159,227], breast cancer [153,257] and with one each for symptom identification [144], use of complementary and alternative medicine (CAM) therapies [258] and the reasons for existing use by health researchers [69].

As a retrospective surveillance tool SGOPE has been used to capture public reaction to health events in terms of emotions and fears [175,242], knowledge [54], attitudes and behaviours [159,160]. Karmegam [239] looked specifically at studies evaluating the potential of SM data to understand the emotional and psychological impact of unforeseen natural disasters in a community. Several reviews focus on using SGOPE data to monitor behaviours, communication patterns and spread of health related concepts, particularly relating to infectious diseases [77,78,159,160,235]. Both the speed and accuracy of tracking are seen to improve on existing surveillance and signal detection systems, although most conclude that SM surveillance should currently be complementary to existing systems rather than replace them [74,77,142,235]. Analysis of SGOPE data has been used to understand the various network mechanisms of information spread, the topics that are discussed, and to identify trends or patterns within the conversations [54,67,69,74,141,153,160,228,229,232,237,239,248,251,258].

A study on chronic disease collated qualitative studies exploring how people shared knowledge within the communities to show how the distinct characteristics of online spaces helped patients self-manage their long term conditions in ways that are difficult to replicate off line, and how these spaces were filling an unmet need for information and or emotional support [54,137].

One of the most frequent use specific use cases was as a new source for identifying adverse drug events (ADE) or reactions (ADR) [74,139,140,142,241,245,248,254,255]. Identified advantages over existing sources include earlier identification [139,142,245,248], reduction of associated economic costs and fatality [255] and the highlighting of ‘mild’ adverse events that may not be seen as serious enough to report through existing routes. Golder [248] found that the prevalence of adverse event reporting on SM ranged from between 0.2 to 8% of the posts, with ‘mild’ events being over represented, while ‘serious’ ones were under represented as compared to other ADE discovery methods. Comparisons show that SGOPE data is generally in concordance with other regulatory sources for most adverse events [74,248], but that at this early stage of method development that it should be used in conjunction with other existing methods [139,245]. Combining SGOPE with EHRs and omics data is seen as an essential method of detecting and predicting ADRs [254]. The additional context from the patient experience narrative adds to existing post-marketing surveillance of interventions [74,140].

Two reviews looked at the misuse of prescription medicines [96,232]. Kim [232] used findings from existing Twitter analysis to create a typology of SM big data analysis on the topic based on the four conceptual dimensions of poster characteristics, communication characteristics, predictors and mechanism for the discussion of problematic use, and the psychological or behavioural consequences of discussing it on social media.

Other use areas included assessing the opportunities, benefits, challenges and limitations that using SGOPE data might offer healthcare providers and researchers [77,137,141,167,187,229,240,242,250,252,253,255]. Benefits identified included providing a new channel for hearing patient perspectives [141,153,242], faster data collection and reduced costs [77,153], improved support for self-management [54]. Zhang [229] categorised SM based papers by their role in public health, with the most frequent use case being as an interactive intervention tool aimed at modifying risky health factors. Classifying studies into five categories encompassing education, disease modification, diagnosis, support and management, Patel [256] evaluated the impact of social media use on outcomes across a range of chronic conditions, concluding that few studies suggested any harm from its use and that as a data source it had tremendous potential to improve patient care. Drewniak [137] looked at the risks and benefits of using patient narratives for patients, relatives and HCPs, finding that they were a promising way of improving patient understanding of their health conditions capable of impacting behaviours and outcomes.

Ru concluded that with improvements in analysis methods, findings from SGOPE would be able to generate new research questions around effectiveness, ADRs and health related quality of life [146].

Vilar [139] evaluated SGOPE as a method of identifying drug-drug interactions (DDI). They conclude that existing DDI resources such as DrugBank, Micromedex and DDI Corpus, although good as knowledge or evaluation bases show little consistency, and that SGOPE had the potential to be instrumental in creating knowledge sets and identifying unknown DDIs.

Calvo looked at the ways and levels that NLP could be used within mental health, including triaging people at risk and diagnosis of specific conditions. At a post level, emotions and risks can be identified, temporal changes can be tracked at the author level, and general trends in sentiment and attitudes established at a population level [175].

One review identified how language markers, such as higher use of pronouns, can be indicators of altered mental state or suicide ideation [224]. Using predefined semantic vocabularies allowed the identification of posts indicating both medium and severe mental illness [140].

Yin [140] looked specifically at SGOPE data as a route to understanding poster experience of health issues, concluding that it gave insights into health factors that often were not recorded in EHR systems. They summarised 103 papers into 5 research categories; those characterising health issues and patients, prediction of events such as suicide, the correlation between SM posts and existing data collection methods, those characterising drug usage/ adverse events/ misuse and detecting sentiment about major health events such as post-partum depression and how this impacted on posting behaviours. Recognising that symptom discussion is large component of SGOPE data, one review focused on papers for symptom extraction [144]. Understanding how symptoms cluster is a recognised knowledge gap [259]. While pain and fatigue were the most common symptoms that were identified, many of their included papers identified symptoms from 10 of the 12 symptom categories, concluding that SGOPE data could help with faster diagnosis and understanding issues such as the recent opioid crisis and pain management.

4.5.4 RQ3: Analysis methods identified by the reviews

Analysis methods have varied widely as new tools and techniques have been developed, and the reviews reflect this [241]. Eleven reviews focused on the methods utilised to analyse this type of data, while 34 looked at both uses and methods (Table 4-4). A study covering the years 2003 to 2017 highlighted the absence of specific trends in either approach, evaluation or performance [144].

Many papers, even recent ones are still using traditional qualitative [54,229,244,256] or mixed methods [160,233,234,251] of analysis on small quantities of data. Of the 42 papers in the Patel review [256], only 3 analysed over 1000 posts, with 26/42 analysing less than 100 texts. Other reviews included papers using a mix of manual and machine learning methods [78,232,235,241,257,258]. Abbe [141] argued that while the debates about qualitative and quantitative analysis continue, the exploratory yet highly automated approach of natural language processing (NLP) can bridge the gap, offering the best of both worlds.

Among the analysis methods used, sentiment analysis was the most commonly utilised [35,66,79,140,143,145,175,187,232,239,240,242,248,252,253,258]. The review found that much of early sentiment analysis was often performed on small volumes of text, using qualitative or content analysis methods [96,234]. Developed originally as a marketing tool for business to understand consumer opinion towards their product [35], sentiment analysis has frequently been used to identify emotions that can signify a posters thinking and mood when trying to identify potential suicide risk [35,79,175,238], to track ADRs and to interpret patient reviews of health care services [145]. Simple automated content analysis has used lexicon based keyword techniques such as the LIWC (Linguistic Inquiry and Word Count) text analysis tool to count the frequency of keywords within the text [257] or compute the percentage of positive or negative emotional terms in a text [175,235].

Machine learning methods have a myriad of different algorithms and techniques of varying levels of complexity in various stages of development. At a basic level they can be divided into either supervised (classification) or unsupervised (clustering) methods. Classification methods were often rule based, looking for a predefined words or patterns of text and the accuracy of the model is heavily dependent on the initial parameters in the choice of words or expressions [143]. The majority of the machine learning studies to date have used supervised methods. Common classification algorithms include Support Vector Machines (SVM), Naive Bayes (NB), Decision Trees (DT) and Random Forest (RF). All these and others are frequently mentioned in the method discussions although SVM was the most popular [77,145,227,249,252]. Gupta [167] noted that SVM was the most promising method for binary classification tasks. Unsupervised techniques using topic modelling which do not require large amounts of labelled data are beginning to become more prevalent, especially for identifying themes and topics within large quantities of text [235,257] but are less frequently utilised [187]. A comparison of all datamining techniques found that they all had various strengths and weaknesses and that the research objective and data should guide the choice of method [249].

The methods of both SGOPE and clinical NLP (looking at the unstructured text in EHRs etc) have similar issues and purposes, but although automatic methods of processing are developing, the unstructured nature, noise, domain specific content, problems with language usage, understanding semantics and the complexities of informal speech mean that there is still a lot of work to be done in developing methods to maximise its usage [66,167,247]. Sarker [96] categorised the methods currently used to identify and monitor such use, concluding that there was still a lack of datacentric pipelines, and proposing a new method based on shared annotation guidelines and labelled datasets. One proposed solution to the issues of 'noise' and irrelevant text within SGOPE is to use a combination of methods to refine the content by using a binary classification method to exclude any irrelevant content and then topic modelling to identify themes from the useful content [66,78,96].

The latest developments in NLP move from rule-based systems to deep learning [66,140,145,187,224]. These aim to improve on the semantic level of understanding possible, by using language models such as word embeddings and distributional semantics [245]. Based on artificial neural networks, deep learning uses 'hidden layers' to extract more detail from the raw input. Within healthcare it is deep learning techniques that are behind the recent advances in automated image processing. As yet they seem to be rarely used within text-based healthcare analysis, with only 1/86 papers applying sentiment analysis methods using deep learning and 4/86 using word embeddings [145] although one review commented on how researchers were starting to use these methods on existing classification and negation identifications problems [224]. Only one review focused on deep learning methods, but these were mostly applied on EHR and biomedical literature data with only a few examples of SM data usage [247].

4.5.5 RQ4: What are the knowledge gaps and areas of future research needed?

All the reviews acknowledge that method development is still at an early stage and that much more work is needed before the full potential of SGOPE can be utilised. Particular challenges identified include algorithm design [77,227], method refinement [233,241] integrating diverse data sources [160,245], pre-processing, coreference and temporal relation extraction [66,241], spelling correction, normalising poster language [241], and reducing bias [140,252]. Studies to date have considerable heterogeneity in methods and outcomes, further work is also needed to define the optimal standards for these.

The general lack of availability of health specific trained or labelled data has implications for the accuracy that can be achieved. Sentiment analysis performance on health related text was found to be lower than that of other domains [145] but that may be because most of the commonly used sentiment lexicons have been developed from publicly available film or restaurant reviews, which do not work as well on health topics [145].

Health domain language can be quite specific to the domain, even at a lay person level. A variety of approaches are being explored to deal with the inconsistencies of 'patient language' such as spelling correction, and attempts to map lay language to medical ontologies [143]. There are calls for the development of annotation guidelines [96] and sentiment analysis tools trained on health care specific corpora [35]. One of the problems with current standard sentiment lexicons is that they are too general for health topics. Attempts to map them to the Unified Medical Language System (UMLS) found that less than 1% of its content is covered by common existing lexicons [145]. Identifying and understanding negation within text remains a major challenge for NLP systems [230]

Several reviews suggest greater sharing of datasets [66,143,145,236,252] and the wider development of shared tasks, where different groups can work towards solving a particular task on the same dataset [66,96,236]. Often used in computer science they are seen as the most comprehensive way to evaluate methods and techniques. All groups share the same training dataset, with the test data being released after the evaluation of the algorithm or pipeline process [175].

At this early stage of method development there are a variety of tools and algorithms available to analyse unstructured text, but a lack of studies that compare their efficiency, and there is therefore a lack of consensus as to which are the most useful [35,145,248]. The frequent lack of clear explanation of the methods used in studies [238,249] and the poor reporting of datasets used [238] means that it is hard to assess the accuracy of many results and may lead to selective outcome reporting or publication bias [248].

Further work is needed in terms of evaluating the findings from SGOPE data, both against existing signal detection methods [142], and to psychosocial, behavioural and physical outcomes [257]. Comparisons of SGOPE data to that in clinical text such as EHRs or biomedical literature identified the potential value of SGOPE but highlighted the particular issues of noisy, irrelevant content, language inconsistencies and ambiguity

[66,141,143,230], but made no comment on how the accuracy of SGOPE data analysis compares with these methods [146].

Other areas for future research were identified including the need to adapt the methods to languages other than English [66,141,175,224,237,238,243], cost- effectiveness studies [248], better understanding of how SGOPE can help posters self-manage [54], maximising the representativeness of the data [144,235], facilitating evaluation [74,258], integrating SM text with audio and video sources [167,237,239], and a better understanding of how SGOPE could integrate with existing systems [241]. Three reviews commented on the lack of demographic analysis, despite geotags being easily accessible from Twitter data [69,74,79,145]. Each individual tweet has potentially 38 data features including detailed metadata such as geotags, but these seem to be unexplored at present [69]. Methods that included temporal analysis could help identify event sequences and causal inferences [238]. Only one review focused on health outcomes [146]. A lack of linking SGOPE interactions and analysis with health outcomes has also been identified [257].

There is a lack of both theoretical [232] and methodological data centric frameworks [96] for SGOPE usage, hindered by the discipline boundaries where researchers in one area often do not know of relevant literature in another. This is compounded by differences in language, terminology and methods that exist [175]. Giuntini suggests that a multi-disciplinary approach could help develop better algorithms [242].

Concerns around the ethics of using social media data posted in public spaces are ongoing and several reviews mentioned the need to be aware of ethical issues [67,69,140,175,231,232,238,251]. The absence of any form of discussion around the ethical implications of this form of data use was highlighted in 23/26 surveillance studies [167] and 13/16 studies on suicide ideation [238]. The need for guidelines and harmonisation of regulation around secondary use of SM data and data donation was identified [231,233] together with a call to analyse data through different socio-cultural lenses [251].

As the issues around privacy and consent begin to be resolved, further questions emerge about the how findings should be incorporated into health care practice. Regarding its use as a method of public health surveillance there is a lack of guidance as to how health organisations should accept or react to data from SM discussions [235]. In the area of mental health questions remain as to if and how any posters deemed to be ‘at risk’ should or could be contacted [175].

The need for interdisciplinary collaboration between NLP and health researchers in order to maximise the opportunities available is highlighted [143,175]. A gap between academic NLP research and the commercial NLP systems as they are beginning to be used on electronic health records (EHRs) has been identified, in that academic work tends to be more advanced [66]. One review looking at the development of methods identified a number of approaches that were in development for analysing SM text but concluded that many NLP developments are not getting as far as being used in applications – 'they are often explored, published and then shelved' [143].

This type of data source has traditionally been seen as lacking credibility, although one review commented that recently several of the major science journals have begun to publish articles supporting its use within health research [67]. One identified limitation is the potential for the content to be influenced by media events or coverage [159,235]. To increase its acceptability efforts need to be made to bridge research and practice by demonstrating translational potential [78,140]. Aligning SGOPE data with clinical EHR data could help to both bridge a credibility gap and help both posters and clinicians reach a better understanding of how health issues impact on individual's lives [140].

4.6 Discussion

In total, 58 review papers were included that answered the research question of how and why SGOPE data is being used in health research in terms of the sub questions (Table 4-1). Of these, 13/58 looked primarily at the purposes, 11/ 59 primarily at the methods, while 34/58 addressed both the purpose and the methods. Despite the heterogeneity of studies included, the early stage of methodology development and the many challenges still to be overcome, there was universal agreement between them of the potential of SGOPE data to improve health and deliver patient centred care. Twitter was by far the most widely used data source, and the majority of studies to date have used either qualitative, quantitative or supervised machine learning methods.

4.6.1 RQ1 Which sites and platforms are being used as data sources?

The high prevalence of Twitter as a data source probably reflects the easy accessibility to large volumes of data that has been accessible through their API, rather than its suitability for health research. Volume extraction restrictions now exist and despite the doubling of the character limit in 2017 to 280 characters per text [260], it still does not offer the unlimited length of post that is available on most discussion forums or blogs. Facebook was another common source, but the recent privacy issues have resulted in far fewer messages being publicly accessible in the last few years [35]. Access to some of the potentially more useful online forums and communities is being restricted, due to a combination of privacy concerns and commercial interests, as the realisation grows as to the economic value of health data [261].

Restricting individual studies to a single data source may be simpler for method development, but it does decrease the overall validity of the studies due to the elements of emotional contagion [168,169] that may be present on a single data source, especially if it is a relatively small community. Even a massive source such as Twitter is still quite limited in the demographics of its posters which has implications for the generalisability of findings based on it [251].

4.6.2 RQ2: What purposes is SGOPE data used for?

SGOPE data is recognised as adding a new dimension to healthcare research [158] and as such has a variety of potential uses. The health topics that have been researched to date reflect both the early stages of methodology and the type of posts that are most available. The most active communities are known to be those with long-term conditions [145]. Much of the use to date has been retrospective or evaluative, but as methods improve higher degrees of semantic meaning can be accurately extracted, and its role as a predictive or triage tool may become more widespread. Simple key word searches are easy to implement and can be very effective when searching through large volumes of text for mentions of selected drugs or conditions, but as methods develop and the volume of available data expands the range of use cases will widen. One of the potential problems of studies based solely or mainly on one data source such as Twitter is that the content posted there tends to be heavily biased by media coverage of events, so that potential use cases for research is driven by the availability of the data [159,160].

4.6.3 RQ3: Analysis methods identified by the reviews

The level of detail reported varied widely between reviews. Some of the reviews that looked at both uses and methods looked in far greater detail at the individual methods that were used than did some of the pure method focused reviews.

SGOPE has potential to add to knowledge of many different aspects of healthcare, and to a certain extent the aims of each project, as well as the resources available in terms of skill levels and time will guide the choice of methods [249]. Although most of the machine learning methods used supervised techniques, these all require large quantities of annotated data, with both volume and the annotation quality having direct influence on the resulting accuracy. Annotating a dataset is a time intensive process, and thus often expensive as it requires considerable domain specific knowledge [66,224].

Results suggest that the best results are often achieved by combining methods in a pipeline process, but that there is no one single combination that is seen as being the most effective [175]. Tasks such as entity recognition, especially using dictionaries, are much easier than the more complex problems of correctly identifying relationships between the entities. Reports from some of the shared task challenges held to date show that within clinical NLP, named entity recognition (NER) can achieve an accuracy of over 70% , but relation extraction methods are much less successful with performances below 50% [236]. These figures are likely to be lower in SGOPE data where the variations in language, grammar and sentence construction are much wider.

Zunic [145] suggests that increased use of shared datasets could increase the use of deep learning methods, improving the performance levels, as well enabling comparison between methods. However the use of complex deep learning methods requires a trade-off between the computational cost involved and the performance levels that can be achieved [243].

4.6.4 RQ4: What are the knowledge gaps and areas of future research needed?

The strongest message from these reviews was confirming how much more work there is to do in this emerging area, with many gaps in our knowledge, identified areas for future research, as well as limitations with existing studies being identified. The lack of systematic testing of methods and results impacts on the credibility of the findings. The importance of reproducibility is a current issue in healthcare research, so future work in this area should make it clear what methods are used.

From the literature to date, there is little evidence of this data source being used to assess or evaluate the patient perspective of the effectiveness of a treatment, intervention or service, other than for detecting adverse events or reactions. Given that so much is still unknown about the relationships between patient characteristics, environment and disease, patterns of symptoms, behaviours or effects any knowledge that can be extracted may give new insights.

Although many of the reviews acknowledged that sharing knowledge between online users was one of the big factors in online health information use [175] only one of the reviews focused on how important this was to those with long term conditions [54]. Very few of the studies had explored or compared unsupervised methods to identify the themes being discussed.

Few reviews looked at identifying causation from social media[142]. This may be because dictionary-based systems that can match interventions and symptoms within text can be a simple and very effective way of identifying potential relationships. Determining that an adverse effect exists is however, different to determining what the actual effect is, especially if as a retrospective event. There has been less focus on assessing causality to identify true drug-ADE pairs [66,241]. One suggestion is that the low quality of the information precludes the evaluation of causal links [142,241], although the quality of the posts in terms of completeness varies widely between sites [241]. Identifying temporal data to sequence events could help distinguish true causal links [66], as will the continued working on the more complex tasks of lexical semantics, coreference resolution and discourse analysis [230].

4.7 Strengths and limitations

Using an umbrella scoping review approach summarises the current state of the art of this fast-moving field. This is the first umbrella review of the topic since one from 2018 looking at the potential uses, benefits, harms of social media as a data source for mental health research, concluding that better research designs were needed [262]. The increased volume of published literature, especially since 2017, shows the significance of SGOPE as a data source for health research.

One of the strengths of this method is that although some of the individual studies may have been included in multiple reviews, each review paper has had different research questions, thus generating a range of different perspectives on any such papers. Seven databases were searched, together with grey literature and reference lists. It is however subject to the usual limitations of the keyword-based searches, in that it is possible that some relevant literature may have been missed. Searches were also limited to reviews written in English.

4.8 Conclusion

This chapter summarised the findings from the umbrella scoping review of literature investigating how and for what purposes is SGOPE data currently being utilised with health research. The review included 58 existing review papers. Despite the raft of suggestions for future research and methodological development that is needed, the consensus from these reviews is that SGOPE is a data source capable of offering considerable benefit to healthcare providers, and that NLP will become an important methodological tool within health research.

Table 4-5: Characteristics of included review papers

Ref	Title	Review Aims/Objectives	Population	Data Source	Review type	No Papers
Abbe 2016	Text mining applications in psychiatry: a systematic literature review	Two specific objectives: (1) to collect and analyze applications from the studies reviewed in order to assess the benefits and limitations of using TM; and (2) to identify new opportunities for use of TM in psychiatry.	Mental health	Online posts, qual studies, EHRs, biomed literature	Systematic	38
Abd Rahman 2020	Application of Machine Learning Methods in Mental Health Detection: A Systematic Review	The main purpose of this paper is to explore the adequacy, challenges, and limitations of a mental health problem detection based on OSNs data. The objective of this systematic literature review is to conduct a critical assessment analysis on detection of mental health problems using OSNs. We also investigated the appropriateness of this pre-mental health detection by identifying its data analysis method, comparison, challenges, and limitations.	Mental Health	Any SM: mostly Twitter	Systematic	22
Al-Garadi 2016	Using online social networks to track a pandemic: A systematic review	This study aims to investigate the adequacy and limitations of pandemic surveillances based on OSN data.	Infectious disease	Any SM: mostly Twitter	Systematic	20
Allen 2016	Long-Term Condition Self-Management Support in Online Communities: A Meta-Synthesis of Qualitative Papers	To understand the negotiation of long-term condition illness, work in patient online communities and how such work may assist the self-management of long-term conditions in daily life.	Chronic	Mostly disease specific / Gen health sites / FB	Systematic	21
Barros 2020	The Application of Internet-Based Sources for Public Health Surveillance (Infoveillance): Systematic Review	aimed to assess research findings regarding the application of IBSs for public health surveillance (infodemiology or infoveillance).	Public Health	SM, search queries	Systematic	162
Calvo 2017	Natural language processing in mental health applications using non-clinical texts	To highlight areas of research where NLP has been applied in the mental health literature and to help develop a common language that draws together the fields of mental health, human-computer interaction and NLP.	Mental health	Any SM: mostly Twitter	Scoping	23

Ref	Title	Review Aims/Objectives	Population	Data Source	Review type	No Papers
CastilloSanchez 2020	Suicide Risk Assessment Using Machine Learning and Social Networks: a Scoping Review	Aims to identify the machine learning techniques used to predict suicide risk based on information posted on social networks.	Mental Health	Any SM: mostly Twitter	Scoping	16
Charles-Smith 2015	Using Social Media for Actionable Disease Surveillance and Outbreak Management: A Systematic Literature Review	1. Q1. Can social media be integrated into disease surveillance practice and outbreak management to support and improve public health? 2. Q2. Can social media be used to effectively target populations, specifically vulnerable populations, to test an intervention and interact with a community to improve health outcomes?	Infectious disease	Any SM: mostly Twitter	Systematic	33
Cheerkoot-Jalim 2020	A systematic review of text mining approaches applied to various application areas in the biomedical domain	To identify the different text mining approaches used in different application areas of the biomedical domain, the common tools used, and the challenges of biomedical text mining as compared to generic text mining algorithms.	Any	EHR, Biomed literature, SM	Systematic	34
Convertino 2018	The usefulness of listening social media for pharmacovigilance purposes: a systematic review	To evaluate the usefulness and quality of signals from social media listening.	ADR	Varied	Systematic	38
Demner-Fushman 2016	Aspiring to Unintended Consequences of Natural Language Processing: A Review of Recent Developments in Clinical and Consumer-Generated Text Processing	To review work over the past two years in Natural Language Processing (NLP) applied to clinical and consumer-generated texts	Any	Clinical & UG texts.	General review	NS
Dobrossy 2020	"Clicks, likes,sharesandcomments"asystematic reviewofbreastcancerscreeningdiscourseinsocialmedia	we had two aims: first, to assess the volume, participants, and content of breast screening social media communication and second, to find out whether social media can be used by screening organizers as a channel of patient education.	Breast Cancer	Any SM: mostly Twitter	Systematic	17
Dol 2019	Health Researchers' Use of Social Media: Scoping Review	To explore how social media is used by health researchers professionally, as reported in the literature	Any	Varied	Scoping	414

Ref	Title	Review Aims/Objectives	Population	Data Source	Review type	No Papers
Dreisbach 2019	A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data	To synthesize the literature on the use of natural language processing (NLP) and text mining as they apply to symptom extraction and processing in electronic patient-authored text (ePAT)	Symptoms	Varied	Systematic	21
Drewniak 2020	Risks and Benefits of Web-Based Patient Narratives: SystematicReview	This review aimed to evaluate whether research-generated Web-based patient narratives have quantifiable risks or benefits for (potential) patients, relatives, or health care professionals	Any	Any SM	Systematic	17
Edo-Osagie 2020	A scoping review of the use of Twitter for public health research	aims to review and synthesize the literature on Twitter applications for public health, highlighting current research and products in practice.	Any	Twitter	Scoping	92
Falisi 2017	Social media for breast cancer survivors: a literature review	To provide a systematic synthesis of the current literature in order to inform cancer health communication practice and cancer survivorship research.	Breast cancer	Online support groups	Systematic	98
Filannino 2018	Advancing the State of the Art in Clinical Natural Language Processing through Shared Tasks	To review the latest scientific challenges organized in clinical Natural Language Processing (NLP) by highlighting the tasks, the most effective methodologies used, the data, and the sharing strategies.	Any	Twitter/ ReachOut forum	General review	17
Fung 2016	Ebola virus disease and social media: A systematic review	Ebola virus disease and social media, especially to identify the research questions and the methods used to collect and analyze social media	Infectious disease	Mostly Twitter & YouTube	Systematic	12
Gianfredi 2018	Harnessing Big Data for Communicable Tropical and Sub-Tropical Disorders: Implications From a Systematic Review of the Literature	To systematically assess the feasibility of exploiting novel data streams (NDS) for surveillance purposes and/or their potential for capturing public reaction to epidemic outbreaks.	Infectious disease	Any SM: mostly Twitter	Systematic	47
Giuntini 2020	A review on recognizing depression in social networks: challenges and opportunities	investigates the state-of-the-art of how sentiment and emotion analysis approaches can identify depressive disorders in social networks.	Mental Health	Any SM: mostly Twitter	Systematic	26
Gohil 2018	Sentiment Analysis of Health Care Tweets: Review of the Methods Used	To review the methods used to measure sentiment for Twitter-based health care studies.	Any	Twitter	Systematic	12

Ref	Title	Review Aims/Objectives	Population	Data Source	Review type	No Papers
Golder 2015	Systematic review on the prevalence, frequency and comparative value of adverse events data in social media	To summarize prevalence, frequency and comparative value of information on the adverse events of healthcare interventions from user comments and videos in social media.	ADR	Mostly discussion forums	Systematic	51
Gonzalez-Hernandez 2017	Capturing the Patient's Perspective: a Review of Advances in Natural Language Processing of Health-Related Text	To review the recently published literature discussing the application of NLP techniques for mining health-related information from EHRs and social media posts. To provide a scope of the trends and advances in capturing the patient's perspective on health within the last three years.	Any	SM + EHRs	General review	87
Gupta 2020	Social media based surveillance systems for healthcare using machinelearning: A systematic review	we review the recent work, trends, and machine learning(ML) text classification approaches used by surveillance systems seeking social media data in the healthcare domain. We also highlight the limitations and challenges followed by possible future directions that can be taken further in this domain.	Any	Any SM: mostly Twitter	Systematic	26
Hamad 2016	Toward a Mixed-Methods Research Approach to Content Analysis in The Digital Age: The Combined Content-Analysis Model and its Applications to Health Care Twitter Feeds	To identify studies on health care and social media that used Twitter feeds as a primary data source and CA as an analysis technique.	ADR	Twitter	Narrative review	18
Ho 2019	Data-driven Approach to Detect and Predict Adverse Drug Reactions	Compares omics, social media and EHRs as sources of ADR knowledge	ADR	Any SM	General review	22
Injadat 2016	Data mining techniques in social media: A survey	Techniques, areas, performance, comparison of techniques, strengths and weaknesses of data mining methods	Any	Any SM	Survey	66
Karmegan 2020	A Systematic Review of Techniques Employed forDetermining Mental Health Using Social Media inPsychological Surveillance During Disasters	Our review aims to analyze the possibility, effectiveness, and procedures of using social media data to understand the emotional and psychological impact of an unforeseen disaster on the community.	Mental Health	Any SM: mostly Twitter	Systematic	18

Ref	Title	Review Aims/Objectives	Population	Data Source	Review type	No Papers
Kim 2017	Scaling Up Research on Drug Abuse and Addiction Through Social Media Big Data	To determine how social media big data can be used to understand communication and behavioral patterns of problematic use of prescription drugs.	Substance misuse	Twitter	Critical	8
Lafferty 2015	Perspectives on social media in and as research: A synthetic review	To summarize findings, opinions and discussion about the use of SoMe in research, including examples from psychiatry.	Mental health	Varied	Systematic	56
Lardon 2015	Adverse Drug Reaction Identification and Extraction in Social Media: A Scoping Review	To explore the breadth of evidence about the use of social media as a new source of knowledge for pharmacovigilance.	ADR	Mainly online fourms +Twitter/blogs	Scoping	24
Lau 2019	Artificial Intelligence in Health: New Opportunities, Challenges, and Practical Implications	To summarise the state of the art during the year 2018 in consumer health informatics	Any	Any SM	General review	14
Lopez-Castroman	Mining social networks to improve suicide prevention: A scoping review	Narrative review of possible suicidal behaviours on social networks	Mental health	NS	Scoping	NS
Mavragani 2020	Infodemiology and Infoveillance:Scoping Review	The aim of this paper is to provide a scoping review of the state-of-the-art in infodemiology along with the background and history of the concept, to identify sources and health categories and topics, to elaborate on the validity of the employed methods, and to discuss the gaps identified in current research.	Any	Any SM: mostly Twitter	Scoping	338
Neveol 2017	Making Sense of Big Textual Data for Health Care: Findings from the Section on Clinical Natural Language Processing	To identify the best clinical NLP papers of 2016	Any	SM + EHRs	General review	5

Ref	Title	Review Aims/Objectives	Population	Data Source	Review type	No Papers
Neveol 2018	Expanding the Diversity of Texts and Applications: Findings from the Section on Clinical Natural Language Processing of the International Medical Informatics Association Yearbook	Summarize recent research / best papers for clinical NLP in 2017	Any	Any SM	General review	15
Patel 2015	Social Media Use in Chronic Disease: A Systematic Review and Novel Taxonomy	To evaluate clinical outcomes from applications of contemporary social media in chronic disease; to develop a conceptual taxonomy to categorize, summarize, and then analyze the current evidence base; and to suggest a framework for future studies on this topic	Chronic	Any SM	Systematic	42
Pourebrahim 2020	Adverse Drug Reaction Detection Using Data Mining Techniques: A Review Article	The aim of this study is to study, review and challenge the methods of ADR diagnosis by data mining on social media, especially Twitter.	ADR	Any SM: mostly Twitter	General	0
Qiao 2020	A Systematic Review of Machine Learning Approaches for Mental Disorder Prediction on Social Media	The purpose of this paper is to provide a systematic overview of SM studies in the mental disorder detection field.	Mental Health	Facebook, Twiiter, Reddit, Tumblr, Instagram	General	0
Ru & Yao 2019	A Literature Review of Social Media-Based Data Mining for Health Outcomes Research	To summarize key points of the area in data accessibility, textual data preprocessing methods, analysis methods, opportunities and challenges.	Any	Any SM	General review	19
Santos 2019	Datamining and machine learning techniques applied to public health problems: A bibliometric analysis from 2009 to 2018	To: (i) analyze the number of papers published from 2009 to 2018 (10 years) due to the increasing number of publications and dissemination of ML in public health; (ii) identify the journals with the greatest number of papers; (iii) determine which techniques, programming languages and software tools are most widely used in the field of DM applied to public health; (iv) identify which countries and databases were targeted by these studies; (v) analyze which public health classes were tackled by these papers and (vi)identify which papers were most frequently cited in the literature.	Public health	Any SM	Bibliometric	250

Ref	Title	Review Aims/Objectives	Population	Data Source	Review type	No Papers
Sarker 2019	Mining social media for prescription medication abuse monitoring: a review and proposal for a data-centric framework	To present a methodological review of social media-based PM abuse or misuse monitoring studies, and to propose a potential generalizable, data-centric processing pipeline for the curation of data from this resource.	Substance misuse	Twitter / Facebook / Reddit	General review	39
Sharma 2016	Identifying Complementary and Alternative Medicine Usage Information from Internet Resources. A Systematic Review	Identify and highlight research issues and methods used in studying Complementary and Alternative Medicine (CAM) information needs, access, and exchange over the Internet.	CAM	Any SM	Systematic	120
Sharma 2020	Sentiment analysis of social media posts on pharmacotherapy: A Scoping Review	The aim of this scoping review was to describe the available evidence as it pertains to SA of Social Media specifically about pharmacotherapy. Themes will be generated about the published uses of SA and the real-world implications of the knowledge generated.	Any	Any SM: mostly Twitter	Scoping	10
Sinnenberg 2017	Twitter as a Tool for Health Research: A Systematic Review	To systematically review the use of Twitter in health research, define a taxonomy to describe Twitter use, and characterize the current state of Twitter in health research.	Health research	Twitter	Systematic	137
Skaik 2020	Using Social Media for Mental Health Surveillance: A Review	This systematic review aims to analyze the literature on using social media posts to predict mental disorders using ML and NLP methods that could be useful for mental health surveillance and presents the cutting-edge techniques in predictive analysis of suicide ideation and depression at the population-level. It also points at the gaps that need further research from the perspective of the data, the models, and evaluation procedures.	Mental Health	Any SM	General	110
Staccini 2017	Secondary Use of Recorded or Self-expressed Personal Data: Consumer Health Informatics and Education in the Era of Social Media and Health Apps	To summarize the state of the art during the year 2016 in the areas related to consumer health informatics and education with a special emphasis in secondary use of patient data.	Any	Any SM	Systematic	5
Su 2020	Deep learning in mental health outcome research: a scoping review	The goal of this study is to review existing research on applications of DL algorithms in mental health outcome research.	Mental Health	SM, EHR, etc	Scoping	57

Ref	Title	Review Aims/Objectives	Population	Data Source	Review type	No Papers
Tricco 2018	Utility of social media and crowd-intelligence data for pharmacovigilance: a scoping review	Review the literatuture regarding using SM conversations for ADR detection	ADR	Any SM	Scoping	70
Vilar 2018	Detection of drug-drug interactions through data mining studies using clinical sources, scientific literature and social media	To review datamining as a method of detecting drug-drug interactions	ADR	SM/ EHRs. FAERS, WHO	General review	NS
Wilson 2015	Using blogs as a qualitative health research tool: A scoping review	To identify how blogs are being used in health research to date and whether blogging has potential as a useful qualitative tool for data collection. Our purpose was to summarize the extent, range, and nature of research activity using blogs.	Any	blogs	Scoping	44
Wong 2018	Natural Language Processing and Its Implications for the Future of Medication Safety: A Narrative Review of Recent Advances and Challenges	To review methods of identifting adverse events from free text	ADR	SM + EHRs	General review	12
Wongkoblap 2017	Researching Mental Health Disorders in the Era of Social Media: Systematic Review	To explore the scope and limits of cutting-edge techniques that researchers are using for predictive analytics in mental health and to review associated issues, such as ethical concerns, in this area of research.	Mental health	Various SM	Systematic	48
Yin 2019	A systematic literature review of machine learning in online personal health data	To systematically review the effectiveness of applying machine learning (ML) methodologies to UGC for personal health investigations.	Any	Any SM: mostly Twitter	Systematic	103
Zhang 2018	Using Twitter for Data Collection With Health-Care Consumers: A Scoping Review	To provide an overview of previously published literature describing Twitter as a data collection method with health-care consumers and provide researchers with considerations when potentially using this data collection approach.	Any	Twitter	Scoping	17
Zhang 2020	When Public Health Research Meets Social Media:Knowledge Mapping From 2000 to 2018	Aims to examine research themes, the role of social media, and research methods in social media–based public health research published from 2000 to 2018	Any	Any SM	Review	3419

Ref	Title	Review Aims/Objectives	Population	Data Source	Review type	No Papers
Zunic 2020	Sentiment Analysis in Health and Well-Being: Systematic Review	This study aimed to establish the state of the art in SA related to health and well-being by conducting a systematic review of the recent literature. To capture the perspective of those individuals whose health and well-being are affected, we focused specifically on spontaneously generated content and not necessarily that of health care professionals.	Any	Various SM	Systematic	86

Table 4-6: Aims, outcomes, key findings, methods and research gaps

Ref	Paraphrased Aims	Focus	Outcomes	Key Findings Paraphrased	Methods mentioned	Future Research
Abbe 2016	Benefits & limitations. Current and potential uses in psych.	Mental health	Objectives of studies, and topic modelling methods /tools used for pre-processing and analysis.	Identified four main areas of application: Psycho-pathology, patient perspective, medical records, medical literature. A data source that cannot be ignored. Techniques and topics heterogenous. Basic capabilities at present but will get better and become a core method.	Mostly rule based systems but some classification.	Improved techniques, apply to more languages than English.
Abd Rahman 2020	Adequacy, challenges, and limitations of SGOPE data for detecting MH problems	Mental Health	DataSources, Conditon, location, Feature extraction methods, analysis methods	22 studies: stress 8, depression 7, suicide 3, MH disorders 4. Geographical: Chna 6, US, 4, Japan 1, Greece 1, unspecified 10. Source: twitter 8, Sina Weibo 5, Facebook 2, others 7. The keywords used to select data often not specified. SVM (13/22) most popular classification, LR & RF (5/22), NB 4/22)	Text analysis, multi method inc questionnaires, accessing respondents OSN accounts. Feature extraction TF-IDF, ngrams, BOW,	Multiple sources, other languages, inclusion of audio, video, photos. Better methods
Al-Garadi 2016	Adequacy / limitations of SM for pandemic surveillance	Infectious disease	Data source and volume, analysis method, study aims and outcomes. Features and classifier performance of supervised methods.	Can complement existing systems / still problems with representivity. Need better algorithms and computational linguistic methods.	Mostly supervised, classification. SVM. Most used ngrams as features.	Better algorithms/ computational linguistics
Allen 2016	Better understanding of how patients with chronic disease share knowledge in online spaces. Possibilities for improving self-management.	Chronic	Network themes and mechanisms	Helpful in encouraging patients to self-manage l/term conditions through sharing collective knowledge, gifting relationships, sociability and disinhibition. Need to understand why people do or don't post.	Qualitative: thematic, grounded theory, content & thematic, IPA, ethnography	Find out why people are reluctant to post and illuminate how these communities help people manage their condition in daily life.

Ref	Paraphrased Aims	Focus	Outcomes	Key Findings Paraphrased	Methods mentioned	Future Research
Barros 2020	To assess research findings regarding the application of IBSs for public health surveillance (infodemiology or infoveillance). Sources, purposes, methods	Public Health	Paper type, year, disease, health topic, forecasting, surveillance, disease characterisation, first person health mention, diagnosis prediction,	Infectious disease the biggest area. We also identified limitations in representativeness and biased user age groups, as well as high susceptibility to media events by search queries, social media, and web encyclopaedias	Correlation analysis (59/162) regression models (46/162). Machine learning 27/162, statistical models 20/162. Manual analysis 18/162, topic analysis 12/162. Deep learning 10/162, linguistic analysis 10/162. Rule-based techniques (n=7), epidemiology theory(n=6), surveys (n=3), and ranking techniques (n=1) were used in less than 10 papers.	Updating keywords to reflect changing search behaviours and health trends. Susceptibility of SM content to media events. Creation of standard datasets to improve method development.
Calvo 2017	What NLP methods used on user generated data in mental health?	Mental health	Objectives of studies, data sources, features extracted	Triaging MH issues seems like a great use but need to find how to react to it in practice. Ethics/ privacy issues. Very interdisciplinary.	LIWC most widely used both for feature extraction and Sentiment analysis. Good methods often a combination of methods/ algorithms. Lots of different tools/ techniques available- could not determine whether any one was superior.	Need to do research into using NLP in different languages. also think about how to make contact with people identified as being at risk from mental health that are identified during the process.
CastiiloSanchez 2020	What ML techniques used to predict suicide from SM data?	Mental Health	Methods, Tools, Techniques	Text classification main objective for 75%. 8/16 studies report explicit datamining techniques. 10/16 using SVM. Papers not reporting time spans of data collection, or number of participants.	LIWC, LDA, LSA for feature extraction, Sentiment analysis	Other languages. Use annotated corpus. Develop new tools. Do temporal studies.

Ref	Paraphrased Aims	Focus	Outcomes	Key Findings Paraphrased	Methods mentioned	Future Research
Charles-Smith 2015	Can SM be used for disease surveillance? Or to test interventions to improve health outcomes?	Infectious disease	Correlation between social media data and national health statistics. Prediction times. Topic / theme identification. Influence on health behaviours.	Earlier prediction of outbreaks. Correlation with existing methods. Topic modelling good for broad topics, but not for lower frequency themes. Lots of gaps in knowledge. Need to look for ways to incorporate SM into PH surveillance.	Topic modelling (LDA). Query selection and thematic analysis to detect lower frequency topics.	Work on who uses what types of social media, so as to get representative data. SM platforms/ preferences change.
Cheerkoot-Jalim 2020	Identify the text mining approaches, tools used in biomedical text. Who benefits? Application areas? What are the challenges?	Any	DataSources, Techniques, Tools and Potential Beneficiaries of research	Looked at who could benefit from SGOPE research	MetaMap, UMLS used - mainly on EHRs and biomed literature. NLP methods; NER and relationship extraction.	Big data paradigms,methods that can scale with the volume of text.Methods of standardsing data across sources. Improving accuracy.
Convertino 2018	Summarise strategies, assess quality of information, potential for early detection from SM.	ADR	Sources, study population, drug Proto-ADE pairs, clinical features, extraction method.	Lots of potential to complement existing regulatory agencies. But utility, validity and implementation are all under-studied. Need standardised methods. Fast moving field. No causality assessment so far.	Keywords, dictionary most popular 37/38.	More work to improve methods. Use in conjunction with other signal detection methods.
Demner-Fushman 2016	Improvements in NLP on patient language, and new opportunities.	Any	SM as a source for quality assessment. Methods	Much more to be done both in clinical and SM NLP. Research moving from capturing trends to addressing individual health-related posts, thus showing potential to become a tool for precision medicine and a valuable addition to the standard healthcare quality evaluation tools.	Sentiment analysis. Rule based RegEx or supervised event extraction most used. More work needed on semantic processing. Using sentences better than words,	Need more publicly available clinical datasets. Work on semantics. Work on porting pipelines across domains. Collaboration between NLP research and EHR suppliers.

Ref	Paraphrased Aims	Focus	Outcomes	Key Findings Paraphrased	Methods mentioned	Future Research
Dobrossy 2020	Assess volume, participants, and content of SM data about breast screening. Potential for patient education.	Breast Cancer	Platforms, volume of discourse, participant roles, discourse content, themes.	Looked at age, role of user types, and the content of the posts. Good source to understand beliefs, attitudes, and literacy of the target population.	NS	NS
Dol 2019	How health researchers are using SM data.	Any	Journals, study country, first author discipline, health topic covered, platforms, study purpose.	81/414 analysing content. Biggest use was recruitment. Generally seen as positive, but concerns re ethics.	NS	Need methods to optimise usage and demonstrate potential.
Dreisbach 2019	Using NLP methods to extract symptoms from SM text	Symptoms	Study purpose, data source, symptom categorisation, evaluation, and performance metrics	Pain and fatigue most evaluated symptoms. Variety of sources. NLP primary methodoly for 15/21 papers. Current focus on extraction of terms. Need to share lexicons to move forward.	21 papers: 14 NLP, 6 text mining, 1 NLP+TM. No breakdown of type of methods.	Future research should consider the needs of patients expressed through ePAT and its relevance to symptom science. Understanding the role that ePAT plays in health communication and real-time assessment of symptoms, is critical to a patient-centred health system.
Drewniak 2020	Does SGOPE research have quantifiable risks or benefits for patients, relatives, or HCPs?	Any	Purposes of the narrative: inform, engage, model behaviour, persuade, comfort	Generally positive benefits although potential risks from misinformation	NS	Future research is needed to define the optimal standards for quantitative approaches to narrative-based interventions.
Edo-Osagie 2020	Current uses of Twitter data in public health	Any	Conditions, data sources, analysis methods, geographical and time trends	Twitter a good data source for 6 aspects of public health: surveillance, event detection, pharmocovigilance, forecasting, disease tracking and geographical identification.	Numerous	Unsupervised methods. Do research into less studied areas

Ref	Paraphrased Aims	Focus	Outcomes	Key Findings Paraphrased	Methods mentioned	Future Research
Falisi 2017	What role does SM play in the health of breast cancer survivors?	Breast cancer	Platforms, ethnicity of study population, analysis method, which aspects analysed, connection between SM content and health outcome.	Focus on psychosocial well-being. Mostly online support forums/ message boards. Few non-Caucasian. Content analyses of social media interactions prevalent, but few articles linked content to health outcomes	40/98 did content analysis. Some manual / some M/L. Pre 2011 = LIWC, post 2011= LDA etc. 37 quant. 3 qual	Should consider connecting SM content to psychosocial, behavioral and physical health outcomes. None of the content analysis articles attempted to do this.
Filannino 2018	What tasks and methods included in the shared tasks?	Any	Task description, data type, data source, dataset size, best performance, measure.	NER & classsification the most used tasks. Clear trend to data-driven solutions. Need more and varied datasets to explore.	NER and classification most common tasks.	Bigger and more varied datasets to share
Fung 2016	What research questions and methods used on ebola related social media?	Infectious disease	Study design, qual or quant, study aim, data collection method, time frame, keywords used, analysis method, main findings and limitations.	12 papers: 8 from Twitter/ Weibo, 1 from Facebook, 3 from YouTube, and 1 from Instagram and Flickr. All studies were cross-sectional. 11/12 articles studied one or more of themes / topics of SM content, post meta-data and characteristics of the SM account. Twitter content analysis methods included text mining (n = 3) and manual coding (n = 1). Two studies involved mathematical modeling. YouTube /Instagram/Flickr studies used manual coding of videos and images. Published Ebola virus disease-related social media research focused on Twitter and YouTube. The utility of social media research to public health practitioners is warranted. No evaluation of the studies utility performed.	Mix of manual coding and frequency analysis using LIWC.	Need a new checklist to appraise quality of SM papers. Future research in the direction of analyzing multiple cross-sectional social media datasets or conducting prospective cohort studies of social media users will provide useful data for analysis of temporal change of social media contents or social media users' behaviors. Need to bridge research and practice.

Ref	Paraphrased Aims	Focus	Outcomes	Key Findings Paraphrased	Methods mentioned	Future Research
Gianfredi 2018	Can SM be used for disease surveillance / predictions? Can they capture public reactions to epidemic outbreaks?	Infectious disease	Data source, disease, study period, geographical location, study purpose, type of analysis and main findings	Out of the 47 articles included, only 7 were focusing on neglected tropical diseases, while all the other covered communicable tropical/sub-tropical diseases, and the main determinant of this unbalanced coverage seems to be the media impact and resonance.	Qualitative, narrative analysis, content analysis, mathematical modelling, correlational analysis, geospatial etc etc	Lots of gaps, possibly due to the media impact of the specific disease. Need further research into ways of integrating diverse data sources.
Giuntini 2020	Sentiment and emotion analysis for identifying depressive disorders. What types of SM data? Which networks? Which methods?	Mental Health	Platform, type of SM, emotion or feeling detection, other disorders inferred, methodology	Most used media is text, then emoticons. Twitter most employed platform. Supervised methods with off the shelf classifiers combined with lexicons such as LIWC.	Supervised (NB, DT, SVM etc) plus LIWC, NRC Word Emoticon, word-Net Affect lexicons	More multidisciplinary studies.
Gohil 2018	What sentiment analysis tools for Twitter / healthcare. Any health specific training, validation, or justification	Any	Health area, sentiment towards, type of method, tool used, manual annotation sample size, sample size	Multiple methods, mix of open source, commercial and bespoke tools. Very few tested for accuracy.	Sentiment analysis. Mix of tools.	This study suggests that there is a need for an accurate and tested tool for sentiment analysis of tweets trained using a health care setting–specific corpus of manually annotated tweets first.
Golder 2015	Prevalence, frequency and value of ADR comments from SM	ADR	Data source type, ADR type, search strategy used, post selection, study aim, ADR prevalence, comparison method	51 studies, discussion forums most used source type. ADR prevalence varied from 0.2% to 8%. General agreement that a higher frequency of adverse events was found in social media and that this was particularly true for ‘symptom’ related and ‘mild’ adverse events.	8/12 used Consumer Health Vocab dictionary. Few evaluation methods	A cost-effectiveness analysis of all pharmacovigilance systems, including social media is urgently required.

Ref	Paraphrased Aims	Focus	Outcomes	Key Findings Paraphrased	Methods mentioned	Future Research
Gonzalez-Hernandez 2017	Show how NLP is developing in regard to capturing the patient perspective from unstructured text.	Any	Types of SM sites, analysis type, types of tasks.	Move from rule based to learning based systems. Work needed on noise reduction and normalisation/mapping. Shortage of annotated shared datasets. Shared tasks useful development tool.	Move from rule based to learning methods. Over 50% papers used lexical content analysis. In SM NLP: regex, LDA topic modelling. Supervised classification. Sentiment analysis	Normalisation of data, co-reference and temporal relation extraction. Need to create and release annotated datasets and targeted unlabeled data sets in distinct languages.
Gupta 2020	What methods, sources, are used for SM based health surveillance. Potential applications, and challenges.	Any	ML Methods, Data Sources, Diseases, Limitations of SM systems	Twitter most used source (64%). SVM most used method (33%) - better at binary classification.	SVM, Decision trees, random forest, NB, Logistic Regression	Noise reduction, Combining SM with other data, theme detection, develop better predictive models for epidemic prediction. Only 3 studies included ethical debate.
Hamad 2016	How is content analysis used in health-related SM studies?	ADR	Keywords and hashtags, sampling and data collection, analysis methods, validation, and presentation of results	Methods used were not purely quantitative or qualitative, and the mixed-methods design was not explicitly chosen for data collection and analysis. Proposes CCA analysis as straightforward method for Twitter analysis	Content analysis (quantitative and qualitative). Infoveillance. Combined content analysis (mix of mixed methods and content analysis)	NS
Ho 2019	Compares omics, social media and EHRs as sources of ADR knowledge	ADR	Study aim, Data & Tool, Method	Data driven approach essential to detect /predict ADRs. Omics data, EHRs and SM all new opportunities.	Datamining. NLP, NER, ontology building. Classification to exclude noise. Aims to reduce false positive rate. Yang = mix of topic + classification. Classification to link effect to drug. UMLS & MetaMap	NS

Ref	Paraphrased Aims	Focus	Outcomes	Key Findings Paraphrased	Methods mentioned	Future Research
Injadat 2016	Techniques, areas, performance, comparison of techniques, strengths and weaknesses of data mining methods.	Any	Domains, Techniques, Research objectives, Strengths and weaknesses of techniques.	19 data mining techniques used to address 9 different research objectives in 6 different industrial and services domains. Most used methods: SVM, NB & DT. Most used in business and social network analysis. Medical/health use only 8%	Datamining. SVM, BN, DT	Research into how techniques are implemented. Need more statistical tests of results. But - many of the tests applied required a normal distribution which was not the case. Health researchers not good about writing about the methods used. Could learn a lot from CRM and HRM domains.
Karmegan 2020	Aims to analyze the possibility, effectiveness, and procedures of using SM data to understand the emotional and psychological impact of unforeseen disaster on the community.	Mental Health	Platform, methods	Twitter most used source. Sentiment analysis used for psychological surveillance. Could not conclude that any one method was superior.	Feature extraction using classification algorithms. Sentiment analysis	Combine text and image processing. Incorporate social network analysis with post content.
Kim 2017	How SM data can be used to understand communication and behavioral patterns of nonmedical or problematic use of prescription drugs	Substance misuse	User characteristics, communication characteristics, outcomes, methodological domain, ethical domain	See lots of potential, but more work needed.	Mixture: manual, qualitative, supervised / non supervised ML to identify themes, patterns, sentiment.	Lots more - sees their review as a base to build on. Identified a lack of theoretical framework for substance misuse monitoring. Consequences of SM engagement understudied.

Ref	Paraphrased Aims	Focus	Outcomes	Key Findings Paraphrased	Methods mentioned	Future Research
Lafferty 2015	How is SM being used in psychiatry? Tools, benefits, and challenges.	Mental health	SM as data, methodological considerations, ethical considerations, SM for recruitment	Observational, real time patient experiences. Can help with development of practice, policy and provision. Opportunities for co-creation of research, patient centric care.	Grounded theory, Social network analysis	Ethical issues. Analyse SM data through different socio-cultural lenses to build theoretical frameworks.
Lardon 2015	Can SM be a new source of knowledge for pharmacovigilance?	ADR	Language, data source, data volume, methods, lexicon,	Identification theme all 11 papers used manual methods. Identified heterogeneity of methods, but also gaps. Included studies failed to assess the completeness, quality, or reliability of the data.	RQ1: All manual /mixed, RQ2: Web scraping, pre-processing, various rule-based methods.	Additional studies are required to precisely determine the role of social media in the pharmacovigilance system. Need methods to assess data quality.
Lau 2019	2018 SOTA of opportunities, challenges, and implication of AI in health informatics	Any	NS	Few 2018 papers reported Artificial Intelligence (AI) research for patients and consumers. No studies that elicited patient and consumer input on AI. Most common use is secondary analysis of social media data (e.g., online discussion forums). The 3 best papers shared a common methodology of using data-driven algorithms (such as text mining, topic modelling, Latent Dirichlet allocation modelling), combined with insight-led approaches (e.g., visualisation, qualitative analysis and manual review), to uncover patient and consumer experiences of health and illness in online communities. There is a lack of direction and evidence on how AI could actually benefit patients and consumers.	Best papers shared a common methodology of using data-driven algorithms (such as text mining, topic modelling, Latent Dirichlet allocation modelling), combined with insight-led approaches (e.g., visualisation, qualitative analysis, and manual review), to uncover patient and consumer experiences of health and illness in online communities	See what patients want from AI in health. More pt involvement to ensure that asking the right questions.

Ref	Paraphrased Aims	Focus	Outcomes	Key Findings Paraphrased	Methods mentioned	Future Research
Lopez-Castroman 2019	Detecting suicide ideation from SM	Mental health	NS	Early days, but SM has important role in suicide prevention. Lots more work needed.	Various: Sentiment analysis, topic modelling, data mining	Add demographic data to text to improve results.
Mavragani 2020	Current state of SM based infodemiology. Validity of methods and research gaps.	Any	Timeline & journals. Data sources, Health topics, Advantages & Disadvantages of SM data	JMIR most used journal. Increasing interest since 2018. Twitter most used platform. Most researched subjects were conditions/diseases, epidemics, healthcare, drugs, smoking/alcohol.	NS	Combine SM data with traditional sources for more complete assessment.
Neveol 2017	Best clinical NLP papers of 2016	Any	Applications of NLP, Directions of progress	Developing applications rather than methods. Starting on the more complex tasks e.g., semantics, coreference resolution, and discourse analysis.	Classification of useful sentences, Information extraction, abbreviation disambiguation, coreference resolution, grounding of gradable adjectives	NS
Neveol 2018	Summarize recent research / best papers for clinical NLP in 2017	Any	NLP of SM data, NLP of HCP text, methods	2017 trends - revisiting old problems such as SM classification and negation with deep learning & neural nets. Production of annotated corpora. Continuing applications rather than methods. Beginning of deep learning. Start of language variants.	Negation detection, corpus annotation, deep learning.	Work in other languages. Increase generalisability.

Ref	Paraphrased Aims	Focus	Outcomes	Key Findings Paraphrased	Methods mentioned	Future Research
Patel 2015	Categorise & summarise existing papers about chronic disease outcomes from SM. Suggest framework for future research.	Chronic	Platform, Taxonomy category, disease, study aim, study design,sample size & description, Method summary, SM effect	85% either Facebook or blogs. 40% for support (social, emotional, or experiential).	Quantitative, Thematic qualitative, Content analysis.	Understand how disease, patient factors and tech can interact to improve outcomes. Reduce potential for bias. Target studies to specific diseases might be the best way to improve clinical care.
Pourebrahim 2020	Datamining methods for ADR detection from SM	ADR	Analysis and evaluation metrics	SM good for early identification of ADRs. Three main stages; Pre-processing, feature extraction and classification	Supervised, regression, unsupervised	NS
Qiao 2020	Overview of SM studies relating to mental disorder detection.	Mental Health	Platforms, collection methods, feature extraction, algorithms, evaluation metrics	Facebook, Twitter, Reddit, Tumblr, Instagram. Most used supervised methods, especially SVM	SVM, Decision trees, random forest, NB, Logistic Regression	Develop systems with lower computational cost to increase speed. Multi-language systems.
Ru & Yao 2019	SGOPE data - methods/analysis opportunities and challenges	Any	Data type, volume, preprocessing method, analysis method, health outcomes	Variety of methods. Outcomes included side effects / effectiveness / adherence / hrqol	NER, mapping, identify concepts, text mining (Ngram, LDA, topic modelling), content analysis, hypothesis testing, supervised, unsupervised	Suggested further research on treatment effectiveness, adverse drug events, perceived value of treatment, and health-related quality of life. The challenge lies in the further improvement and customization of text mining methods. Only 6 discussed ethics.

Ref	Paraphrased Aims	Focus	Outcomes	Key Findings Paraphrased	Methods mentioned	Future Research
Santos 2019	Numbers of papers / journals, countries / databases, methods/tools, which public health issues looked at	Public health	Year, Journal, Study purpose, health area, techniques, software/ programming language, study country	Results showed a slight increase in the number of papers published in 2014 and a significative increase since 2017, focusing mostly on infectious, parasitic, and communicable diseases, chronic diseases and risk factors for chronic diseases. JMIR and PLoS ONE published the highest number of papers. Support Vector Machines (SVM) were the most common technique, while R and WEKA were the most common programming language and software application, respectively. The U.S. was the most common country where the studies were conducted. In addition, Twitter was the most frequently used source of data by researchers.	SVM, Decision trees, random forest, NB, most used techniques. R, WEKA and Python most used languages/ apps.	In depth analysis of variations in techniques (deep learning / ensemble etc)
Sarker 2019	Look at existing methods of SM based medication abuse or misuse, propose new data centric pipeline.	Substance misuse	Data source, dataset size, medication studied, study objectives, methods, and findings.	39 studies, 80% published since 2015. Twitter most used source. Earlier studies manual qualitative, but growing trend towards NLP methods.	Supervised, unsupervised	Develop shared annotation guidelines and annotated datasets. Will help the direct project and enable comparison across methods. Show agreement for manual annotation. Reduce noise in data.

Ref	Paraphrased Aims	Focus	Outcomes	Key Findings Paraphrased	Methods mentioned	Future Research
Sharma 2016	Identify and highlight research issues and methods used in studying Complementary and Alternative Medicine (CAM) information needs, access, and exchange over the Internet.	CAM	NS	Significant interest in developing methodologies for identifying CAM treatments, including the analysis of search query data and social media platform discussions	Qualitative, thematic, content analysis, keyword searches, regex, Consumer health vocabulary	Little work done on using SGOPE to understand CAM user's perspectives / prevalence of CAM use. Lots more work required.
Sharma 2020	Can sentiment analysis be conducted on social media platforms to understand public sentiment held towards pharmacotherapy?	Any	Author, Year, Journal, data source, conditions, pharmacotherapy, SA method used, potential clinical use.	Lack of consistent approach. Opinion on medication (7/10) and ADRs (3/10) Lexicon based more used than ML for sentiment. (Lexicon 6, ML 1). Combining SA with other ADR methods improved results. Lots of untapped potential.	Lexicon, ML. Combining	No gold standard methods yet. Early stage of development. Accuracy rarely assessed.
Sinnenberg 2017	How and why health researchers using Twitter?	Health research	Ways Twitter data used by researchers, ways that Twitter platform used in health research, Publication date, research topic, ethics and funding	The primary approaches for using Twitter in health research that constitute a new taxonomy were content analysis (56%; n =77), surveillance (26%; n =36), engagement (14%; n=19), recruitment (7%; n=9), intervention (7%; n=9), and network analysis (4%; n= 5).	Content analysis, network analysis	Future work should develop standardized reporting guidelines for health researchers who use Twitter and policies that address privacy and ethical concerns in social media research. New opportunities to characterise users from metadata such as demographics.

Ref	Paraphrased Aims	Focus	Outcomes	Key Findings Paraphrased	Methods mentioned	Future Research
Skaik 2020	Recent trends and tools for using social media posts to predict mental disorders using ML and NLP methods. Identifying research gaps.	Mental Health	Collection methods, applications, best practices and gaps	25 papers papers looking at population level mental health classification techniques. 15/25 depressive disorders, 10/25 suicide-ideation. Twitter most used data source, SVM most used model. Heterogeneity of methods and feature selection.	Models: SVM, Ensemble, LR, RF, DT, LSTM. Features: WEKA, LDA, TF-IDF, Sentiment, Lexical, Syntactic, Demographics, Word embedding, Topic modelling	Improve identification of risk factors.
Staccini 2017	Uses and challenges for secondary use of health data	Any	Data donation, uses of SGOPE data	Secondary use of patient data (apart from personal health care record data) can be expressed according to many ways. Requirements to allow this secondary use to have to be harmonized between countries, and social media platforms can be efficiently used to explore and create knowledge on patient experience with health problems or activities. Machine learning algorithms can explore those massive amounts of data to support health care professionals, and institutions provide more accurate knowledge about use and usage, behaviour, sentiment, or satisfaction about health care delivery.	NS	Very early days, lots to work on. Socio-ethical concerns, increased adoption in health care. Need to check AI /SM is asking the right questions. Need a formal framework for consent and secondary use of data. Far from massive adoption in health practice.
Su 2020	Deep learning in Mental Health	Mental Health	Methods, Tools, Techniques	A growing number of studies using DL models for studying mental health outcomes. Particularly, multiple studies have developed disease risk prediction models using both clinical and non-clinical data, and have achieved promising initial results.Lots of potential but lots of challenges	CNN, RNN, Autoencoders	Reduce bias, improve methods

Ref	Paraphrased Aims	Focus	Outcomes	Key Findings Paraphrased	Methods mentioned	Future Research
Tricco 2018	Using SGOPE for ADR detection. Types / characteristics of platforms? How valid or reliable are the conversations?	ADR	Data sources, document characteristics, health conditions, methods, types of listening system, outcome results	46/70 documents (66%) described an automated or semi-automated information extraction system to detect health product AEs from social media conversations (in the developmental phase). Seven pre-existing information extraction systems to mine social media data were identified in eight documents. 19/70 documents compared SM reported AEs with validated data: consistent AE discovery in 17/19. No evaluation of methods or reliability.	Supervised 15/70, Rule based 6/70, unsupervised 4/70, deep learning 1/70, other ML 5/70, Manual or NA 32/70. Dictionary/ lexicon based most used.	Further research is required to strengthen and standardize the approaches as well as to ensure that the findings are valid, for the purpose of pharmacovigilance. Studies required to look at uses / utility over a longer time period. Need standardised methods. Fast moving field.
Vilar 2018	To review datamining as a method of detecting drug-drug interactions from pharmacovigilance sources, scientific literature. Challenges and limitations compared.	ADR	Data source, methods	SGOPE offers new possibilities for identifying DDIs. Current emphasis has been on ADRs not DDIs.	Dictionary matching, association mining, supervised LR.	More studies are necessary to really prove and understand the potential of social media resources and their role in pharmacovigilance.
Wilson 2015	Understanding how blogs could be used for qualitative health research	Any	Geographical location, study aims, now data used in health research.	Used for data collection and recruitment. Good for accessing out of reach populations. Potential for significant improvement of health equity. Sees blogs as 'central part of global transformation'. Need to develop knowledge and skills to take advantage of this new resource.	Purely qualitative	Look for innovative methods to develop qualitative research.

Ref	Paraphrased Aims	Focus	Outcomes	Key Findings Paraphrased	Methods mentioned	Future Research
Wong 2018	To review methods of identifying adverse events from free text	ADR	Definition of NLP tasks, evaluation metrics, challenges in applying NLP to medication safety, data source, methods	Time saving/ real time. Limited by lack of data sharing inhibiting large-scale monitoring across populations. SM good for groups such as children, pregnant women, often not included in trials. Data is Pt reported outcomes, values / preferences - more patient focused.	Supervised, CRF classifier, unsupervised k-means clustering. Linguistic based, standardising text with UMLS. Statistical based.	Integrate data sources from different domains to improve ADR detection. Ethical issues. Increased volume of open-source data.
Wongkoblap 2017	Scope & limitations of new predictive method using SM. Ethical concerns.	Mental health	Key characteristics, data collection techniques, data preprocessing, feature extraction, feature selection, model construction, and model verification.	Methods work across languages. Despite an increasing number of studies investigating mental health issues using social network data, some common problems persist. Assembling large, high-quality datasets of social media users with mental disorder is problematic, not only due to biases associated with the collection methods, but also with regard to managing consent and selecting appropriate analytics techniques.	Most common method was text analysis with LIWC. Sentiment analysis. . Supervised / predictive models. Only 1/58 used deep learning,	Move towards open science standards - share datasets / workflow /code. Ethical aspects of using SM data not clearly defined. Lack of models for detecting stress or anxiety disorders. Combining SM content with confirmed patients rather than self-reported ones. Network analysis to investigate prevalence.

Ref	Paraphrased Aims	Focus	Outcomes	Key Findings Paraphrased	Methods mentioned	Future Research
Yin 2019	To systematically review the effectiveness of applying machine learning (ML) methodologies to UGC for personal health investigations.	Any	Methods, Objectives, DataSource, Health issue, Language, Dataset size	103 eligible studies, summarized with respect to 5 research categories, 3 data collection strategies, 3 gold standard dataset creation methods, and 4 types of features applied in ML models. Popular off-the-shelf ML models were logistic regression (n=22), support vector machines (n=18), naive Bayes (n=17), ensemble learning n=(12), and deep learning (n=11). The most investigated problems were mental health (n=39) and cancer (n=15). Common topics were treatment experience, sentiments and emotions, coping strategies, and social support. Clinical credibility an issue. Application in practice - who should monitor UGC. Conflicting advice from peers / HCPs an potentially interesting avenue. SGOPE can learn health information not in EHRs. Processing and ethical challenges unresolved.	Logistic regression (22), SVM (18), Naïve Bayes (17), ensemble learning (12), deep learning (11)	Ethical aspects of analyzing personally contributed data, bias induced when building study cohorts and dealing with natural language, interpretation of modeling results, and reliability of the findings.

Ref	Paraphrased Aims	Focus	Outcomes	Key Findings Paraphrased	Methods mentioned	Future Research
Zhang 2018	Consideration of Twitter as a data source for health researchers.	Any	Research design, collection techniques, analytic methods, tools, author's opinion on Twitter as a health research method.	17 papers: Quantitative (n =2), qualitative (n =7), and mixed methods (n = 8). Health topics and research questions included pain, migraines, and cancer, social discourse of conditions like perceptions of portrayal of seizures, and cyberspace compared to real-world phenomena. Twitter currently used to search and mine research data. Utilizing Twitter as a recruitment and data collection tool in health research remains largely unexplored. Data collection predominantly passive and covert data collection. Challenges include verification, ethics - overt or covert collection.	Qualitative, quantitative, mixed methods	Creates new questions about data collection, verification, ethics for researchers.
Zhang 2020	Role of SM, themes and methods used in SM based public health research.	Any	Publication trends, themes, role of SM, research methods	Growing number of publications and journals including studies.	Still mostly qual or quant, with little use of computational methods.	Need to develop the methodological potential.

Ref	Paraphrased Aims	Focus	Outcomes	Key Findings Paraphrased	Methods mentioned	Future Research
Zunic 2020	Data sources, roles, motivations and demographics of posters. Topic areas. Practical applications, methods, and current performance levels of sentiment analysis.	Any	Data sources, role of post author, demographic features recorded, health area, ML algorithms used for SA, classification performance, lexical resources	86 studies. Majority of data from social networking/ Web-based retailing platforms. Primary purpose of online conversations is information exchange/social support. Communities tend to form around health conditions with high severity / chronicity rates. Topics include medications, vaccination, surgery, orthodontic services, individual physicians, and health care services in general. 5 poster roles identified: sufferer, addict, patient, carer, and suicide victim. Only 4 reported demographic characteristics. Many methods used for SA. Mainly supervised. Only 1 study used deep learning. Performance less than achieved by general sentiment analysis methods. F-score, below 60% on average. Few domain-specific corpora and lexica are shared publicly for research purposes. Unclear if performance issues are because of the intrinsic differences between the domains and their respective sublanguages, the size of training datasets, the lack of domain-specific sentiment lexica, or the choice of algorithms.	Sentiment analysis. Mix of tools. A wide range of methods were used to perform SA. Most common choices included support vector machines, naïve Bayesian learning, decision trees, logistic regression, and adaptive boosting. Only 1 study used deep learning.	Improved methods. Performance less than achieved by general sentiment analysis methods. Lack of domain specific datasets / lexicons. Need to create and share large, anonymised domain specific datasets. More inclusion of demographic data.

Chapter 5 Phase 3 - Methodology and methods for the main study (P3)

5.1 Outline

In this chapter I build on what was learnt from both the analysis of the exploratory dataset and the findings from the umbrella review to develop, describe and justify the specific methodology and the methods for the main study.

5.2 Aims and objectives

In line with the overall project aim, the aim of this phase is to determine the extent to which the depth of findings of the exploratory study can be achieved at scale using computational text analysis methods on the much larger dataset. The objectives for this include:

- Identifying the themes contained in the posts
- Evaluate effectiveness using sentiment analysis
- Use linguistic methods to explore inferred causal links
- Comparing methods of theme and effectiveness identification

5.3 What have I learnt so far?

The exploratory study (P1) identified eight main themes within the posts including the reason for taking Modafinil (condition or symptom), the impact of the symptoms (physical, emotional, or social), the method of acquisition, reported dosages, side effects, details and comparisons with other interventions tried, the perceived effectiveness and the outcome of taking it on quality of life. These eight themes represent those that are central to many research questions about the effectiveness of an intervention or service. In that study, the effectiveness of Modafinil was evaluated as being 68% positive, 18% negative. Entities in the form of conditions, symptoms, side effects, and acquisition method were identified. The division of the corpus into pre and post modafinil, ngram extraction and the reported rapid onset of any effect allowed the identification of text indicating perceived causality. It also demonstrated that basic NLP methods could highlight facets of the data that could be mapped to the qualitative methods.

The umbrella scoping review (P2) highlighted several factors that have influenced or confirmed the subsequent design of this project. Virtually all the papers commented on how the methodology for this type of data is still at a very early stage of development, with much more work that can and needs to be done. Previous work has tended to be either smaller qualitative studies concerned with understanding how and why SGOPE data could help improve healthcare, or much larger ones that were focused on extracting features or entities from the data. Disciplinary differences were identified, with non-health researchers tending to focus on technical aspects applying various techniques to make small increases in the accuracy of their model, while health researchers focused more on the clinical aspects and implications of its use for both patients and service providers. Although the majority of the studies used supervised methods, very few studies were reported as using deep learning methods. The lack of explanation as to the methods used by many of the individual studies impaired the validity of some of the findings. There was little evidence of any work that either compares the results of qualitative and NLP methods on the same data set or brings the ideas together. The other important finding was the lack of any discussion around the ethical implications of using SGOPE data [67,167,175], which I have addressed in section 2.11 of Chapter 2.

5.4 Overall approach - Supervised, unsupervised, deep learning or ....?

One of the main decisions to be made in any NLP project is to whether to use supervised or unsupervised methods. Despite the review finding that most previous SGOPE studies used supervised methods, I have chosen to take an unsupervised approach, combining NLP with linguistic analysis. The rationale for this includes both practical and philosophical reasoning.

Supervised methods are based on a set of statistical techniques for identifying and classifying parts of speech, entities, sentiment, and other features of text that are developed on a manually annotated 'training' set of data. The general idea is that once the model has been trained by the exposure to the annotated text, it can then be used to predict the same type of labels on previously unseen text [263]. Manually annotating the quantities of data required to give good results is, as has already been shown in the review, both time consuming and resource expensive [264,265]. Supervised methods require large quantities of labelled data to be effective. Current practice suggests that to achieve reasonable results, 80% of any dataset should be labelled for the training stage, and that the model is then tested on the remaining unseen 20% [266].

The use of testing and training datasets does allow for some accuracy evaluation of the models, assessing how well the model can predict on unseen data. However, the annotation process itself is inherently subjective. Any form of labelling introduces the potential for bias, whether conscious or unconscious, to be introduced into the data. A recent extreme example of this was that of the Google online photo service classifying black people as gorillas [267]. It follows that predictive models will exacerbate any errors [65] or bias [268,269], which has to lead to questions as to how much any accuracy figures can be relied on. Ideally any training dataset will also be balanced, in that it contains an equal number of occurrences of the labels of interest so that it can recognise them in the unseen data. Poor accuracy results are often the result of the labelled training data not containing enough or any examples that may be of interest.

Another relevant common theme from the review was that at present there is a lack of trained health specific annotated data suitable for use in supervised models [270]. This issue is compounded by the fact that even if the ethical and practical issues in sharing existing labelled health datasets with other researchers were overcome, they are often labelled for such specific research questions that they are not generalisable to other research questions [271].

An additional disadvantage to using supervised methods for this project is that they follow a top-down deductive methodology where the training data is labelled with what the researchers or annotators think is important at the time and in that data. Following this approach unusual, unexpected, or infrequent features being unrecognised or disregarded as outliers. While there are undoubtably times where a research question requires identification of specific pre-known features or to guide respondents to a predefined path, these methods can miss the opportunity to discover novel patterns or features [272] and are therefore not suitable for an exploratory project such as this.

NLP has advanced significantly in the last few years with the introduction of large transformer models such as GPT3 and the complex neural nets of deep learning. Based on the neural networks of the human brain, deep learning uses relevant training data to learn basic patterns from which the system then independently infers more complex patterns [273]. From a technical perspective, deep learning techniques have led to significant performance advances in perceptual tasks such as image recognition, speech recognition or question answering systems. However, despite the accuracy improvements, they remain less capable of fully understanding the meaning in the text [274], or of identifying causal relationships [275]. Large language models such as BERT [276], introduced from 2019, reduce the amount of labelled training data that is needed for supervised tasks, but at a greatly increased cost in terms of the computational power needed to run them (Table 5-1), and for the very latest models such as GPT3, a big environmental cost. Estimates suggest that training a single BERT base model on GPUs consumes as much energy as a trans-American flight [277].

Table 5-1: Overview of recent language models adapted from [269]

Year	Model	# of Parameters	Dataset Size
2019	BERT	3.4E+08	16GB
2020	GPT-3	1.75E+11	570GB
2021	Switch-C	1.57E+12	745GB

In addition to the increased computational complexity requirements, there has also been criticism of the perceived 'black box' of deep learning techniques, where it is impossible for anyone other than the developers to understand how the algorithm works, the assumptions it makes and the decisions that flow from the assumptions [141,200]. As shown in the review, deep learning methods are not yet widespread within health research. One possible reason for this could be that no matter how accurately the algorithm performs, in order for machine learning or AI systems to become trusted and accepted in healthcare, clinicians, researchers and patients need to feel that the conclusions the systems make and the way that they reach them are interpretable and explainable [278].

Many of the existing use cases for machine learning within healthcare are predictive. Examples include identifying patients at risk of developing a condition or making an early diagnosis from imaging. For those type of research questions, supervised or deep learning systems that continually learn from annotated data continue to improve and can make a significant contribution to health care generally. However, the aim of this study is not to predict anything, but to explore the potential of SGOPE data to augment existing knowledge. Taking an unsupervised approach for this project aligns more to the inductive method as utilised by grounded theory in qualitative work. By not defining any assumptions or definitions to the model about the outcomes or variables, the data tells its own story [279] and is therefore more suitable for this exploratory research question. Unsupervised methods use clustering ideas to learn by observation rather than example. The input data does not have either a pre labelled class or a class attribute. This avoids the need for manual annotation of the data that is required to create a trained dataset for the algorithm to learn from. The algorithms used are generally based on distance, enabling any type of attributes can be associated together. This can help to identify previously unobserved naturally occurring patient groups that may not be considered in preconceived cohort classifications [280].

5.5 Issues with a pure NLP approach – combine with linguistics

In addition to evaluating the effectiveness of Modafinil by manual and NLP sentiment analysis, the exploratory P1 study was able to show how posters were expressing their belief that Modafinil was the cause of whatever, if any, effect they noticed. Despite the perceptions that the development of ever larger language models is bringing us closer to the ultimate goal of natural language understanding (NLU) rather than NLP, the current reality is that no matter how big a language model or amount of computing power that is used, the level of understanding that is required to identify causal text is beyond pure NLP systems [274,275].

Humans interpret language in a completely different way to machines. The way that humans or animals learn and develop understanding is not from having everything around them being labelled, but by observing and interacting with their environment which leads naturally to the understanding of the relationships between the component parts [281]. Making the links towards understanding the relationships between all the object types requires looking at them in the context of the surrounding text. This skill is one that many human readers can perform naturally as part of the reading and understanding process but is one that is proving to be much more challenging to do computationally on much larger datasets.

One of the reasons for this is many of the methods used within NLP rely much more on the tokenistic features of the text such as word frequency, and rather less on the syntactic features that help identify the semantic meaning of the text. The pre-processing techniques such as stopword removal, stemming or lemmatisation or 'bag of words' (BOW) representations are all inherently subtractive and remove meaning from the text. Stopwords such as 'from', 'to', 'and' or 'the' often indicate relationships between entities. Stemming and lemmatisation both remove inflectional features such as tense, as well derivative features such as the 'er', 'ee' or 'ment' from the root 'employ' which make it harder to understand who or what is being referred to. Any of the NLP methods that utilise a bag-of-words representation then lose all sense of word order [200].

There is an argument that being able to develop methods capable of achieving the real understanding of text requires will require stepping back from the development of ever larger language models towards combining rules-based methods with ideas from corpus linguistics [200,282]. Incorporating some linguistic analysis into the methods also helps bridge the gap between personal experiences of the posts and the impersonal world of the biomedical model. As mentioned earlier, there is no standard approach or set of procedures in corpus linguistic methodology [47], but I will again use the combination of keywords, ngrams , collocation and concordance as used in the exploratory study to explore how posters are expressing their view of any perceived causal links.

The statistical methods behind keywords and terms extraction have been shown in the exploratory study to result in a reliable and replicable overview of the themes. However, lists of words in isolation do not show how the words are used in context, or explain why they are relevant. There are also differences between an entity, defined as ‘a thing with distinct and independent existence’ [283], and keywords which can include words that are spelt the same but that have very different meanings e.g., apple (computer or fruit). Looking at relevant extracted words and terms with concordance techniques will identify specific examples and show the context in which they are used.

5.6 Justification of method choices

For this phase of the project, my aim is to take an inductive approach allowing the data to generate its own themes as in qualitative studies but on a much larger scale. Discovering the latent ideas within the data is particularly important for projects such as this which aims to discover unknown unknowns from the data that can be used to augment existing knowledge. This would then be more generalisable to other use cases as well as being a critical step in the path towards genuine natural language understanding rather than processing [281].

One of the overall aims of this project is to develop a general methodology that can be applied to other forms of SGOPE or electronic health record (EHR) data. As pointed out in the umbrella review, previous studies have been criticised for not explaining their methods. For this type of analysis to gain trust within healthcare, it is important that both the systems and the methods are as transparent as possible. I therefore want to use methods that are flexible, explainable, inductive and that can be run on individual computers, rather than requiring expensive cloud-based facilities or having a large environmental cost. The comparison of methods within each task, allows me to identify the strengths and limitations of each. Where the methods are time intensive, I have compared the time to run each model. Balancing computational complexity with the output generated will help in the development of a general methodology that can be generalised to exploring patient perspectives of other health topics from other SGOPE or similarly unstructured clinical datasets.

5.7 Data used in the main study

The data analysed in the main study (P3) comprises a flat file, in csv format, received from Treato Ltd [284] of all 69022 publicly available social media posts and threads including the term Modafinil, Provigil, Armodafinil or Nuvigil as of July 2017. The file was extracted by them from their database of over two billion health related public domain posts and covered the years 2011 to 2016. All posts were written in English. Utilising this dataset gives the project additional validity as it covers a far wider range of data sources than it would have been practical for me to acquire in the timespan.

5.8 Coding process

Development of the analysis code was done using Python v3.8.5 in JupyterLabs v3.0.15. The coded methods were first developed by me using the smaller exploratory dataset which reduced both testing and processing time. Once the individual components were complete, the main dataset was then analysed using the same code. This approach also allowed me to compare the results from the coding methods to those found in the exploratory study.

5.9 Cleaning and pre-processing

The choice and implementation of pre-processing steps in NLP can have significant impact on the results, but this importance is currently often overlooked, both in application and its reporting in academic papers, with details often being unreported [285]. Data quality can be defined across six parameters: existence, validity, consistency, integrity, accuracy and relevance [286]. Cleaning the data before analysis is one of the most important components of any NLP project, and is especially crucial where the raw data is as noisy and messy as SGOPE [161].

The cleaning process is somewhat of a black art – each dataset will have its own characteristics and each project will require specific features from the dataset to answer the research question, but it is important to try to maximise the quality of the processed dataset for each subtask. For instance, in topic modelling the aim of pre-processing is to reduce noise and incoherence from the data [287] allowing the themes to emerge. Stemming and lemmatising words to their root form enables this, whereas when assessing effectiveness, it is important to retain all relevant detail to understand the nuanced context within the text. Taking too blunt an approach can result in the loss of potentially useful data. I decided to take a staged approach here, initially performing only the bulk of the transformation and parsing of fields to obtain a dataset that would retain an optimal level of quality and flexibility (Appendix I).

After making copies of the original dataset, the next stage was to explore the dataset, using OpenRefine v3.3 [288] both to understand its structure and to determine what cleaning processes were required before further analysis and model building [202].

The time-stamp field, originally in format 2011-01-01 00:00:00 UTC, was simplified to that of PostYear representing the year of the post. Line breaks, paragraph breaks and other additional spaces were removed.

The URL field was parsed to identify the main name of the website or forum. New fields were created for subsite names. Having extracted the site name it became obvious that many of the URLs contained either the name of the condition that was of primary interest to the poster and /or the title of the thread or question that they were referring to. Utilising the variety of clustering techniques within OpenRefine [288] enabled the grouping and extraction of this detail from the URL (Appendix J). Three new fields were created to represent the second level of domain name, the focus condition of the site if applicable, and the extracted thread titles. All references to dosage amount in mg were standardised to xxxmg.

The variety of data sources included had resulted in inconsistencies as to the location of the thread titles within the dataset. Those from reddit were included as a separate field, with some, but not all, also included at the beginning of the text field, while those from other sites were extracted from the URL parsing. To maximise the options for analysis the cleaned data was structured to include three additional new fields; Text, the response only, Title, the title of the thread, and TextWithTitle, a field where the thread title precedes each response. The final structure of the cleaned data is shown in Appendix K.

Duplicate detection

Removal of any exact duplicate records is a basic cleaning task, but the identification of near duplicates, where just one or two words are different or in different positions within the text is less straightforward. In short posts occurrences such as this may be coincidental, with genuine posters using the same few words to create the post content, or it could also be due to a manipulative action. Within the threaded conversations of social media, partial duplication by the inclusion of quotes from previous posts may be common [289].

Decisions about how to balance the need for clean data with the resource constraints required to identify every instance of partial duplication need to understand how the over representation of these instances may impact on the results. A recent study that explored this impact on unsupervised topic models found that the effect of duplication depends on factors including the number of distinct repeated strings and the similarity of the repeated strings to the rest of the corpus [289]. They also found that the impact was dependent on the choice of model, but that both types were capable of sequestering duplicate text if it did not overlap too heavily with the topics of genuine interest. If large scale duplication is suspected, then a combination of unigram deduplication and ngram deletion can reduce any negative effects [289]. Bearing this in mind, I identified and removed obvious duplicates within the cleaning process, accepting that some would remain. Posts with exact or truncated duplicate text were deleted if they were collection errors i.e., where forum, author, and date were also duplicated. Posts with duplicate text but posted on different sites were not deleted. A further 31 posts all by the same author comprising random irrelevant words to 'blastmytwitter.com' were also deleted.

Spam removal

In an ideal world, all the individual posts would be referring to an individual's experience of Modafinil, but the nature of social media means that the dataset will contain some spam messages from companies trying to promote or sell their products. A common cleaning step for spam removal is based on identifying all text including either "www" or "http". This assumes that these links appear only in spam posts. However, they are also included in non-spam posts, given as links for further information which may be useful in understanding where post readers are being directed to. As an initial step towards the identification and removal of marketing or spam, posts of authors making over 100 posts were identified, and manually checked to see if they were obvious spam or appeared genuine. Posts including either "www" or "http" were similarly sampled. Checking for potential spam identified 2205 posts included "www" and 1115 included "http". Manual reading of a random selection of 100 posts including 'www' showed that only 36 were obvious spam. The majority included links to sites for further information, while others included spelling variations such as "Awww", "whewww" or "meawwwwwwwwwwoww" so these were not removed in the initial cleaning process. After being used for potential spam detection the authorID field was deleted.

The restructured file was saved in csv format for the next stages. The TextOnly and Title fields were exported as two separate corpora text files for linguistic analysis. Keeping them distinct avoided the possibility of the repetition of the title words skewing any frequency-based analyses.

Normalisation

After the initial data cleaning, further pre-processing or normalisation is used to try to bring an element of standardisation to the inherently random nature of natural language. Without losing any of the meaning it is possible to reduce some of the variations, which allows for more efficient processing. Reducing the number of variables also reduces the dimensionality of the data, which is important for memory requirements, reduces the number of decisions that need to be made in the code and can enable the generation of more accurate statistics from the dataset. For textual data normalisation steps tend to focus on two main areas: sentence structure and vocabulary [290]. The order of normalisation is important. Dividing the steps into two groups: "pre tokenization" (for steps that modify sentence structure) and "post tokenization" (for steps that only modify individual tokens) is useful in considering the implications of any normalisation [290].

Below I list some of the common pre-processing / normalisation steps and considerations of their use.

Table 5-2: Normalisation steps and considerations

Step	Example	Considerations
Lower case	Hello >> hello	Standardising all text to lower case is useful for frequency-based techniques such as TF-IDF but - Initial capitalisation useful for identifying entities such as proprietary drug names Capitalisation can be used in text to show emphasis – often of emotion.
Whitespace removal	string.split(text)	Too blunt. Sees a word differently depending on the punctuation following it – stop stop. stop, stop!, stop? would all be classed as different tokens.
Remove punctuation		Possible loss of context? How to evaluate an ellipsis? 'Nt expresses negativity - an important concept
Remove non-alpha		How to identify dosage if numbers removed? Data may contain emojis URLs, hashtags or @twitter mentions within the posts may be useful
Repeating characters	Baaad > bad	
Correct homoglyphs	$tupid > stupid	Need to consider in conjunction with removing non-alpha characters. Might removal make text harder to interpret?

Step	Example	Considerations
Text expansion abbreviations		Easier to recognise or expand abbreviations before lower-casing
Remove pluralisation		Might some plurals be important? No of times an ADR occurred
Emoticons		Used widely within some SGOPE data. Can be a useful way to analyse emotions in SA. But can we always assume that the author has chosen the correct emoticon for their emotion?
Remove URLs from text		Often used as a method of spam detection, but maybe a blunt tool. Might be useful to understand where readers are being signposted to.
Remove hashtags / @mentions		A frequent feature of datasets from Twitter. Noisy, but can give clues as to meaning of text.
Remove stop words		Which common stop words might infer causal links?
Stemming		Removes the end of the word – lose comprehension?
Lemmatisation		Uses the root form of the word – lose tense?

Spelling

Misspellings are a significant issue in all forms of written language but are a particular problem in text from social media. Generally misspellings fall into two types, non-words and real world errors [291]. Non word errors can be classified as words that do not appear in dictionaries, while real word spelling errors relate to those that exist but may be used or interpreted in the wrong context. Generic solutions have been proposed such as using a distance measure such as Levenshtein, fuzzy logic or frequency based noisy channel model [292] but as yet there does not seem to be a standardised solution that can reliably be used [293]. Relying on generic methods can reduce the data quality by introducing two types of errors into the dataset. Those that check against sources such as WordNet may not identify words that are correctly spelt but used in the wrong context. Overcorrection can replace correct spellings used in the intended context with something else.

5.10 Descriptive Statistics

Initial descriptive statistics were compiled regarding the number of sources, number of posts from each source and the temporal trends of both postings and sites. Corpus statistics were generated. Basic statistics on the lengths of the posts were calculated and plotted. Unique condition names parsed from the URL field were identified and frequencies generated.

5.11 Theme identification

To identify the themes within the posts the first approach was to use topic modelling, a widely used unsupervised machine learning technique for discovering topics, themes, or subjects within large quantities of unstructured textual data [294–296].

In the exploratory study, which was too small to feasibly expect good results from topic modelling, keyword and term extraction were shown to map well to the manually identified themes. Keyword and term extraction techniques are fundamental tasks used to extract the words and phrases that are typical of the specific corpus. Topic modelling differs from keyword extraction in that it can be described as a method of finding groups of words, not necessarily keywords, that represent a topic [296].

5.11.1 Topic Modelling

Particularly suitable for exploratory and descriptive analyses topic modelling is a technique that can be used as a method of determining what people are talking about in social media by looking for underlying structure within the text [295]. Combining an inductive approach with quantitative measurement, topic modelling is a useful method for obtaining an insight into the concepts that are contained within documents in a similar manner to grounded theory [210,297], although it is not yet widely used in clinical NLP [280]. As a method, it is useful not only for retrospective analysis, but is also capable of identifying potential new areas of research interest [296].

Unlike clustering methods such as k-means clustering, where each document is considered as relating to a single topic, topic modelling utilises the idea that every post can contain one or more topics, and that every topic is a mixture of words [298]. It also improves on co-occurrence analysis as any latent semantic connections between words can be revealed without the words having to be in the same document [295]. The algorithms work by looking at how words and phrases co-occur with a corpus of text [299]. Identifying any patterns within enables them to then create groups or clusters that represent the characterization of the text. Generally the precision of the models increases with the size of the dataset [294] making them a pragmatic choice for this study.

Various methods of topic modelling exist, but a recent comparison of four of the most widely known methods, Latent Semantic Analysis (LSA), Latent Dirichlet Allocation (LDA), Non-Negative Matrix Factorisation (NMF), and Principal Component Analysis (PCA) on short text data such as that found in SGOPE, concluded that LDA and NMF were seen to deliver the most meaningful extracted topics [296] so were selected for comparison in this study.

Both these methods have the same input in the form of a BOW document-term frequency matrix and generate similar output although the mathematical processes are different. Although the LDA method is the most widely used method used for patient experience feedback [65], a previous study found that NMF gives better results than LDA when used for short texts [300]. Other comparisons between the two found that LDA output was more semantically interpretable with more distinct categories [301], while NMF was faster and therefore less resource intensive [302]. However another comparison found the opposite [303]. As part of the project involves identifying a methodology for this type of data that can be developed to use on other datasets, I ran both methods and compared the findings as they relate to SGOPE data.

The narrative field used for analysis was the combined TextWithTitle, as this contained the most context. Initial pre-processing included transforming the text field to lower case, removing punctuation, and tokenising the text to words. Although the removal of URLs within the text is often suggested as these are generally seen as irrelevant [295,304], the cleaning process had shown how many of these links were to other resource sites given as suggestions of further information rather than for marketing purposes, so I did not remove them.

The standard nltk stopword list was extended to include removal of the most common terms for Modafinil (Appendix L), as these are expected to be in each post and as such would be over-represented in the results [305]. Bigrams and trigrams were identified. Restricting the dictionary to the POS tags of noun, adjective, verbs and adverbs reduced the dictionary size, and computational complexity [306]. Lemmatization reduced individual words to their root form. Lemmatizing was chosen rather than stemming because the resulting processed words are easier to read and comprehend [307]. Each post was analysed as a separate document to enable the distribution of topics to be viewed.

The ability of topic modelling methods to achieve a meaningful segregation between topics is dependent on five main factors: the quality of the pre-processing, the variety of topics that the text contains, the choice of algorithm used, the number of topics the algorithm is asked to look for and the specific tuning parameters of the algorithm [295,308].

Both LDA and NMF have different parameters that can be set, but for comparison purposes I tried to standardise these as much as possible. Each require a value for the desired number of topics to be entered as a parameter. The SGOPE posts of this dataset are likely to contain any or all of text referring to any of the eight themes identified in the first study

– reason for taking Modafinil (symptoms / conditions), impact of symptoms, acquisition, dosage, any ADEs, effectiveness and outcome, so I used eight as an initial value. I then generated a range of models using different methods and parameters and then used both quantitative and qualitative evaluation techniques to select the most relevant to the research question [294].

For the LDA models, coherence and perplexity measures were calculated for each model after it had generated the topics. I also ran a separate NMF coherence testing measure aimed to detect the optimum number of topics from a range of 5 to 50. Timings were taken for each of the model generations for comparison.

The first method used genism v3.8.3 [309], currently one of the most widely used Python libraries for this LDA modelling [308]. From the input data the algorithm creates a bag-of-words corpus and a dictionary. The dictionary was filtered to removed words that appeared in fewer than 5 or over 50% of the posts to avoid over or under-representation in the output. The number of unique tokens in the dictionary was recorded, as this along with the other parameters impacts on the memory requirements for processing.

LDA is a probabilistic method that represents documents as a mixture of topics that contain words with certain probabilities of occurrence [302]. It uses two probability values: P (word | topics) and P (topics | documents) to obtain the cluster assignments. The LDA algorithm has two important parameters, document concentration (alpha) and topic concentration (beta).

- High alpha – each document contains mix of most topics rather than just one – documents more similar in terms of the topics they contain.
- Low alpha – each document contains just one or only a few topics.
- High beta – each topic contains mix of most of the words – topics more similar in terms of the words they contain.
- Low beta – each topic contains just a few of the words.

The data is first transformed into two formats. One being a corpus, comprising each post as a document. The second, a term frequency document term matrix, where each row represents a document, and each column either a word or ngram. This matrix is used to perform topic modelling on the text. After an initial random assignment for each word in the document, the probability values are calculated multiple times in an iterative process until the convergence of the algorithm.

The gensim LDA models additionally require a value for the number of passes, or iterations through the posts that the model will perform. To find the LDA model that would give the clearest indication of the topics included in the data I ran the model repeatedly with different input parameters of topic number, number of passes, and the amount of data loaded into memory at any one time, calculating coherence and perplexity scores for each variation.

This model outputs the ten most salient terms for the number of topics, together with the weighting that it calculates for each word within the topic. Outputs for each included word lists of the top ten words for each topic, the number of documents having each topic as their main topic, and a bar chart showing the main topic for each document. Each model was saved for further analysis if needed. Outputs were visualised using pyLDAvis v2.1.2 [310], which generates an interactive visualisation to represent the topics found. The visual model shows the top 30 words per topic. The prevalence of each topic is indicated by its size and ideally the topic bubbles are distributed equally over the quadrant of the diagram with a degree of separation between them [308]. The visualisation can be adjusted to show varying values of α, the balance between words per topic and topic per document. Low values highlight lower frequency words that are more specific to the topic, while higher values show higher frequency words that may appear in multiple topics. Although the optimal value for topic interpretation will be dependent on the dataset, the authors of pyLDAvis suggest a value of 0.6 [311].

The next stage comprised using the same pre-processed dataset for both LDA and NMF analysis using scikit-learn [312] (version 0.23.1). Using NMF this library can be used to calculate the highest mean coherence value over a user defined range of possible topic numbers. I ran it over a range with minimum value 5 and maximum value 50 and used the suggested value to run the gensim models again, as well as the sklearn models. Other than the number of passes which is not required as input in this algorithm, similar parameters were used to generate the LDA model. Words appearing in less than 5 documents were removed and the maximum number of features capped at 10,000.

The initialisation method was set to 'nndsvd' as this is optimised for dealing with sparse data [312] is faster to execute [313] and gives more reliable results as it does not begin with random initialisation [298].

Outputs included the top ten words per topic, number of documents for each main topic, and the percentage of the corpus that each topic had been assigned to.

NMF is a based on linear algebra, and factors high dimensional vectors into a low-dimensionality representation. An NMF models aims to find the k number of topics in the selected data and to determine the degree to which each document and individual word belong to each topic. Using the concept of matrix factorization from linear algebra, the algorithm outputs two matrices:

- Matrix W contains scores that indicate how much each document belongs to a topic.
- Matrix H contains scores that indicate how much each word belongs to each topic.

Determining the optimal number of topics

One of the problems with current methods of topic modelling is the initial hyperparameter setting of the number of topics. Within qualitative analysis the iterative manual coding processes leads to the natural emergence of both the number and the subjects of the relevant themes. However, in topic modelling the number of anticipated topics must be set as a parameter before processing. Generally a higher number will return narrower and more specific topics, although too high a number will result in similar entities that cannot be distinguished between by humans in a meaningful way [295]. Choosing too low a number may return topics that are too broad [65]. Once the selected value has been entered as a parameter, the algorithm then attempts to fit the text to the specified number. Deciding on the optimal number of topics is recognised as being challenging [65]. An iterative approach, adjusting the pre-processing or parameters may be needed to fine tune the output.

Interpreting and evaluating topic models

Topics are often thought of as being discrete values such as food, sport, or religion, but in reality, this is far from the case. Topics can be broad or granular, and as in any informal output, whether text or speech, there can be a large degree of overlap between them. As with any form of result output, evaluation is crucial to ensure that the models accurately capture the internal structure of the corpus.

Selecting the most appropriate topic model from the variety of methods and models that exist involves a variety of trade-offs and judgements by a human researcher. Obtaining meaningful and understandable results is an iterative process. These trade-offs include processing time and capacity and the degree of granularity that is required [314]. Adjusting the granularity, by subdividing a corpus by a logical division, can be used to answer more complex questions, while comparing the topics identified from separate but related corpora can be used to show similarities and differences between them [294]

No matter how complex the mathematical underpinnings of each algorithm, the purpose of using topic modelling in this study is to identify the themes that are contained in the posts. To an extent this is a subjective decision as to whether the topics identified are relevant to the research question. The simplest interpretation method is to manually review the words with the highest probabilities (top words) for each topic and see if there is a label that can be applied that covers the substantive content [295]. Words that overlap or appear in several of the suggested topic groups, can make it hard to make sense of the output [315].

The previously accepted and commonly used measure of how well a probabilistic method evaluates topic models is known as a perplexity score [316]. This is calculated as the inverse of the geometric mean per-word likelihood, with lower values indicating better models [280], but it has recently been shown to be unreliable in correlating with human judgement [317]. Alternative measures based on coherence have evolved. In human terms coherence is seen as the extent to which the words of the topic make sense as a reflection of the content. Mathematically, topic coherence looks at each word identified in a topic and measures the degree of semantic similarity between them. Coherence is seen as a more reliable measure than perplexity, but human interpretation is still needed to relate the topics to the research question [295]. Generally, there is no gold standard list of topics to compare against, but in this case the themes identified from the exploratory study can be used as an evaluation metric.

Evaluation and comparison between methods

The initial comparison was between the LDA and NMF methods. As the statistical methods of calculating perplexity and coherence measures are known to be flawed, evaluation and comparison comprised a manual process of seeing which topics made the most sense, together with a comparison of the times each model took to run. The intelligibility was tested by looking at the top terms generated by each model to see if a brief label could logically be used to summarise them.

Top2Vec

My original intention was just to compare the LDA and NMF methods, but a new method, Top2Vec [318] was released in late 2020 which removed one of the main limitations of the existing methods, so I included it to the methods for comparison. Rather than requiring a number to be input for the desired number of topics, it takes a different approach and calculates for itself the number of topics within a corpus of text.

Using word vectors rather than the bag of words method, it assumes that many semantically similar documents are indicative of an underlying topic. By first creating a joint embedding of document and word vectors, it then looks for clusters of documents before identifying the words that attracted the documents together. This also removes the need for any specific pre-processing, stopword removal, stemming or lemmatisation [319]. The important parameters that require setting relate to speed. 'Fast-learn' is the fastest but generates the lowest quality vectors, 'Learn' produces better quality vectors but takes longer, while 'Deep Learn' aims to produce the best quality vectors but can take a significant amount of time to train [318].

Using version 1.0.24 on the original cleaned corpus, but without any additional pre-processing, as the LDA and NMF models, I ran it using both the learn and deep learn speeds to see which gave the best balance between accuracy and performance time. Using the 50 topic words output for each topic, I then manually categorised the generated topics from the optimal deep learn model into a single main theme, mapped where possible to those identified in the exploratory study. Many of the individual identified topics contained words that could apply to multiple themes, so a subjective decision was made as to which was most relevant, or it was marked as unknown.

5.11.2 Keywords/Keyterms

As an additional comparison I also used the technique from the exploratory study of using a corpus linguistic tool to identify the top 1000 keywords and keyterms specific to the post corpus. When generating the specific keywords from the dataset that represent the themes, it is important that the reference corpus used for comparison should represent the same type of language as that of the focus corpus to ensure that the statistical analysis identifies the words and phrases that are lexically distinctive and specific. In the exploratory study, I used the English Web corpus 2013 (enTenTen13) [203], a corpus of 19 billion words collected from online sources. For consistency in this study I will use the 2020 version of this reference corpus (enTenTen20), which now contains 38 billion words from online sources [320].

5.12 Effectiveness evaluation using sentiment analysis

One of the objectives of the whole project is to evaluate the posters view of the effectiveness of Modafinil. To assess these views I compared results from two widely used unsupervised sentiment analysis libraries, TextBlob v0.15.3 [208] and VADER v3.3.2 [321]. Sentiment analysis focuses on evaluating the subjective opinion of the author, rather than any objective “factual” information within the post. It can be a somewhat blunt tool. The accuracy is limited by factors such as ambiguous language or sarcasm [322] and is also dependent on the cleaning and pre-processing of the text. A paper that compared fifteen possible steps for sentiment analysis of Twitter data found that generally stemming, removal of repeated punctuation and number removal gave the best results, while punctuation removal, handling capitalised words, replacing negation with antonyms and spelling correction appeared to reduce the performance, although they conclude that results vary depending on the data and the subsequent processing task [293].

The original cleaned ‘TextOnly’ field was selected for the sentiment analysis as this contained only the responses to the posts. Capitalisation, punctuation and stopwords were retained for this part of the analysis as they each can contribute meaning or intensity to the analysis. Wordcounts were calculated for each post. The results from each are compared against each other and the manual analysis of the exploratory study.

5.12.1 TextBlob

One of the most widely used rule-based sentiment analysis libraries in Python currently is TextBlob [208]. Widely used to detect consumer opinion towards a product or service, it calculates values for polarity and subjectivity for each post. The lexicon it uses derives from a separate library used in NLTK. It focuses on adjectives from customer product reviews that have been tagged by humans for polarity and subjectivity [323]. Subjectivity analysis assesses how objective or subjective the text is, whereas polarity classification determines whether the text is positive or neutral [324]. It uses the sentiment lexicon to assign scores for polarity and subjectivity for each word, which are then averaged out using a weighted average to give an overall sentence sentiment score. Basic statistics were generated for both values, and the numerical polarity score converted to categorical values of Positive (>0), Neutral (0), and Negative (<0). To aid comparison with the exploratory study which included a Mixed category, I also tried two variations on this categorisation, one assessing Mixed as including polarity values of as >-0.01 to <0.01 and the other >-0.05 to <0.05. The data with the calculated columns was then saved to be used as the input for the VADER analysis to allow for comparisons between the two methods.

5.12.2 VADER

The methods behind the design of the VADER library make it possibly a better choice for sentiment analysis on social media type posts [325]. Rather than calculating the polarity and subjectivity of a post, it scores each post on four aspects, positive, negative, neutral, and compound. The positive, negative, and neutral scores represent the proportion of the post that fall in these categories. The compound score is calculated from the other three, normalised to a value between -1 and 1 and represents the overall sentiment of the post [321]. The lexicon it uses is based on general language rather than reviews [323] and contains around 7500 words. Validation was achieved using a team of ten human annotators to rate each word both as being positive or negative, and then scored on a value of +4 to -4 in terms of the intensity of it [321].

Although the basic sentiment is calculated on the individual words, VADER looks at the whole text and can take account of negations [326]. This can help to give a balanced assessment when the post contains contradictory words out of context as in Figure 5-1.

```
In [46]: # VADER example

         sentiment_dict = sid.polarity_scores("This shit is wonderful")
         sentiment_dict

Out[46]: {'neg': 0.387, 'neu': 0.215, 'pos': 0.398, 'compound': 0.0258}
```

Figure 5-1: VADER contradiction example

It is intended to take account of some of the characteristics often seen in SGOPE data where features such as repeated punctuation or the use of capital letters can be used to signify stronger sentiment [321] as shown in (Figure 5-2).

```
In [46]:  # VADER  Punctuation and upper case intensifier example

           print(sid.polarity_scores("This made me feel better"))
           print(sid.polarity_scores("This made me feel better!"))
           print(sid.polarity_scores("This made me feel better!!"))
           print(sid.polarity_scores("This made me feel better!!!"))
           print(sid.polarity_scores("This made me feel BETTER!!!"))

           {'neg': 0.0, 'neu': 0.556, 'pos': 0.444, 'compound': 0.4926}
           {'neg': 0.0, 'neu': 0.534, 'pos': 0.466, 'compound': 0.5399}
           {'neg': 0.0, 'neu': 0.514, 'pos': 0.486, 'compound': 0.5826}
           {'neg': 0.0, 'neu': 0.47, 'pos': 0.53, 'compound': 0.6714}
```

Figure 5-2: VADER punctuation and capitals

The VADER lexicon comprises the validated positive and negative words [321] and can be further modified with domain specific words. Words that it does not recognise are categorised as neutral. I initially ran it with the default lexicon. After assessing the positive and negative words it had identified from a sample of posts at each end of the range spectrum, I modified the lexicon as shown in Table 5-3. The words removed were frequently used by marketing /spam posts, so their repeated use was skewing the post score.

Table 5-3: VADER lexicon modifications

Added to 'Negative' words	Added to 'Positive' words	Removed from 'Positive' words
headache	awake	credit
jittery	focus	free
rash	concentrate	accepted
tired	normal	approval
harmful	productive	
dissappointed	helped	
sleepy	grateful	
nightmare	miracle	
intolerable	lifesaver	

5.12.3 Comparison between methods

A sample of posts from each end of the range of the two methods were identified for evaluation. As both TextBlob's polarity score and VADER's compound score have ranges from +1 to -1, the results from each method will be compared at a post level to each other on the P1 dataset and also to the manual evaluation of the P1 dataset.

5.13 Identifying causal perceptions

The automatic identification and extraction of causal relations from text is an essential component of human cognition but is a very complicated challenge for machine-based systems and as yet remains unsolved [275]. 'Proving' causality is one of the holy grails of computational analysis methods, subject to both statistical and philosophical debate [80] with no single method being guaranteed to show all the relevant causal links [81]. While appreciating the complexity of the debates, the aim of this research is not to provide a statistical proof of effectiveness across the whole patient population, but to generate a better understanding of the patient experience of using Modafinil, by exploring individual patient's perspective of whether or not it is effective for them.

The basic task focuses on identifying cause-effect relationships between nouns [327]. Particularly in the English language there are few explicit lexico-syntactic patterns that reliably correspond to causal relations, while there are large numbers of examples where a causal relation can be seen although the pattern is far from unique, so it cannot be used as a general extraction rule [328].

Any expression of causal belief can either be explicit or implicit. Explicit is where both cause and effect are explicit in a single sentence, e.g. 'Modafinil made me feel sick ' or ' John killed Bob'. Implicit causality is harder to identify as the lack of connective words means that the reader needs to use background knowledge or reasoning to identify any cause > event sequence [327]. Explicit links can also be intra-sentential, inside the sentence or inter-sentential with the cause and the effect being in different sentences [327].

Within SGOPE data a significant proportion of the causal links between factors in a post may be inferred rather than defined explicitly. Particularly for some of the shorter posts, any dictionary or collocation analysis purely of the narrative content of the post may miss out on the meaning of the post, if the dictionary words or phrases either are either implicit or are situated further away from each other than the limits set for the co-occurrence or collocation analysis. In many cases it may the situating of the post in a particular forum or online space that implies the belief of causation, rather than any word combinations or phrasing that could be located.

To look for examples of causal belief in the main study, I used the same technique of comparing the SGOPE corpus with the English Web reference corpus (enTenTen20) [203] as in the exploratory study, I generated lists of the top 1000 keywords, keyterms and ngrams specific to the dataset to help identify both themes and examples of causal text. For each of the words or terms in the lists I included its frequency in the focus corpus, the number of posts it appeared in, and a calculated score based on the relative frequencies in each corpus. I then classified the top scoring 100 of each of the keywords and keyterms into themes and summarised the results to see how this technique compared to the topic modelling. It was apparent that the majority of the key ngrams could have many interpretations. It was however possible to identify those that indicated cause / effect belief, and those that indicated a temporal dimension. Combining these selected ngrams with concordance techniques revealed specific relevant sentences that expressed the poster's understanding of these sequential events.

5.14 P3 Study Design

Figure 5-3 shows the processes and methods that were used in the P3 study and the outputs that were generated.

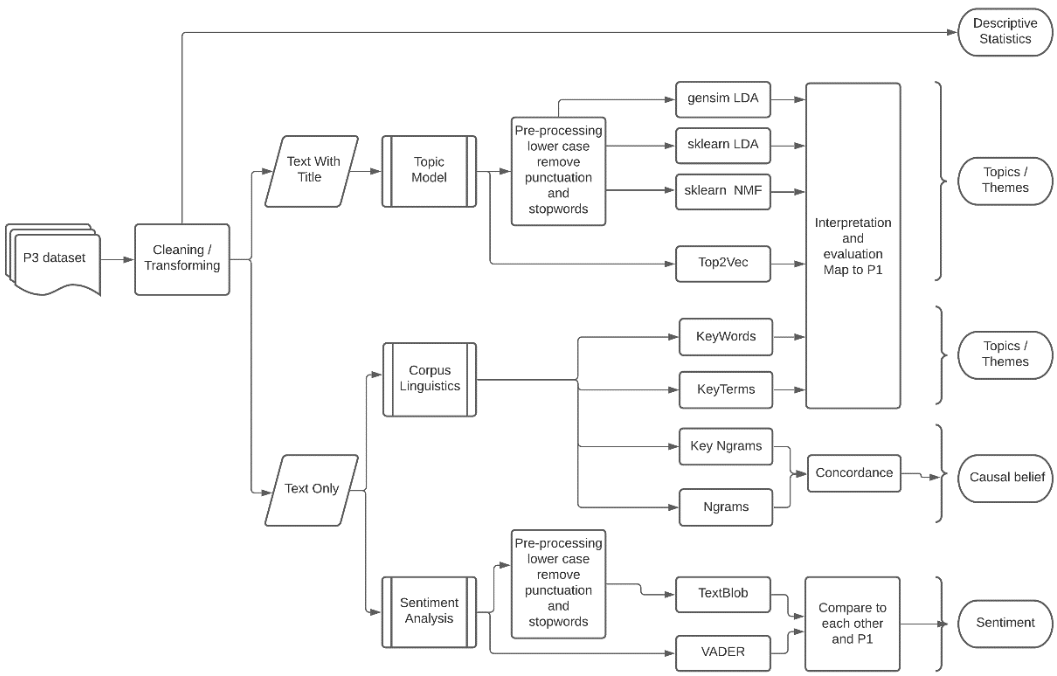


Figure 5-3: Study design for P3 main study

5.15 Evaluation of findings between P1 and P3

The final stage of the study is to compare the findings from the two studies and identify areas of agreement or gaps. One of the frequent comments from papers from the umbrella review referred to the lack of explanation as to how the various models and processes were evaluated. The project design address this by building evaluation into the process as shown in Figure 5-4. This enables the comparison of results from the automated methods to the manual findings of the first study. The code used for the analysis of the main dataset was developed using the exploratory set to reduce the processing load and times.

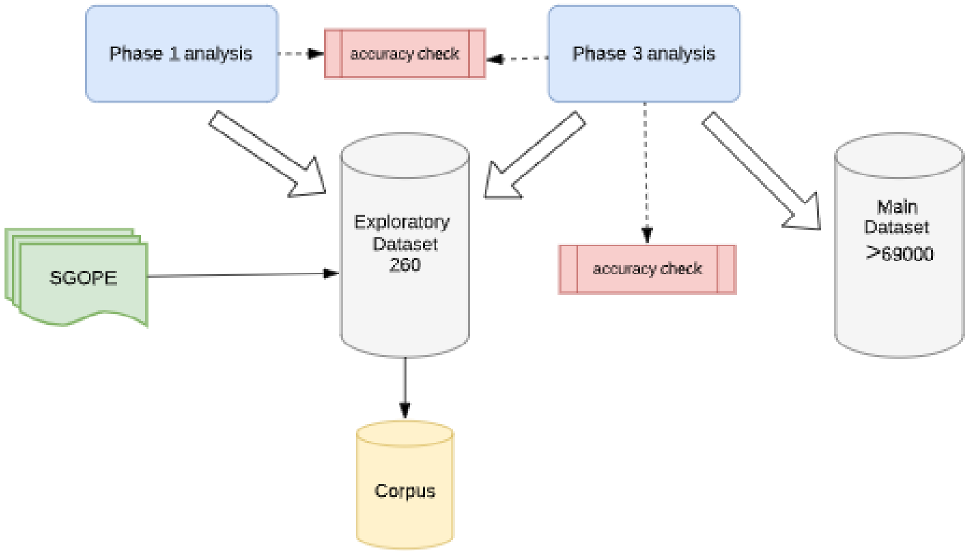


Figure 5-4: Evaluation between P1 and P3

Chapter 6 Main Study (P3) Results

6.1 Outline

In this chapter I describe the results for the main study. This phase of the study aims to develop a method to build on the findings from Phase 1 capable of performing a larger scale analysis of patient generated data relating to Modafinil experiences, and that can be adapted for wider use. Html links for the interactive visualisations of the results are included.

6.2 Aims and objectives

In line with the overall project aim, the aim of this study is to determine the extent to which the depth of findings of the exploratory study can be achieved on a much larger scale using computational text analysis methods. The objectives include:

- Identifying themes using topic modelling and keyword/keyterm extraction
- Evaluate effectiveness using sentiment analysis
- Use linguistic methods to explore inferred causal links
- Comparing methods of theme and effectiveness identification

6.3 Descriptive Statistics

The original dataset as supplied contained 69,022 records from a 6-year period of 2011 to 2016. The dataset included 27,760 distinct author names, with the most prolific author having made 204 non identical posts. A manual inspection of the 13 posters (0.0005 %) who had contributed over 100 posts, showed only one of the thirteen posters contributing obvious spam. Initial cleaning removed 301 duplicates (262 exact duplicates + 39 truncated), 1 post with no text and 161 obvious spam records leaving 68559 posts. The number of posts from each year is shown in Table 6-1.

Table 6-1: Posts by year

Year	No of posts	% of total
2011	8874	13%
2012	7219	10%
2013	9329	14%
2014	10380	15%
2015	16461	24%
2016	16296	24%
Total	68559	100%

After cleaning, 790 unique top-level sites were identified with the number of posts per site ranging from 25,355 to 1. Reddit.com was the largest overall source, with 25,355 posts (36.98%) from a total of 213 sub reddits, each of which represents a separate community. Five of the subreddits contributed more than 1000 posts, with the largest being the afinil subreddit with 12,870 posts.

The transient nature and popularity trends of online spaces is demonstrated in the analysis of the number of posts each year for some of the most frequent sites (Table 6-2).

Table 6-2:Individual site usage trends

Site	2011	2012	2013	2014	2015	2016
reddit	40	88	239	2939	10508	11542
facebook	841	1030	2303	910	816	619
Drugs.com	533	265	216	141	178	133
Talkaboutsleep	162	38	12	24	22	29
Apneasupport.org	81	41	34	18	-	-
Msworld.org	461	324	200	169	80	70
livingwithnarcolepsy	-	-	32	68	124	56

The posts were extracted from threaded conversations having 20,822 unique thread titles. The dataset included 8253 titles as a distinct field, with another 12,569 extracted from the URL address during the cleaning process.

Table 6-3: Data structure after cleaning

Cleaned Field	Unique Values	With value	Empty
PostID	68559	68559	0
PostYear	6	68559	0
SiteName	790	68559	0
SubSite	1042	32763	35797
SiteCondition (from SiteName or SiteCondition	166	18070	50490
TextOnly	68559	68559	0
Title	20,822	53970	14590
TextWithTitle	68559	68559	0
Author	27,760	68559	0
PostDate (retaining for duplicate checking)	2192	68559	0

Post lengths ranged from 1 to 1577 words, with a mean of 100 and an interquartile range between 34 and 132 words. The composition of the two corpora is shown in Table 6-4.

Table 6-4: Components of Text and Title corpora

	Title	TextOnly
Sentences	9162	480564
Words	330,420	6,836,862
Tokens	374,143	7,988,227
Unique words	16,301	104.565
POS Tags	62	63
Average post length (words)	6.2	100
Minimum post length (words	1	1
Maximum post length (words)	86	1577

Parsing the URL for the site or forum name, resulted in 166 separate health conditions being identified (List in Appendix N) from 18070 posts. Figure 6-1 shows analysis by the number of posts to the ten most condition specific sites. This does not assume that the specified condition was the primary or sole condition that the poster had but indicates a degree of choice that was made by the poster in choosing where to make their contribution.

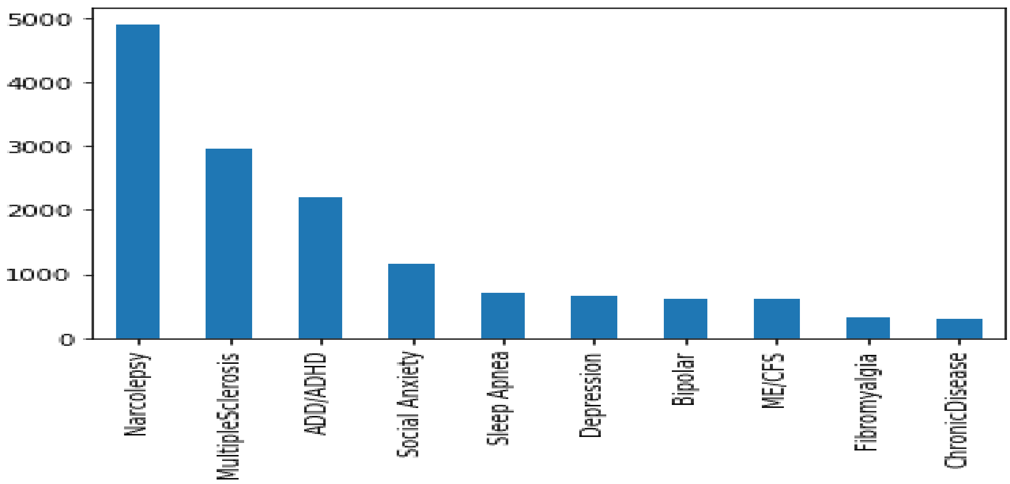


Figure 6-1: Post frequency to condition specific sites from data source URL parsing

6.4 Theme Detection

6.4.1 Gensim LDA

Topic modelling was main method of theme detection. First, I used the gensim LDA library. After pre-processing and dimensionality reducing methods, the dictionary comprised 12792 tokens. Initial parameters were set to 8 topics (as per the number of main themes identified in the exploratory study) and 50 iterations.

The default output for the library is the top 10 words per topic, together with the weighting of each word within the topic. The visualisation library for the genism LDA returns the 30 most salient words from the corpus and for each topic in the interactive visualisation (links below) and numbers the topics in order of the percentage of the tokens that are included within it. Although the returned topic word lists (Table 6-5) could all be seen to relate to the poster's experience, they did not appear to be clearly distinguishable from each other. The image from the visualisation as shown in Figure 6-2 indicates a significant overlap of topics 1-4 which between them represented 72.7% of the tokens.

Table 6-5: LDA t8p50 topic words

Topic	Word	% of Tokens
1	last night sleep stay work week dose keep feel long day really still go make first awake hour morning take well try need start tired time wake get even much	23.40%
2	could want work good see may help doctor think year really tell bad go make month take well try know say fatigue give time would medication find get med thing	22.70%
3	could experience use dose work seem low stimulant amphetamine feel people good high may really think make effect side take caffeine well drug try would find increase less even much	17%
4	thread experience want use seem come look people good information see think thank go question place first guy read know try order say time post would find new site get	9.60%
5	result prescribe study sleep use many people treat apnea condition disorder show may patient cause doctor think overnight make medical narcolepsy drug say case test pdoc medication diagnosis treatment even	7.90%
6	vitamin eye leave stuff use weight diet gain water lose exercise lot level make eat go play car food codeine drink part body healthy brain way big get supplement thing	6.60%

Topic	Word	% of Tokens
7	symptom study memory problem mind ability concentration abilify dream add cause improve cognitive help antidepressant social depression anxiety motivation brain parnate issue mental dexedrine focus bipolar mood increase thing life	6.40%
8	purchase cash free online ship shipping pill delivery company generic day pay fedex pharmacy prescription brand buy insurance script sale discount cheap order price need cost next cod require get	6.20%

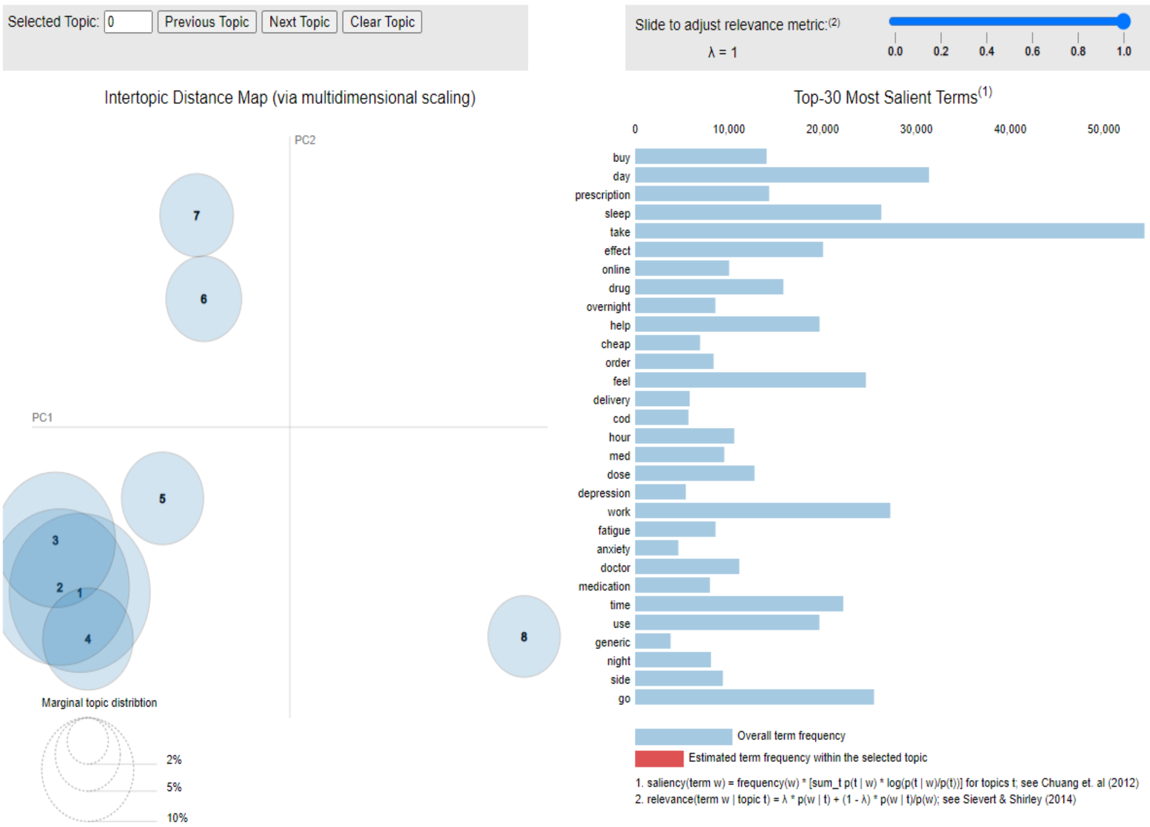


Figure 6-2: LDA 8 topics 50 passes

[LDA t8 p 50.html]

Initial coherence model testing using the NMF method on a range of values between 5 and 50 suggested that the optimal number of topics for this dataset was 27 (Figure 6-3), so I ran the models again with the number of topics set at 27 but varying the number of passes across the data.

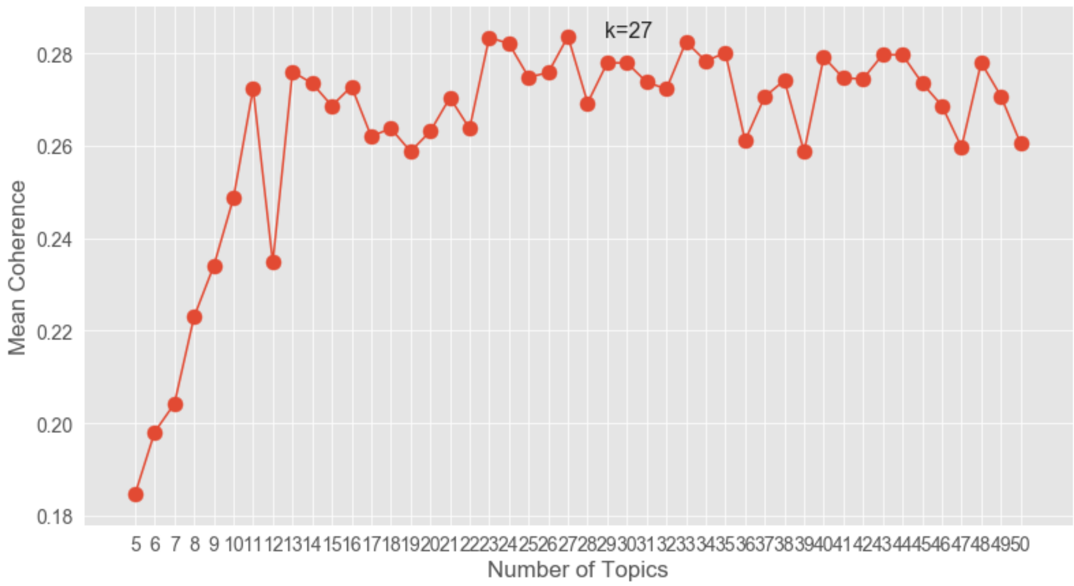


Figure 6-3: Coherence testing - range 5 - 50

Running the LDA model with parameters of 27 topics and 200 passes (Figure 6-4) showed a clearer distribution of topics, but still with a substantial degree of overlap of topics one to six. Increasing the number of passes to 1000 did not seem to give a significant improvement to the visual evaluation (Figure 6-5), although it took over five times as long to run (Table 6-7).

Although both the visualisations show some distinct topic circles that are not overlapped by others, categorisation of the topics into themes was not possible as the majority of them could have multiple interpretations. The top 10 topic words for each of the 27 topic models are shown and the attempted mapping are shown in Table 6-6.

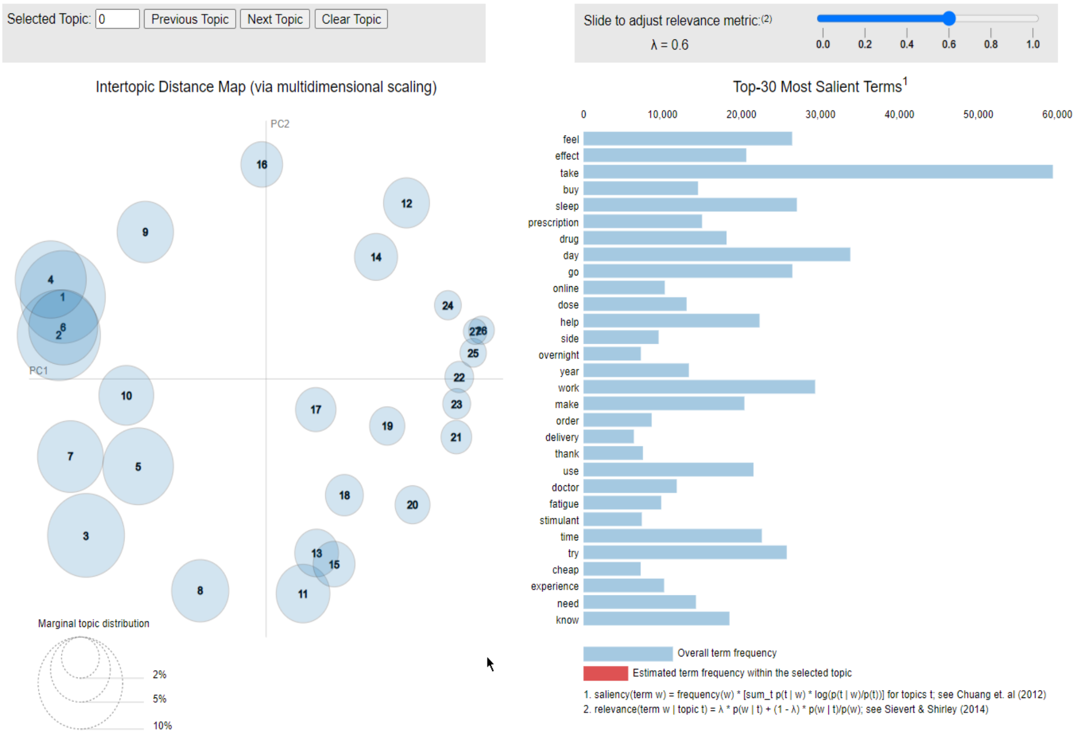


Figure 6-4:LDA t27 p200

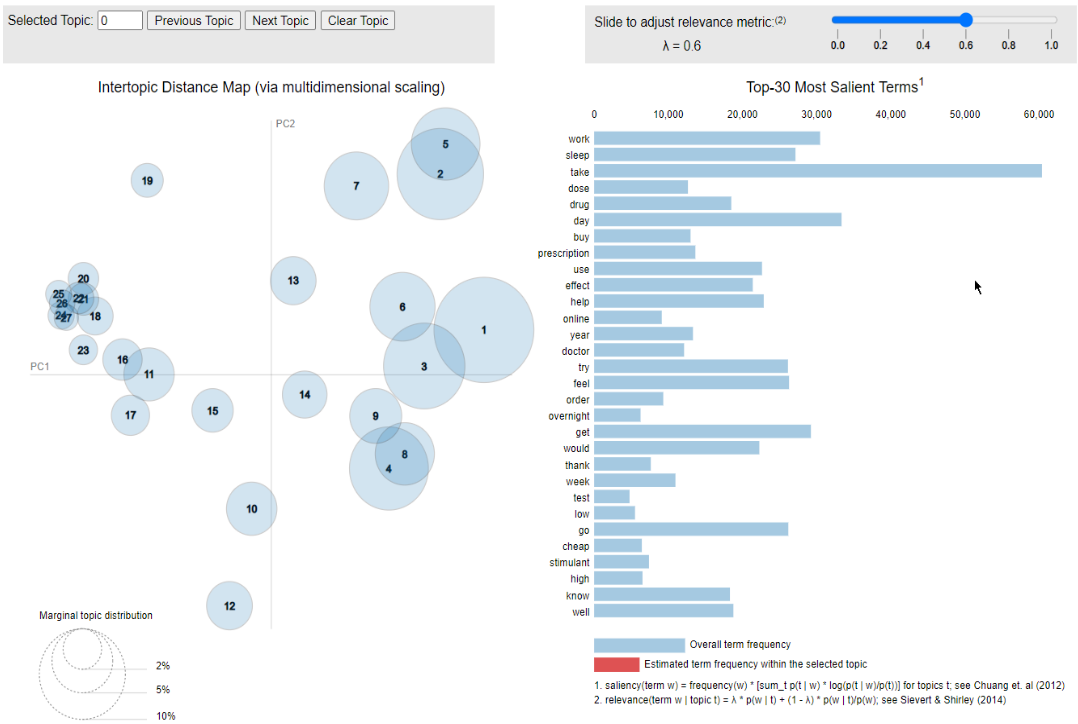


Figure 6-5:LDA t27 p1000

Table 6-6: Top 10 topic words for LDA models with 27 topics

Topic	T27 P200 Top 10 Words	Theme	Topic	T27 P1000 Top 10 Words	Theme
0	look http offer send check post find site new product	Spam?	0	result request process future testing test multiple access message impossible	
1	people addiction study risk take use smart report drug research		1	abilify dosage lower high half small dose increase level low	Dosage
2	try feel first start time day take dose work week	Dosage	2	know go make say feel would really think get thing	
3	know could people effect would may use think seem much		3	study memory improve focus motivation help cognitive increase energy concentration	Effect
4	thread need fast shipping ill next non group cash overnight	Spam?	4	people effect experience may side would take use cause drug	Effect
5	amantadine present pre alternative adderal parent functional deliver solution card	Spam?	5	eat appetite weight food loss lose exercise supplement diet healthy	
6	step delivery awesome up discount trial alot free complex comparison	Spam?	6	stimulant try amphetamine xyrem add depression would anxiety help well	
7	stimulant amphetamine make high add vyvanse dose drug abuse low		7	need dosage exam tolerance build day take use daily week	Dosage
8	try good fatigue pain find take help work well med	Effect	8	feel make awake day morning take help stay work keep	Effect
9	know go make time hard really work get thing life		9	job sleeping shift disorder home bipolar disease schedule work well	
10	theanine eat water drink moda alcohol cup caffeine lot coffee		10	thread http people list helpful safe use smart drug research	
11	know try go say could want would take think get		11	try symptom fatigue find medication take help doctor year med	Symptom impact
12	go last month start time ago take stop year week	Temporal	12	get go feel first start time day take year week	Effect
13	question appreciate experience advice post info read share wonder thank	Info Sharing	13	withdrawal mental stress self physical walk body health activity brain	
14	say ask prescribe narcolepsy would sleep medication see doctor tell		14	lol man cocaine fibro kill powder benzos enjoy abuse woman	
15	good focus memory improve energy help motivation use supplement brain	Effect	15	try know look question experience post would see read thank	
16	depression effect experience cause side anxiety take bipolar	Condition	16	go nap night time day study sleep hour wake get	

Topic	T27 P200 Top 10 Words	Theme	Topic	T27 P1000 Top 10 Words	Theme
	medication antidepressant				
17	return social mental combo handle smoke room physical push pattern		17	feel make effect experience drink take caffeine get coffee much	Effect
18	combine curious measure fail success ritalin narcoleptic mix hurt will		18	order price buy delivery generic day pharmacy online cheap prescription	
19	eye relapse line highly daughter fully chemical bupropion word recall		19	say month ask insurance prescribe expensive tell doctor get pay	Acquisition
20	abilify cymbalta reaction tip effexor pdoc crazy cfs worried skin		20	parnate natural simply learn stage parent attention placebo listen recall	
21	blood stress heart body test dopamine system increase level brain		21	treat disorder narcolepsy patient cataplexy medication doctor antidepressant treatment overnight	Condition
22	cover cost insurance company script generic drug pay pharmacy ship	Acquisition	22	tab heart machine touch pain evidence block leg concerned ms	
23	try get feel make seem bit really well bad much		23	fast tianeptine money bottle game waste alcohol total deliver previous	
24	order price buy day purchase receive online cheap prescription modalert	Spam?	24	order place pill good receive use site get product modalert	Acquisition
25	feel night awake time day hour sleep take work get	Effect	25	return legal arrive country afraid complete addict letter page detail	Acquisition
26	go husband right back start sale call see today get		26	appointment dream have move old kid diagnosis year be school	

HTML links to the three gensim LDA models are at:

LDA t8 p 50.html

LDA t27 p200.html

LDA t27 p1000.html

Timings

In terms of the processing load, the timings of the genism LDA models were impacted far more by the number of iterations through the data than the number of topics selected (Table 6-7). Adjusting the memory handling parameters reduced the processing time significantly but gave the highest coherence score to a model with just 2 topics and 10 passes which did not seem a plausible result. Using the default memory settings showed that a model set at 8 topics and 50 passes gave the highest coherence score, although no clear topics could be discerned from that model as shown in Table 6-5.

Table 6-7: LDA model timings

No of topics	No of passes	Time to run (default)	Time to run (adj for memory)
8	50	32 minutes	13 minutes
27	200	2 hr 16 mins	1 hr 44 mins
27	1000	11 hr 6 min	8 hr 13 mins

6.4.2 Sklearn LDA and NMF

Running the same 27 topic model with the sklearn library enabled a direct comparison of the LDA and NMF methods. Table 6-8 compares the top 10 words per topic, the number of posts each model classified as belonging to each topic, together with the percentage of the corpus per topic in descending order for each method. It also includes my evaluation of the theme that the topic words most closely indicated. As with the earlier gensim LDA models, trying to map each of the returned topic word lists to the themes from P1 was complicated by the degree of overlap in most of the lists. The bar graphs below (Figure 6-7 and Figure 6-7) show that the NMF method returned topics that were distributed slightly more evenly throughout the corpus, whereas the LDA version identified some topics that were much less represented. The sklearn LDA model allocated 94.45% of the posts to just 8 topics. The remaining 19 topics each represented less than 1% of the posts. In comparison, the largest NMF topic was assigned to 16.6% of posts, with the remaining 26 ranging from 5.4% to 2.0%. Future work could look at going back to the posts included in some of the smaller topics to assess their relevance to the research question.

Table 6-8: sklearn LDA and NMF topic distributions

sklearn LDA 27 topics						sklearn NMF 27 topics				
Topic No	No Docs	% of Docs	Top 10 topic words	Theme		Topic No	No Docs	% of Docs	Top 10 topic words	Theme
15	13774	20.09	fatigue take day sleep get work help like taking feel			0	11361	16.57	would know think good one also really could much something	
26	10685	15.59	like would take caffeine also noopept get day effects taking			26	3731	5.44	time first took got started years ago today last back	
11	10558	15.40	take day feel taking 200mg like effects dose caffeine 100mg	Dosage		1	3573	5.21	sleep night apnea study hours awake asleep wake narcolepsy cpap	
2	7421	10.82	order generic modalert uk buy get online brand prescription com	Acquisiti on		2	3010	4.39	fatigue ms helps amantadine chronic neuro pain helped prescribed years	
24	7195	10.49	sleep narcolepsy doctor get apnea know fatigue help work would			13	2731	3.98	generic modalert brand name sun order modvigil pills pharma price	
13	5994	8.74	depression wellbutrin adderall meds take anxiety help adhd like bipolar			7	2559	3.73	effects side effect negative term headaches experience long bad experienced	SideEffec ts
8	4658	6.79	dopamine adrafinil effects would like amphetamine liver drug effect drugs			20	2535	3.70	caffeine coffee theanine drink moda noopept nicotine cup tea energy	
23	4471	6.52	like get people sleep time would think drugs take really			16	2535	3.70	anxiety depression meds wellbutrin adhd bipolar add mood lamictal treatment	
9	554	0.81	fatigue anyone take helps tried help helped rs4680 works work			9	2449	3.57	insurance cover pay company doctor month expensive cost covered generic	Acquisiti on
25	447	0.65	buy online prescription overnight cod delivery cheap order without http	Acquisiti on		6	2383	3.48	dose 200mg 100mg 50mg morning mg dosage low doses half	Dosage

sklearn LDA 27 topics						sklearn NMF 27 topics				
Topic No	No Docs	% of Docs	Top 10 topic words	Theme		Topic No	No Docs	% of Docs	Top 10 topic words	Theme
12	424	0.62	www http com gov nlm ncbi html https nih greater	Info		23	2341	3.41	http www com html https org uk nlm ncbi reddit	Info
10	298	0.43	insurance entry div commenttext itemprop content class cover generic post			5	2316	3.38	adderall vyvanse vs xr adhd amphetamine dexedrine add ir mg	Other Int
21	281	0.41	vitamin water oil bulletproof diet coffee supplements vit magnesium drink			12	2115	3.08	feel like makes tired feeling awake better really felt normal	Effect
5	245	0.36	birth narcolepsy drug control world smart safe decision concluded creatively			10	2048	2.99	anyone tried else experience know wondering thanks ever taken experiences	Question
20	217	0.32	nardil heart parnate blood blunting pressure anxiety quote irritable emotional			11	1964	2.86	adrafinil liver prodrug 300mg vs metabolized noopept 600mg legal powder	
14	209	0.30	smart bbc limitless nightmare drugs documentary movie nzt semax journalist	Media		24	1942	2.83	ritalin adhd concerta tried stimulants methylphenidat e stimulant amphetamines vs add	Other Int
6	202	0.29	nsi 189 tapatalk chemically ltp sent tinnitus iphone induced saphris			18	1907	2.78	use days tolerance week daily using long term build phenylpiraceta m	Dosage
19	161	0.23	mdma ly master monk asshole mode oral smarter testosterone gabapentin			8	1904	2.78	day every awake night twice per hours next stay one	Dosage
22	152	0.22	sjs rash skin reaction allergic johnson syndrome stevens throat itchy	SideEffec ts		4	1884	2.75	take helps morning also need days sometimes everyday wake needed	
16	119	0.17	hair loss regrow sores contraception			14	1861	2.71	drug narcolepsy people smart prescribed medication	

sklearn LDA 27 topics						sklearn NMF 27 topics				
Topic No	No Docs	% of Docs	Top 10 topic words	Theme		Topic No	No Docs	% of Docs	Top 10 topic words	Theme
			tbi strike nuvagil msm expired						drugs safe world prescription	
17	101	0.15	trashy addicts phentermine favourite assistance thread claritin xd dexamphetamine nootriment			17	1826	2.66	taking started stop stopped week daily months weeks headaches years	Dosage
3	89	0.13	banned esports tournaments italy ban psychoactive uk bill doping microdoses	Legal		3	1805	2.63	buy online prescription overnight cod delivery order cheap without pharmacy	Acquisition
1	85	0.12	ireland geoff rx_rex custom hillary swd aching disease reporter gotmilk			15	1663	2.43	work shift disorder working well job home shifts worked full	
0	69	0.10	stacks er visit fear finasteride vs trying rc caffeine wat			19	1621	2.36	help need please awake energy stay focus meds pain tried	
18	55	0.08	pregnancy jaw teeth authoritative breastfeeding clenching nursing guide grinding ltheanine			21	1567	2.29	get done need go able bed back prescription script hard	
7	47	0.07	netherlands kg cons pros hype burnout xxx shortage procrastination dpd			25	1521	2.22	try give see want might maybe going could thanks next	
4	40	0.06	pots tension intas forehead hcg agomelatine muscle reminder friendly ppap			22	1372	2.00	works well great better worked hope best wonders find tried	Effect
28	8	0.01	None			28	35	0.05	None	
Total	68559	100				Total	68559	100		

Mapping the topics found by both models, even at a superficial level, to the themes from the P1 study was problematic. For the sklearn LDA model only 7/27 could be mapped to the general themes. The NMF model was slightly more interpretable with 14/27 that could be seen as relating to themes.

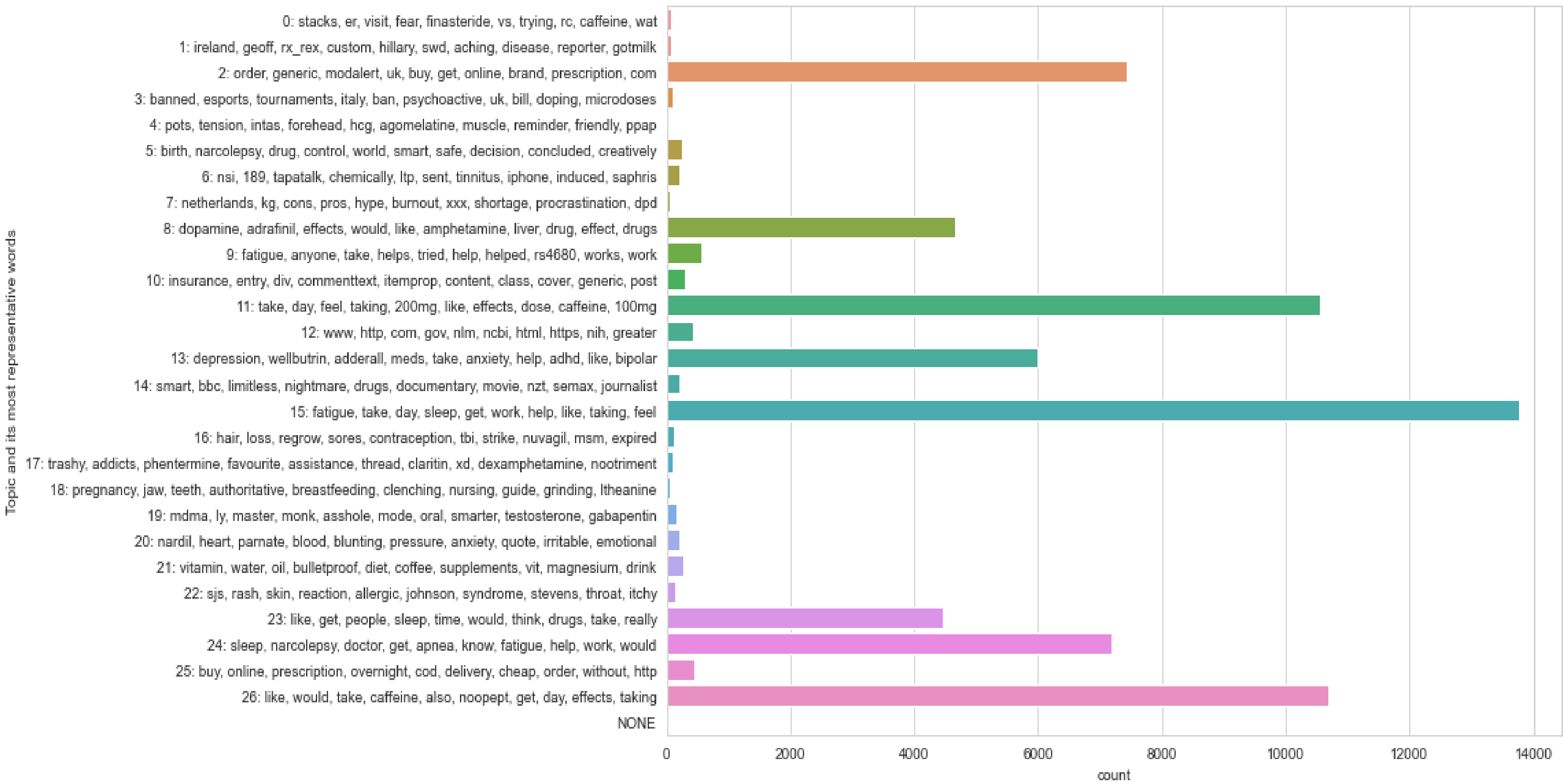


Figure 6-6: sklearn LDA 27 topics

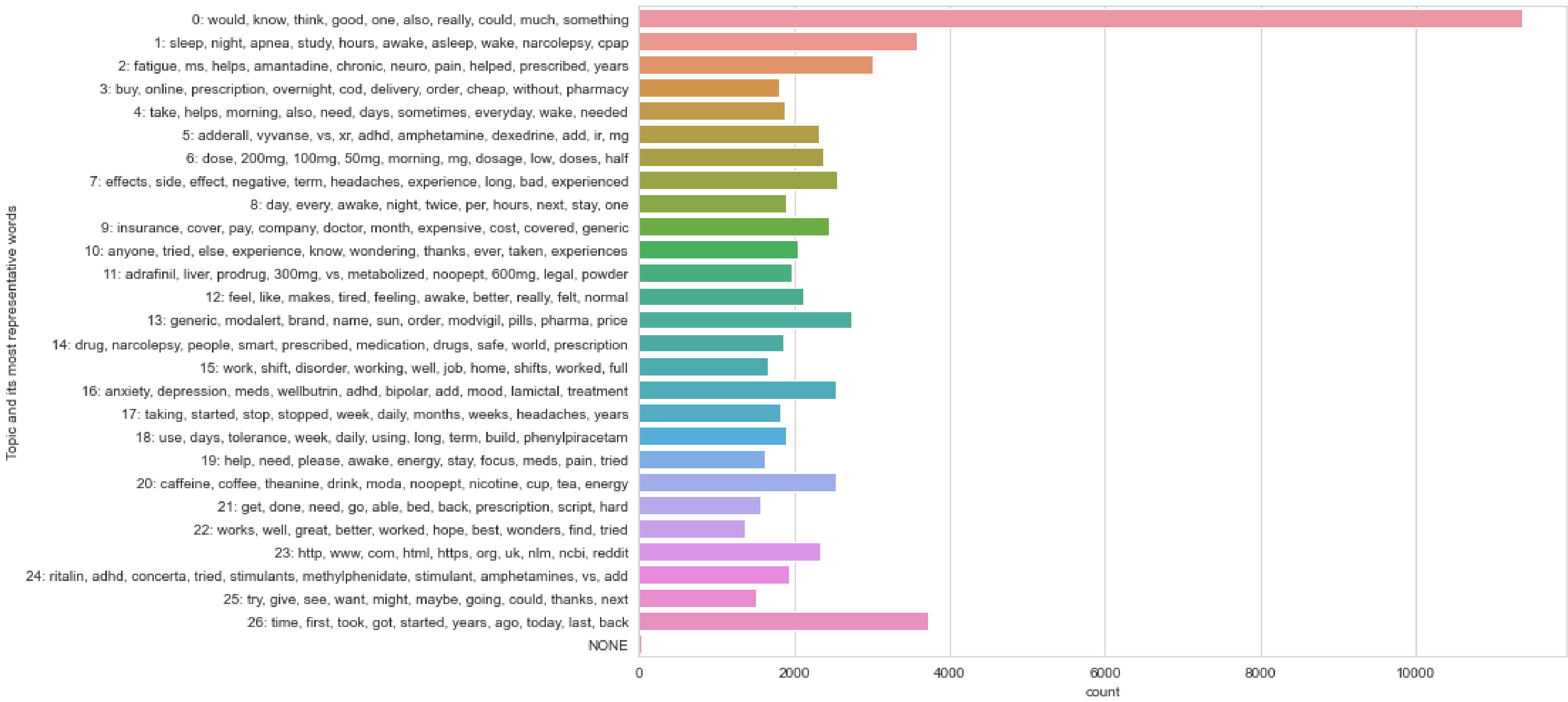


Figure 6-7: NMF 27 topics

6.4.3 Top2Vec

The Top2Vec library was substantially quicker than the LDA methods. By default, it returns the number of detected topics, the top 50 words per topic, and the number of posts per topic. The optimal DeepLearn parameter took 2hrs 15minutes and generated 367 topics, while the Learn parameter took 19 minutes to generate 566 topics from the dataset.

Results from the DeepLearn model were used for analysis. The number of posts per topic ranged from 2017 (0.029%) in the largest group to 45 (0.0007%) in the smallest. Overall 257/367 (70%) of the posts could be mapped to either the P1 themes or the codes used during the thematic analysis. 186/367 (51%) of the topics representing 38637/ 68559 (56%) of the posts could be mapped to the P1 themes. A further 71/367 (19%) of the topics representing another 15557/ 68559 (23%) of the posts were mapped to the codes.

In total 110 (30%) of the topics representing 14345/68559 (21%) were initially categorised as being uninterpretable without taking a deeper look at the specific posts. 31/367 (3913 posts) combined multiple themes so were classed as mixed, 50/367 (7019 posts) were uninterpretable so labelled unclear, while 29/367 (3413 posts) contained words indicating that the topics related to possible spam posts. Looking at these could help in further cleaning of the data. The full list of the topics, words, number of documents per topic and associated theme is included as Appendix M. Word clouds of the 20 largest topics in decreasing size order are shown in Figure 6-8.

Table 6-9: Top2Vec topics to P1 themes

Mapped to P1 themes	No Topics	No Posts
Condition / Reason	36	8491
Symptom Impact	12	4310
Acquisition	33	7135
Dosage	12	2446
SideEffects	25	4853
Other Interventions	30	6954
Effect	32	3503
Outcome	6	945
Totals	186	38637
	51%	56%

Table 6-10: Top2Vec topics mapped to P1 codes

Mapped to P1 codes	No Topics	No Posts
Mechanism of action	5	1876
Alcohol	2	748
Food	3	841
Education	3	959
Modafinil	3	1098
Information sharing	7	1383
Recreational	7	1124
RCTs	1	343
Vitamins	1	315
Emotions	4	716
Pregnancy /Birth	3	614
Enhancement	6	996
Temporal	2	336
Smoking	2	419
Ethics	1	250
Addiction	1	248
Information Sources	2	419
Biohacking	3	445
Web pages	1	230
Military	1	230
Circadian	1	199
Tests	1	196
Exercise	2	376
Withdrawal	2	340
Interactions	1	165
Tolerance	1	156
Media	3	384
Supplements	1	98
Pharma	1	73
Totals	71	15577
	0.1934605	0.227206

Figure 6-8: Top2Vec -20 largest topic wordclouds

Following topic generation, the posts can be searched by topic, and both topics and posts searched by keywords, enabling the easy identification of the posts that are of interest. As an example, one of the findings from the exploratory study related to a possible tolerance effect from long term daily use of Modafinil for some posters. A sample of topics related to dosage with the keyword 'tolerance' returned the following posts:

> *"Modafinil Cycle To Reduce Tolerance Build-Up Hi there, I have been taking 50mg of Modafinil a day for a few months to fight off Chronic Fatigue which has plagued me for the last 6 years. I originally started on a higher dosage of 200mg but tinkered with the lower dosage and found the drug to have the same outcome in relation to keeping me awake throughout the day without needing naps. However 50mg seemed to also allow me to get to sleep easier at night without any insomnia which was quite evident on the higher dosage. As this drug has really saved my life, I'm concerned that I may build up a tolerance to it which many users seem to report. Does anyone have any experience with cycling this drug for example 3 weeks on – 1 week off, etc, to lessen 137uvigil137137 buildup. Thank you "*
>
> *Document: 37450, Score: 0.2558336853981018*

> *"I feel like Modafinil no longer works With the half life of Moda, that isn't enough of a break to get it out of your system. You would need to take a 2-3 day break every 2 days to avoid all tolerance issues. It builds up in your system quite fast (as in, 2 days)."*
>
> *Document: 18382, Score: 0.5526922941207886*

> *"Modafinal and Socialbility? As with just about any other drug, you are going to rapidly build tolerance if using it for multiple consecutive days. If you don't want to build tolerance, only use modafinil once every 1-2 weeks."*
>
> *Document: 12505, Score: 0.5343003273010254*

"How do you stack/rotate modafinil? I don't use it anymore, but 4 days on, 3 off, or else 1 day on, 2 off when tolerance started to build."

Document: 29494, Score: 0.5251915454864502

As a vector-based model it also allows the identification of alternative words from the data that it calculates as being closest to a specified keyword. Using the example of negative emotion 'hopeless' as one of those that were identified in describing the impact of symptoms on the quality of life, it returned the following words that could be used to determine other posts of interest.

- depressed 0.3676547468326155
- afraid 0.33197382292634453
- scared 0.33171643134990514
- life 0.3301474397403381
- emotions 0.3171251355911894
- helping 0.3144332617646932
- everything 0.31367146761366105
- feel 0.31022423484375694
- point 0.3099103441929625
- self 0.3088865562090225
- living 0.3065375928464693
- fear 0.3064296832508233
- myself 0.30410450965727753
- struggle 0.30352447100804103
- become 0.29617116416773753
- desperate 0.29581570820160336
- come 0.2953263016354549
- feeling 0.29485036902249395
- happy 0.2938490165207627
- exercise 0.2915999637498477

6.4.4 Keywords/Keyterms

The top 1000 keywords (Appendix P) and keyterms (Appendix R) as calculated using the corpus linguistics method of comparing this corpus to a reference corpus were extracted. The top 100 keywords and keyterms are summarised below (full lists in Appendix O and Appendix Q). Of the keywords 74/100 mapped directly to the themes identified in the P1 study, 10 were name variations of Modafinil, and 6 reflected sub-themes.

Table 6-11: Top 100 keywords – corpus specific

Theme	No of keywords		No of docs	Frequency
Other Interventions		46	49249	68310
Stimulant	14		28396	39111
Antidepressant	7		2212	3504
Racetam	6		4164	5540
Nootropic	3		3956	5323
CNS depressant	2		1808	2962
SNRI	2		899	1260
Supplement	2		570	707
Acetamide	1		623	789
Anticonvulsant	1		386	686
Device	1		748	1095
Tranquilizer	1		595	771
Antipsychotic	1		407	572
Antiepileptic	1		598	795
MAO-B inhibitor	1		307	480
Nonstimulant	1		416	627
Plant	1		340	436
Adamantanes	1		804	957
Reason for taking		15	25007	33938
Modafinil		10	72604	110855
Acquisition		5	9278	25437
Dosage		4	16920	24980
Side Effects		2	925	1091
Effect		2	5701	6867
Amino Acid		2	1142	1333
Corporate Body		1	496	697
Investigation		1	414	606
HCP		1	1108	1339
Nutrient		1	925	1091
Mixed		3	15670	28957
?		4	8382	12021
Not relevant		3	975	975
Totals	46	100		

Other interventions were by far the largest group of keywords identified with 46 of the top 100 being mentioned in 49249/68559 (72%) posts. The comparison of the relative single word frequencies from this dataset to a reference corpus automatically highlighted nouns such as drug names. Further classification into the type of drug demonstrates the range of drug classes that are mentioned within the posts. After stimulants, antidepressants were the second largest class of drug mentioned.

Grouping the top 100 keyterms showed very similar results (Table 6-12) with 80/100 mapping directly to the original eight themes.

Table 6-12: Top 100 corpus specific keyterms

Theme	No KeyTerms	No Docs	Frequency
Acquisition	44	4812	12120
Reason	20	6916	7761
Dosage	10	2862	3088
HCP	3	775	897
Effect	2	344	374
Investigation	2	1152	1427
Modafinil	2	45	387
Side Effect	2	1316	1453
Outcome	1	158	159
Other Int	1	278	300
Mixed	5	1410	1590
?	4	1984	2048
Not Relevant	4	775	916
Total	100		

Table 6-13: Top 100 keyterms mapped to themes

KeyTerm	Freq	No of Posts	Score	Theme
sleep study	1273	1023	130.9	Investigation
overnight delivery	969	197	98	Acquisition
Saturday delivery	766	213	90.3	Acquisition
brain fog	854	714	89.5	Reason
sleep apnea	1633	1315	87.2	Reason
daytime sleepiness	779	713	83.5	Reason
low dose	835	767	63.6	Dosage
sleep disorder	597	547	59.3	Reason
shift work	563	524	56.4	Mixed
next day delivery	554	193	52.2	Acquisition
prescription buy	404	141	48.3	Acquisition
day delivery	585	202	47	Acquisition
chronic fatigue	654	598	45.7	Reason

KeyTerm	Freq	No of Posts	Score	Theme
buy provigil	371	46	45.6	Acquisition
ms fatigue	369	323	45.3	Reason
side effect	1214	1120	42.9	SideEffect
half life	408	365	41.5	Dosage
buy nuvigil	330	19	40.8	Acquisition
excessive daytime sleepiness	345	334	40.5	Reason
div itemprop	325	325	40.3	NotRel
sleep doctor	325	281	39.5	HCP
sleep specialist	320	293	38.1	HCP
post entry-content	304	304	37.7	NotRel
online pharmacy	658	296	36.3	Acquisition
cheap provigil	286	33	35.4	Acquisition
prescription provigil	276	37	34.3	Acquisition
empty stomach	414	383	34.3	Dosage

KeyTerm	Freq	No of Posts	Score	Theme
overnight shipping	299	165	34.2	Acquisition
generic provigil	254	183	31.4	Mixed
sleep doc	252	201	31.2	HCP
prescription nuvigil	245	22	30.6	Acquisition
online buy	294	143	30.1	Acquisition
non prescription	260	149	30	Acquisition
delivery provigil	232	28	29	Acquisition
141uvigil141 provigil	229	26	28.7	Modafinil
work disorder	226	207	28.1	Reason
excessive sleepiness	231	202	28	Reason
cheap nuvigil	223	17	27.9	Acquisition
cod Saturday	225	123	27.9	Acquisition
cod delivery	222	111	27.6	Acquisition
sleep schedule	246	227	27.5	Mixed
overnight cod	217	121	27	Acquisition
taking provigil	214	208	26.8	Dosage
overnight fedex	213	128	26.6	Acquisition
delivery cod	212	106	26.5	Acquisition
prescription cod	212	124	26.5	Acquisition
small dose	241	225	26.2	Dosage
sleep deprivation	340	296	25.5	Mixed
shift work disorder	202	189	25.3	Reason
extreme fatigue	222	216	25.2	Reason
delivery buy	201	98	25.1	Acquisition
141uvigil cod	196	20	24.7	Acquisition
high dose	300	287	24.4	Dosage
prescription overnight delivery	191	124	23.9	Acquisition
order provigil	188	31	23.6	Acquisition
141uvigil141 cod	186	24	23.5	Acquisition
fedex delivery	187	113	23.4	Acquisition
work sleep	187	180	23.2	Mixed
cod nuvigil	182	17	23	Acquisition
shift work sleep	175	168	21.9	Reason
delivery nuvigil	172	19	21.8	Acquisition
obstructive sleep apnea	227	212	21.8	Reason
work sleep disorder	173	166	21.6	Reason
obstructive sleep	234	219	21.4	Reason
first dose	214	205	20.9	Dosage
day supply	176	162	20.5	Acquisition
shift work sleep disorder	162	155	20.3	Reason
free fedex	160	84	20.2	Acquisition
141uvigil nuvigil	158	19	20.1	Modafinil
dry mouth	239	196	19.5	SideEffect
miracle drug	159	158	19.1	Outcome
cod provigil	150	29	19.1	Acquisition
fish oil	300	278	19	OtherInt
cod Saturday delivery	150	103	19	Acquisition
sleep latency	154	129	18.9	Investigation
extended release	161	146	18.9	Dosage
order nuvigil	148	22	18.9	Acquisition
good sleep	189	182	18.8	Effect
prescription order	152	104	18.8	Acquisition
term memory	185	162	18.8	Effect
severe fatigue	151	144	18.5	Reason
cognitive enhancement	150	130	18.4	?
online cod	152	95	18.1	Acquisition
second dose	164	140	18	Dosage
good luck	1585	1564	18	?
enough sleep	183	174	17.7	?
taking nuvigil	137	136	17.6	Dosage
insurance company	705	604	17.4	Acquisition
fedex cod	136	88	17.4	Acquisition
prescription next day	136	103	17.3	Acquisition
quote name	139	30	17.2	NotRel
post count	148	116	16.7	NotRel
short term memory	151	143	16.7	Reason
fatigue syndrome	193	179	16.6	Reason
smart drug	130	116	16.4	?
cod next day	125	77	16	Acquisition
prescription fedex	125	96	16	Acquisition
chronic fatigue syndrome	183	172	16	Reason
overnight buy	125	90	16	Acquisition
prescription next day delivery	124	95	15.9	Acquisition

6.5 Sentiment Analysis

6.5.1 TextBlob

The TextBlob library returns values for both polarity and subjectivity. Of the 68559 posts, the initial results for polarity were 47,282 (69%) positive, 6,229 (9%) neutral and 15,048 (22%) negative. Polarity scores extended across the whole range from +1 to -1 with a mean of +0.1003. Subjectivity scores also covered the entire range of 0 to +1, mean +0.4638.

Table 6-14:TextBlob basic stats

	Polarity	Subjectivity
count	68559	68559
mean	0.10030264	0.46380733
std	0.19852109	0.19614805
min	-1	0
25%	0	0.3787037
50%	0.08888889	0.48184848
75%	0.2	0.57081148
max	1	1

A paired plot showing the distribution and relationship between the polarity and subjectivity scores is shown in Figure 6-9.

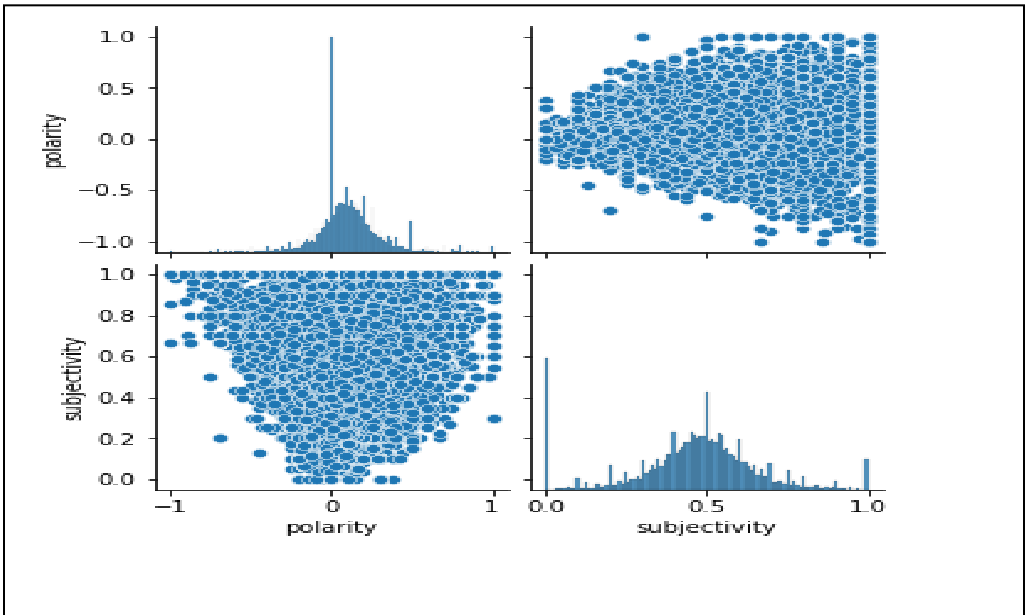


Figure 6-9: TextBlob polarity and subjectivity (all posts)

6.5.2 VADER

Using the same parameters of >0 being positive, < 0 being negative, the initial results returned from the standard VADER were 43,898 (64%) positive, 4,592 (7%) neutral and 20,070 (29%) negative. Modifying the lexicon resulted in 44,610 (65%) positive, 4,417 (6%) neutral and 19.533 (28%) negative. Compound score values ranged from +0.9997 to -0.9991, with a mean of +0.2825. The distribution is shown in Table 6-15.

Table 6-15: Basic stats for extended VADER

	Compound	Positive	Neutral	Negative
count	68559	68559	68559	68559
mean	0.2825079	0.11785168	0.8144244	0.06772396
std	0.61562543	0.09204523	0.1018511	0.06403353
min	-0.9991	0	0	0
25%	-0.1779	0.059	0.759	0.012
50%	0.4515	0.107	0.82	0.058
75%	0.8407	0.16	0.876	0.101
max	0.9997	1	1	0.67

6.5.3 Comparison between methods

Although the results from both methods were similar, with both showing a majority of posts being assessed as positive, comparing the distribution shape of the sentiment values between the methods showed distinct differences between the two. Both are skewed towards the right, indicating the positive mean value, but whereas TextBlob showed a normal type of distribution of polarity apart from those classified as neutral (Figure 6-10), Vader showed a similar peak at 0 but seems to assess more of the posts as being at the extremes of the available range (Figure 6-11).

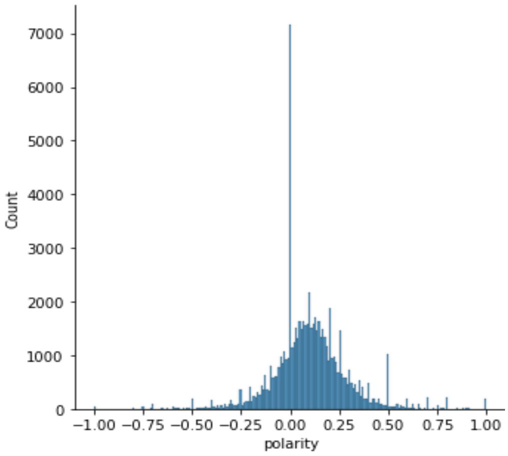


Figure 6-10: TextBlob polarity distribution (P3)

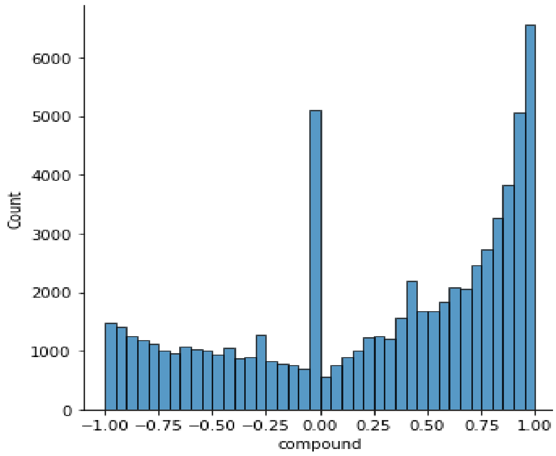


Figure 6-11: VADER compound sentiment distribution

Similar results were seen when comparing the length of the post and the sentiment score as shown below (Figure 6-12 and Figure 6-13).

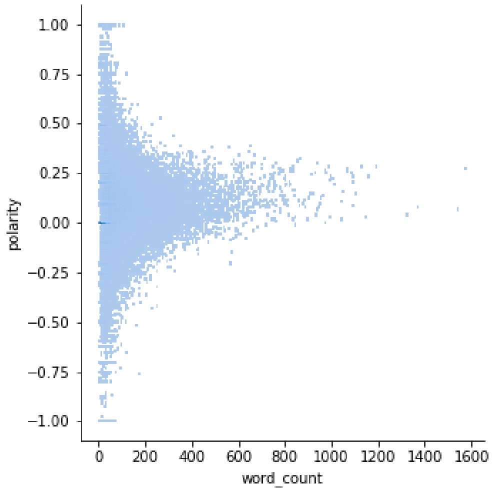


Figure 6-12: TextBlob- Word count to polarity

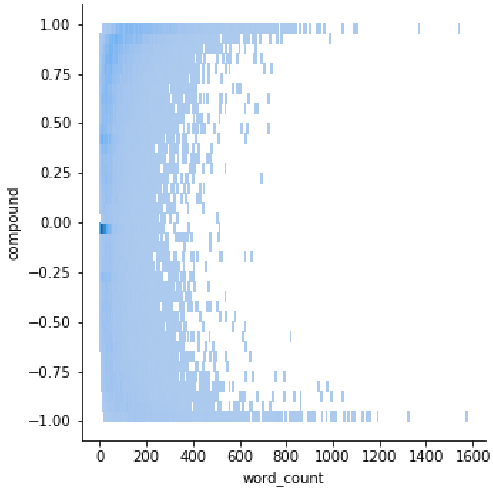


Figure 6-13: VADER – Word count to compound

The average word count of the ten highest rated posts by VADER was 704, and of the lowest ratings was 1095. For Textblob the average word count of the ten highest posts was 39, and the of the lowest ratings it was 23. VADER is reported as performing better on short texts [321]. The P3 dataset contained 1232 posts with a word count of over 400 and 8496 posts longer than 200 words. However, running it again on the reduced datasets showed little difference in the percentages of posts rated in each category (Table 6-15).

Table 6-16:Vader results at reduced word counts

	All Posts Standard	All Posts Extended	<400 Words Extended	<200 Words Extended
No of posts	68559	68559	67327	60063
Mean compound	0.2658	0.2819	0.2816	0.2658
25%-75%	-0.2040 +0.8250	-0.1794 +0.8404	-0.1779 +0.8438	-0.1655 +0.7984
Positive %	64.0	65.0	65.0	64.2
Neutral %	6.7	6.4	6.6	7.3
Negative %	29.3	28.5	28.4	28.5

6.6 Identifying causal text and linguistic analysis

Using the corpus linguistic tool SketchEngine to generate 1000 key ngrams (full list in Appendix T) specific to the SGOPE corpus identified many phrases that could infer a form of causality. Attempting to map the key ngrams to the individual themes was problematic. Of the top 100 (Appendix T) of the most specific ngrams to the corpus, only 16 could be mapped to themes directly. A full analysis would require looking at the ngrams in the context of the post. The key ngrams are however helpful in detecting expressions of causality. Unlike the individual words which all have a POS tag that can indicate tense (Appendix E), ngrams are combinations of words. It was possible to label many of them as relating to past, present, or future tense, or as indicating a possible belief. Examples are shown below in Table 6-17:

Table 6-17: Key ngrams indicating possible belief

Key Ngram	Freq	NoDocs	Score	Theme	Tense	Possible Belief
keep me awake	406	396	50	Effect	Present	Yes
works for me	408	398	49.2		Present	Yes
i have found	458	440	48.8		Past	Yes
but it does	488	485	48.3		Present	Yes
i find that	403	388	46.1		Present	Yes
was able to	610	579	46		Past	Yes
that i can	474	460	45.4	Outcome	Present	Yes

Key Ngram	Freq	NoDocs	Score	Theme	Tense	Possible Belief
i felt like	396	377	45.2	Effect	Past	Yes
i find it	407	400	44.4		Present	Yes
gave me a	395	389	44.3		Past	Yes
in my experience	377	365	43.8			Yes
because i have	381	380	42.6		Present	Yes
because i was	377	363	41.2		Past	Yes
because of the	576	561	34.7			Yes
and i think	368	363	32.3		Present	Yes
in my opinion	301	293	29.7			Yes
and it seems	258	257	29.7		Present	Yes
i have noticed	242	235	29.1			Yes
but i feel	241	237	28.7		Present	Yes
it gives me	230	226	27.9		Present	Yes
to kick in	225	217	27.9	Effect		Yes
seems to work	229	226	27.6		Present	Yes
it seems to be	237	237	27.4		Present	Yes
has helped me	225	219	27.2		Present	Yes
because i do	236	233	27.1		Present	Yes
effect on me	216	212	26.9	Effect		Yes
me feel like	220	216	26.9	Effect		Yes
it gave me	218	213	26.7	Effect	Past	Yes
changed my life	216	209	26.5	Outcome	Past	Yes
but it seems	231	231	26.3		Present	Yes
gives me a	216	210	26.3	Effect	Present	Yes
think it is	255	247	26.3		Present	Yes
as soon as i	227	223	25.9		Present	Yes
i can say	229	218	25.6		Present	Yes
it does help	205	204	25.6		Present	Yes
for me is	212	208	25.5		Present	Yes
i still feel	206	200	25.4	Effect	Present	Yes
my experience with	204	201	25			Yes
and i know	228	225	24.8		Present	Yes
thought i was	211	204	24.7		Past	yes
thought it was	237	233	24.6		Past	Yes
and it helps	196	194	24.4		Present	Yes

Key Ngram	Freq	NoDocs	Score	Theme	Tense	Possible Belief
know if i	208	204	24.3			Yes
i felt like i	198	188	24.1	Effect	Past	yes
i found it	209	202	24		Past	Yes
i thought it	229	227	23.9		Past	Yes
seems to have	242	234	23.5		Present	Yes
it helps with	185	183	23.2		Present	Yes
it has helped	187	185	23.2		Past	Yes
it seems that	232	227	23.2		Present	Yes
i know this	200	197	23.2		Present	Yes
feel like it	190	186	22.9		Present	Yes
because of my	191	188	22.9			Yes
am able to	189	178	22.9		Present	Yes
great for me	182	182	22.8			Yes
i can sleep	181	177	22.8	Effect	Present	Yes
i started to	197	186	22.8		Past	Yes
and it worked	186	186	22.7	Effect	Past	Yes
have found that	198	195	22.7		Past	Yes
give you a	228	226	22.7			Yes
and i felt	188	184	22.6	Effect	Past	Yes
it wears off	176	172	22.2	Dosage		Yes
a huge difference	183	180	22.2	Effect		Yes
better for me	177	176	22.2			Yes
this is a	642	627	22.2		Present	Yes
i found out	187	181	21.7		Past	Yes

The ngram ‘have found that’ was shown to be indicative of causal expression in the exploratory study. Using it on the P3 dataset and filtering out any of the sentences that did not explicitly mention Modafinil or one of its name variants in the concordance sentence returned the following examples.

Table 6-18: ngram concordance: have found that

PostID	**ngram concordance: have found that**	Theme

6289	I have been on Nuvigil for about 2 years now, and I have found that I have to skip my medication at least one day per week in order to not lose its effectiveness.	Tolerance
7711	I have found that I get visuals from modafinil anyways, for the first few hours of it's effects I have mild visuals and a solid body load.	Side Effects
26660	After taking modafinil 200mg next day i have found that i have a skin rash on the right hand and itchy skin on both hands.	Side Effects
29323	Forgetting and False Memories I am on Nuvigil, and I have found that I become a 'zombie' when they have my dosage too high.	Dosage
53387	I have found that I have been able to reduce my Prozac dosage while taking Provigil.	OtherInt /Effect
59900	I also have found that I am much more confident since started on provigil (200mg/day).	Outcome
67037	I have tried Adderall and Provigil and have found that I prefer a sister drug to the Provigil called Nuvigil, but my insurance company won't pay for it so I'm stuck with the Provigil or Adderall.	Comparison

The word sketch tool shows how any word or phrase is used within the corpus. Many of the key ngrams for this corpus relate to an observation the poster has made, or effect they have noticed in relation to the subject of their post. The most frequent key ngram in the corpus is ‘in the morning’ which appears 3016 times in 2627 posts. Using the corpus query language (CQL) to filter down to only those concordances that included Modafinil in the same sentence returned 183 examples of dosage patterns, amounts, drug combinations timing advice and effect. As with the P1 study, posters report how the standard dose can be excessive for some people:

“my Dr prescribed starting dose of 200mg modafinil ..once in the morning ...with the instruction that if the200mg did not keep me awake that I should double the dose to 400mg once a day in the a.m....the 200mg was too much all at once..all it did was enhance the side effects to the point that I wasn't able to notice if the medicine was doing what it was supposed to.because I was too busy cradling my cracked feeling skull and drinkn insane amounts of water...” DocID 3209

Another frequent lemma relating to effectiveness in the ngrams is ‘feel’ which can be used by post writers in many ways. As a verb it is used 22,767 times in the corpus. Splitting the occurrences into grammatical categories as in Table 6-19 shows the categories, some of the most frequent examples each phrase from the corpus, and the number of occurrences of each category. A visual representation of the most frequent adjectives and objects of the verb 'feel' are shown in Figure 6-14, with a diagram of the most frequent collocates in Figure 6-15. The size of each circle represents the frequency of the collocate. With 'good' being the largest adjective collocate of feel, this supports the The full list of collocates of 'feel' together with their frequencies in the corpus is included in Appendix U.

Table 6-19: Grammatical categories of 'feel'

Grammatical category	Examples	Frequency
pronominal subjects of feel	I feel, you feel, made me feel, it feels	12026
modifiers of feel	don't feel, I still feel, I just feel, really feel	6842
adjectives after feel	feel better, feel tired, feel worse, feel great, feel sleepy, feel normal	5342
objects of	feel the effects, feel a bit, felt nothing	4354
prepositional phrases	feel like, feel in, feel on, feel though	2163
subjects of	I feel, my body feels, I don't feel	2032
pronominal objects of feel	feel it, you feel you, feel myself	689
complements of feel	feel a lot better, felt it more, felt a bit weird	289
wh-words following feel	feel when, feel what, I feel that, feel how, feel normal which	179
feel and /or	sleep and feel, yawning & feeling	150
-ing objects of feel	felt taking, felt amazing	81
particles after feel	feel up to it, feeling down,	74
infinitive objects of feel	it feels to be	37
particles after feel with object	feel hyped up, to feel out	19

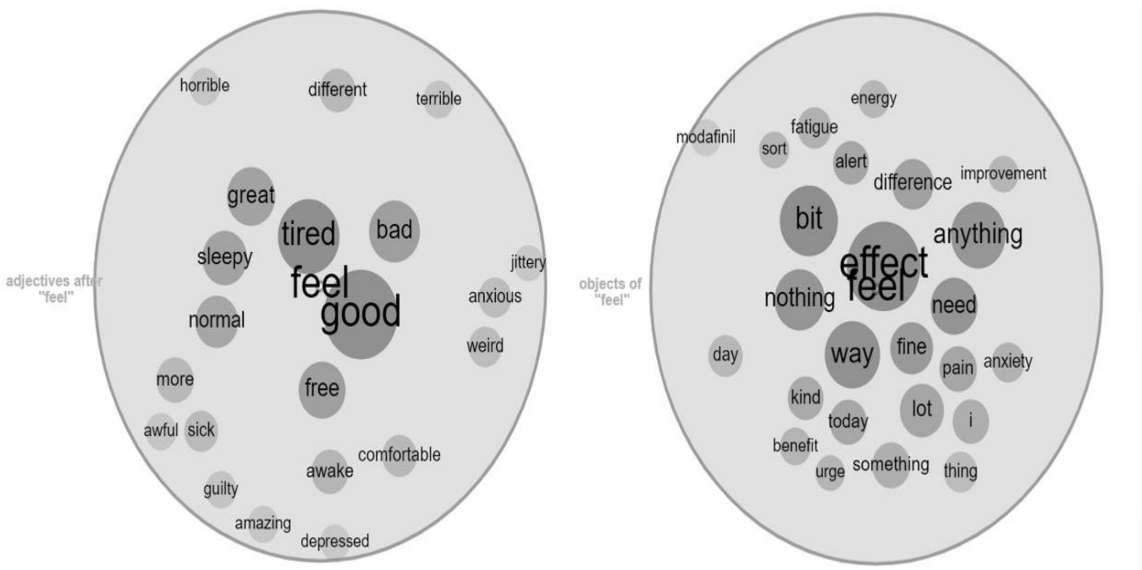


Figure 6-14: Most frequent adjectives and objects of 'feel'

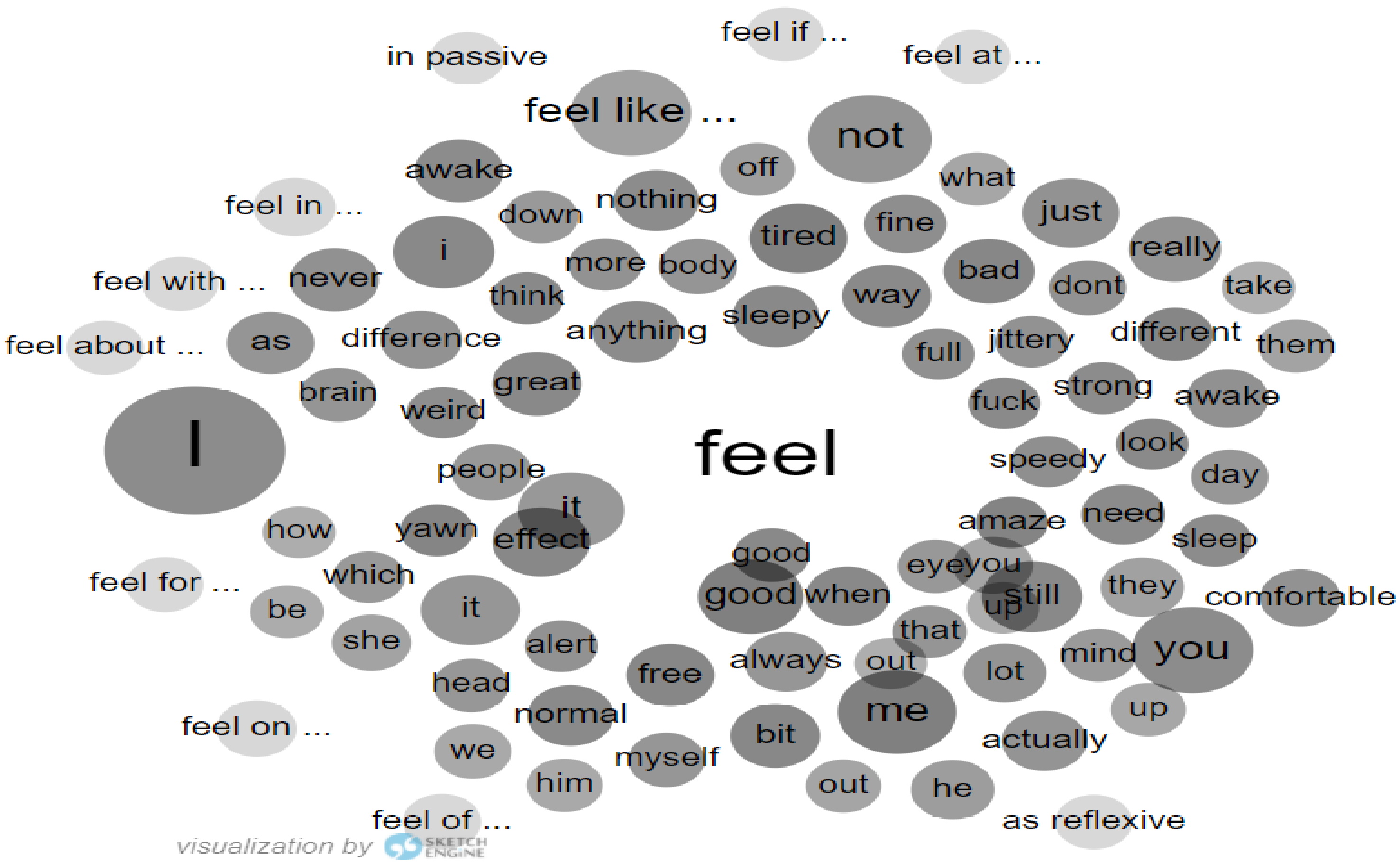


Figure 6-15: Word sketch of the verb 'feel'

Feeling normal was identified as being an important outcome for some posters in the earlier study. Table 6-20 shows examples of the 'make me feel' ngram concordances, filtered by 'normal'.

Table 6-20: Concordance 'makes me feel' and 'normal'

Example: concordance of ngram 'makes me feel' filtered by 'normal'
I have never noticed excessive energy or anything out of the ordinary; it just makes/make me/me feel/feel like a normal person would.
Taking the whole thing almost makes/make me/me feel/feel normal for a while.
Anything that makes me less sleepy makes/make me/me feel/feel more '' normal '' (i.e., less tired), and not high (course I am not shooting it in my arm or anything).
While Modafinil *feels* like a some sort of drug-induced happiness, Zoloft actually makes/make me/me feel/feel naturally normal and happy.
Cheers. :) I am on Modafinil which makes/make me/me feel/feel normal most of the time. @Nicole - I''m showing my age, but as a student it was ProPlus every time for me!
It just makes/make me/me feel/feel closer to normal.
At first I did feel speedy but now it just makes/make me/me feel/feel normal (ish)!!
Doesn''t jack me up or give me jitters - just makes/make me/me feel/feel as '' normal people normal'' as I can imagine.
My epileptologist has just put me on nuvigil for sleepiness and it really helps, there is only a day here and there it doesn''t but it''s awesome now most of the time I have the energy that my family has (2 kids) doesn''t make me hyper just honestly makes/make me/me feel/feel more normal.
I take Nuvigil, and, unlike stimulants, it just makes/make me/me feel/feel normal without the waves of crippling exhaustion or a crash at the end of the day.
Nuvigil makes/make me/me feel/feel like a normal person again and without it, my quality of life is severely decreased.
I love nuvigil and it makes/make me/me feel/feel '' normal '' and have a ''normal'' life but somedays I feel like I could use another pill and if its *safe* to take it twice a day then that may help me ALOT!!
I have read posts where people talk about feeling revved up from it but for me it just makes/ make me/me feel/feel normal .
The provigil makes/make me/me feel/feel normal.
It just makes/make me/me feel/feel normal which is perfect....no jitters.
It makes/make me/me feel/feel normal.
It makes/make me/me feel/feel pretty normal like I used too.
I usually take it around noon at work during the week and it makes/make me/me feel/feel normal, and I can get through the rest of the day.
It makes/make me/me feel/feel normal.
I am taking 200mg an hour before work and it makes/make me/me feel/feel normal .

I try not to take it every day, but it definitely helps.... makes/make me/me feel/feel normal almost.
I''ve now been feeling like it makes/make me/me feel/feel more '' normal '' (normal energy & focus) for a few hours past my dose (8am and 2pm) and the other times are like a complete drop in energy, not even normal tired....just SO exhausted.
(It wasn''t my first choice.) The only thing that makes/make me/me feel/feel close to normal is use of stimulants such as Nuvigil, but those give me serious insomnia.
I hate that a pill/pills makes/make me/me feel/feel normal .
It doesn''t make me feel buzzed or jittery, it just makes/make me/me feel/feel '' normal .

In the manual coding of P1 I was able to easily split each post into 'Pre' and 'Post' Modafinil which helped determine the base state, action, consequences for identifying causal text. This was not possible with the P3 dataset, but for future work I plan on using the tense of the POS tags to see if this can be replicated.

6.7 Comparison between P1 and P3 studies

The themes or topics had to be assessed using different methodologies due to the differing size of the datasets. In terms of theme identification, the Top2Vec model returned topics that were both the easiest to interpret, and to compare with the themes of the exploratory study. Although the mapping process was subjective, I assessed 186/367 (51%) of topics representing 38,637/68559 (56%) of the posts directly to the eight main themes of the exploratory study. A further 28 clear minor themes were identified, most of which had appeared as codes in the qualitative analysis rather than themes in their own right. As in the P1 study the top 100 keyword and keyterm extractions could also be successfully mapped.

As a guide for comparison Table 6-21 shows the number and percentage of each of the models or methods topics that were either uninterpretable or could generally be mapped to those from the P1 study. It should not be taken as a definitive guide, as the mapping decisions could be debated, and I only mapped the top 100 of the key words and key terms. The remaining 900 of each may alter the results from those shown below.

Table 6-21: Mapping topics from LDA and NMF models to P1 themes - percentages

	LDA 8 p50	LDA 27 p200	LDA 27 p1000	sklearn LDA 27	sklearn NMF 27	Top2Vec	Key Words	Key Terms
No Topics / Words	8	27	27	27	27	367	100	100
Mixed / Unknown	100%	52%	56%	74%	48%	22%	10%	13%
Spam		19%				8%		
Effect		11%	19%		7%	9%	2%	2%
Acquisition		3%	11%	7%	7%	9%	5%	44%
Reason for taking		3%	3%			10%	15%	20%
Dosage		3%	7%	3%	15%	3%	4%	10%
Symptom Impact			3%			3%		
Other Intervention s					7%	8%	46%	1%
Side Effects				3%	3%	7%	2%	2%
Outcome						2%		1%
Other Clear Topics		7%		11%	11%	19%	6%	5%
Modafinil				3%		1%	10%	2%

In terms of sentiment analysis, the results from both datasets were significantly positive, in line with the manual evaluation of the P1 dataset.

Table 6-22: SA All methods

	P3			P1		
	Pos	Neu	Neg	Pos	Neu	Neg
Te^{x}tBlob	69%	9%	22%	72%	4%	24%
VADER std	64%	7%	29%	55%	1%	44%
VADER ext	65%	7%	28%	57%	1%	43%
Manual				68%	14%[1]	18%

[1] Manual evaluation of P1 data included a 'mixed' category which I have categorized as neutral for the purpose of this table

The results from the TextBlob method were the closest to those from the manual evaluation. Matching them at an individual level in a confusion matrix as shown in Table 3-6 showed that at a post level on the four-level scale there was only 64% agreement, although this did raise to 85% agreement if I allowed for one level difference. The standard

VADER lexicon returned significantly lower numbers of positive ratings and higher negative ratings, although the modifications to the lexicon did improve this slightly.

Standardising the manual scale to a positive, neutral, negative grading by categorising the 'mixed' and 'unclear' grades as neutral and then re-analysing the agreement levels between each method and the manual grading of the P1 posts, showed that again that the TextBlob method had a higher overall agreement level than VADER (Table 6-23). However, comparing the agreement for each rating showed that while TextBlob identified 84% of the manually evaluated positive P1 posts, 8% of the neutral ones, it only matched 57% of the negative posts. VADER was more consistent, matching 72% of the positive posts and 77% of the negative posts. It failed to match any of the neutral / mixed posts, which resulted in lowering the overall agreement. Refining the range of values classed as neutral would improve this.

Table 6-23: Confusion matrices - TextBlob and VADER to P1 dataset

Manual	TextBlob NLP			Total		Manual	VADER NLP			Total
	Positive	Neutral	Negative				Positive	Neutral	Negative	
Positive	146	4	24	174		Positive	125	2	47	174
Mixed	28	2	15	45		Mixed	13	0	32	45
Neutral	1	1	3	5		Neutral	1	0	5	6
Negative	13	2	20	35		Negative	8	0	27	35
Total	188	9	62	259		Total	147	2	111	260
Accuracy	Positive	Neutral	Negative	Overall		Accuracy	Positive	Neutral	Negative	Overall
	84%	8%	57%	64%			72%	0%	77%	58%

In total 14/260 (5%) of the P1 posts were classified as positive manually, but as negative by both NLP methods. Conversely 6/260 (2%) were classified as negative manually, but as positive by both TextBlob and VADER. To try to understand why some of these variations were occurring I looked at two examples where the manual evaluation was positive but both TextBlob and VADER classed the post as negative, another two examples where the manual evaluation was negative but both TextBlob and VADER classed the post as positive, and one where the manual was positive, but the NP methods disagreed to compare how the methods differ in how they categorise the individual words.

Post AAP078 - Manual Positive, TB Neg, Vader Neg

> *Post: "Even after I started CPAP therapy I was still very tired and difficulty driving to work and staying awake when I got there. Provigil really helped. I'm still tired more than I would like to be but know*

without Provigil I don't think I could hold down a job. SideEffects No side effects that I'm aware of. I'm on a lot of other medication. "

VaderPos: {'helped', 'awake', 'like'}

VaderNeu: {'of', 'than', "n't", 'driving', 'medication', 'Provigil', 'more', 'a', 'work', 'very', 'to', 'staying', 'think', 'on', "'m", 'that', 'was', 'do', 'other', 'there', '.', 'Even', 'job', 'side', 'I', 'after', 'could', 'got', 'be', 'aware', 'CPAP', 'effects', 'would', 'started', 'still', 'hold', 'without', 'and', 'down', 'SideEffects', 'lot', 'but', 'really', 'know', 'when', 'therapy'}

VaderNeg: {'No', 'difficulty', '**tired**'}

Scores: {'neg': 0.122, 'neu': 0.795, 'pos': 0.082, 'compound': -0.437}

TextBlobPos: {'more', 'really', 'aware', 'very'}

TextBlobNeu: {'on', 'medication', 'know', 'think', 'lot', 'after', "n't", 'like', 'there', 'Provigil', 'would', 'do', 'driving', 'I', 'difficulty', 'got', 'effects', 'to', 'helped', 'still', 'that', 'when', 'work', 'than', 'side', 'could', 'awake', 'hold', 'a', 'staying', 'and', 'be', 'Even', 'CPAP', 'job', 'started', 'but', 'therapy', '.', 'No', 'of', 'without', 'was', 'SideEffects', "'m"}

TextBlobNeg: {'down', 'other', '**tired**'}

Post Polarity: -0.035793650793650794

D265 – Manual positive, TextBlob negative, VADERr negative

Post: "I just started taking Nuvigil. Today is day 2. Before Nuvigil I have been suffering for the past 3 years or so with marked fatigue. First I decided I should do a sleep study. They discovered that I have sleep apnea and I stop breathing about 50 times an hour. I got the CPAP machine and used it every night for the past 3 months. It didn't make the difference I was hoping. So, I talked to my Dr again and since I have young children I didn't want a sleeping medication and so he prescribed Nuvigil. Its been a life changer. I have the energy I used to have when playing with the kids and taking them outside instead of laying on the couch everyday."

VaderPos: {'hoping', 'energy', 'playing'}

VaderNeu: {'about', 'taking', 'should', 'night', 'to', 'been', ',', 'them', 'past', 'for', 'I', 'got', 'couch', 'did', 'started', 'sleeping', 'Its', '2', 'or', 'want', 'when', 'Before', "n't", 'discovered', 'a', 'Dr', 'on', 'that', 'do', 'is', '.', 'laying', 'Nuvigil', 'instead', 'study', 'and', 'every', 'changer', 'day',

```
'talked', 'make', 'of', 'apnea', 'medication', 'They', 'children',
'was', 'he', 'machine', 'times', 'outside', 'years', 'my', 'since',
'So', 'young', 'again', 'decided', 'breathing', '50', 'so', '3',
'It', 'difference', 'the', 'it', 'First', 'have', 'sleep',
'everyday', 'an', 'CPAP', 'just', 'with', 'prescribed', 'months',
'Today', 'hour', 'life', 'used', 'kids', 'marked'}

VaderNeg: {'suffering', 'fatigue', 'stop'}

Scores: {'neg': 0.065, 'neu': 0.885, 'pos': 0.05, 'compound': -
0.2699}


TextBlobPos: {'First', 'young', 'marked'}

TextBlobNeu: {'sleep', 'breathing', 'medication', 'taking', 'study',
'prescribed', 'Before', 'decided', 'kids', 'again', ',', 'hour',
'playing', 'and', 'CPAP', 'times', 'apnea', 'suffering', 'months',
"n't", 'it', 'about', 'I', 'got', 'night', 'Dr', 'to', 'couch',
'changer', 'that', 'started', 'talked', 'discovered', '3', 'did',
'hoping', 'difference', 'my', 'do', 'or', 'So', 'used', 'want',
'instead', 'when', 'Today', 'the', 'a', 'make', 'for', 'machine',
'been', 'life', '.', 'with', 'was', 'sleeping', 'is', 'on',
'fatigue', 'an', 'he', '2', 'children', 'outside', 'Its', 'day',
'Nuvigil', 'them', 'It', 'years', 'should', 'since', 'laying', '50',
'have', 'just', 'stop', 'every', 'so', 'of', 'energy', 'They'}

TextBlobNeg: {'past', 'everyday'}

Post Polarity: -0.03571428571428571
```

AAPNuv146 - Manual negative, TextBlob positive, VADER positive

Post:- First day was great (started ar 150 dose) then falling asleep during day. Increased to 250, didn't fall asleep during day but very nervous and couldn't sleep at night. Going to breakup dosage to see if that helps SideEffects Itching, cant sleep at night

```
VaderPos: {'Increased', 'great', 'helps'}

VaderNeu: {'150', 'then', 'if', 'dose', 'during', 'was', 'First',
'.', 'at', 'day', "n't", ',', 'Going', 'started', 'breakup',
'SideEffects', '(', '250', 'see', 'could', 'ar', 'very', 'that',
'did', ')', 'sleep', 'to', 'and', 'asleep', 'dosage', 'but',
'Itching', 'night', 'fall', 'cant'}

VaderNeg: {'nervous', 'falling'}

Scores: {'neg': 0.086, 'neu': 0.766, 'pos': 0.147, 'compound':
0.4785}


TextBlobPos: {'great', 'First', 'very'}

TextBlobNeu: {'sleep', 'ar', 'dosage', 'Increased', 'fall', 'did',
'breakup', "n't", 'dose', 'asleep', 'then', 'Going', 'night',
'falling', 'to', 'helps', '250', 'day', 'that', 'see', 'could', ',',
'(', 'Itching', 'and', 'nervous', 'started', 'but', '150', 'cant',
'.', 'at', 'was', ')', 'SideEffects', 'if', 'during'}

TextBlobNeg: set()
```

```
Post Polarity: 0.41666666666666667
```

AAP133 – Manual negative, TextBlob positive, VADER positive

> *Post: I suffer from severe CFS (exhaustion, constant sleepiness) as well as sleep apnea. Was taking 10 mg. of Adderall morning and noon and did great on that. But sleep doctor wanted me to try Nuvigil since it's supposed to be better... 150 mg. did nothing for my sleepiness so he upped it to 250 mg. This DID start really helping the sleepiness - like others, I still felt tired but like it was masked and I could at least stay awake during the day to get some things done. However, after being on it for 3 months now, I've started having really bad tremors (hands shake and I can't write straight!) - and also ringing in my ears (a buzzing sound that is CONSTANT) - this is driving me NUTS. Did anyone else get this ringing in the ears?! I stopped the Nuvigil a little over a week ago and the tremors are gone, but the ringing is still going on in my left ear - right ear finally stopped ringing 2 days ago. I'm hoping my left ear will quit ringing soon. I hope my doctor will put me back on the Adderall - I was not abusing it in any way and it at least was helping w/o these bizarre side effects! SideEffects Tremors (shaking hands) and constant tinnitus (buzzing in ears)*

```
VaderPos: {'helping', 'hoping', 'better', 'straight', 'awake',
'like', 'hope', 'great', 'well'}
```

```
VaderNeu: {'taking', "'ve", 'Was', 'hands', 'morning', 'felt',
'during', 'to', 'However', 'This', 'Did', 'from', 'Adderall', "'m",
'supposed', ',', 'this', 'for', 'side', 'I', 'start', 'anyone',
'ears', 'did', 'else', 'finally', 'started', 'wanted', '-',
'others', 'buzzing', 'things', '2', 'w/o', '10', "n't", 'driving',
'doctor', 'mg', 'a', "'s", 'masked', 'are', 'on', 'also', 'that',
'sound', ')', 'But', 'is', 'constant', '.', 'least', 'stay', 'over',
'Nuvigil', 'effects', 'tremors', 'not', 'gone', 'mg.', 'still',
'and', 'write', 'little', 'but', 'quit', 'any', 'day', 'of', 'ca',
'apnea', 'will', 'put', 'some', '(', 'days', 'tinnitus', 'being',
'ago', 'having', 'he', 'was', 'right', 'after', 'DID', 'my', 'be',
'since', 'going', '!', 'at', '150', 'week', 'CONSTANT',
'SideEffects', 'me', 'so', '3', 'now', 'done', 'CFS', '250',
'Tremors', 'the', 'ear', 'it', 'try', 'nothing', 'soon', 'way',
'...', 'these', 'noon', 'get', 'as', 'sleep', 'in', 'could', 'back',
'sleepiness', 'ringing', 'left', 'months', '?', 'really', 'upped'}
```

```
VaderNeg: {'severe', 'exhaustion', 'abusing', 'bad', 'stopped',
'bizarre', 'tired', 'shaking', 'shake', 'suffer', 'NUTS'}
```

```
Scores: {'neg': 0.121, 'neu': 0.745, 'pos': 0.134, 'compound': 0.7467}
```

```
TextBlobPos: {'First', 'young', 'marked'}
```

```
TextBlobNeu: {'sleep', 'breathing', 'medication', 'taking', 'study', 'prescribed', 'Before', 'decided', 'kids', 'again', ',', 'hour', 'playing', 'and', 'CPAP', 'times', 'apnea', 'suffering', 'months', "n't", 'it', 'about', 'I', 'got', 'night', 'Dr', 'to', 'couch', 'changer', 'that', 'started', 'talked', 'discovered', '3', 'did', 'hoping', 'difference', 'my', 'do', 'or', 'So', 'used', 'want', 'instead', 'when', 'Today', 'the', 'a', 'make', 'for', 'machine', 'been', 'life', '.', 'with', 'was', 'sleeping', 'is', 'on', 'fatigue', 'an', 'he', '2', 'children', 'outside', 'Its', 'day', 'Nuvigil', 'them', 'It', 'years', 'should', 'since', 'laying', '50', 'have', 'just', 'stop', 'every', 'so', 'of', 'energy', 'They'}
```

```
TextBlobNeg: {'past', 'everyday'}
```

```
Post Polarity: 0.05516917293233084
```

AAPProv22 - Manual positive, TextBlob positive, VADER negative

Post:- I posted 10/05 in this forum. I am still on 300-400mg daily. Taking it 5-6 years. Stopped for 4 months (insurnce company), experienced vivid nightmares where I would wake up screaming. I wouldfall asleep for a milisecond and be yanked awake with such intensity I would wake up screaming. This cycle could be repeated several times at the start of a night. It was horrible. Get sleepy, fall asleep, wham scream wake, sleepy, sleep, wham scream wake. Some nights, I sat up watching tv until my body just gave out and slept. I found it fascinating that taking provigil for daytime sleepiness improved the quality (no nightmares, scremaming) of my sleep. The nightmares were in blazing technicolor and almost seemed real. Thenak god they are gone. I have found no addictive properties to it. SideEffects jaw clench, smelly urine, have to watch combining it with caffeine, can't take it after say about noon, if I want to sleep

```
VaderPos: {'god', 'improved', 'awake', 'fascinating'}
```

```
VaderNeu: {'10/05', 'if', 'found', 'where', 'are', 'was', 'such', 'ca', 'milisecond', 'would', 'blazing', 'gone', 'Taking', 'repeated', ',', 'watch', 'body', 'in', 'for', '(', 'were', 'intensity', 'could', 'yanked', 'the', 'to', 'and', 'asleep', 'I', 'slept', 'night', 'out', '4', 'forum', 'fall', 'still', 'watching', 'daily', 'The', 'quality', 'about', 'months', 'It', 'after', 'technicolor', 'Thenak', 'Get', 'times', 'gave', 'am', 'it', '5-6', 'nightmares', 'a', 'just', 'want', 'on', 'wake', 'nights', 'be',
```

```
'cycle', 'take', 'scremaming', 'noon', 'jaw', "n't", 'properties',
'clench', 'of', 'Some', 'real', 'vivid', 'daytime', 'they', 'say',
'combining', 'have', 'experienced', ')', 'sat', 'almost', 'company',
'smelly', 'sleepiness', 'seemed', 'my', 'addictive', 'until',
'This', 'posted', 'insurnce', '.', 'at', 'wouldfall', 'with', '300-
400mg', 'SideEffects', 'provigil', 'several', 'years', 'tv',
'urine', 'up', 'that', 'wham', 'this', 'caffeine', 'start',
'taking', 'sleep'}

VaderNeg: {'sleepy', 'horrible', 'scream', 'screaming', 'no',
'Stopped'}

Scores: {'neg': 0.134, 'neu': 0.794, 'pos': 0.072, 'compound': -
0.8807}


TextBlobPos: {'fascinating', 'vivid', 'experienced', 'real'}

TextBlobNeu: {'for', 'where', '10/05', 'noon', 'ca', 'a', ',',
'nights', 'smelly', 'improved', 'Get', "n't", 'months', 'provigil',
'sat', 'my', 'seemed', 'at', 'body', 'and', 'SideEffects', 'have',
'tv', 'clench', 'found', 'daytime', 'with', 'were', '.', '4',
'quality', '300-400mg', 'in', 'wham', 'The', 'such', 'am',
'watching', 'times', 'fall', 'say', 'daily', 'watch', 'that',
'awake', ')', 'gone', 'night', 'forum', 'start', 'are', 'I',
'wouldfall', 'on', '5-6', 'this', 'be', 'cycle', 'take', 'years',
'Some', 'asleep', '(', 'up', 'god', 'several', 'yanked', 'almost',
'company', 'wake', 'was', 'insurnce', 'if', 'sleepy', 'Taking',
'gave', 'would', 'could', 'jaw', 'caffeine', 'screaming', 'want',
'posted', 'properties', 'milisecond', 'slept', 'combining', 'no',
'This', 'just', 'technicolor', 'taking', 'urine', 'after',
'nightmares', 'intensity', 'blazing', 'out', 'they', 'sleep', 'It',
'about', 'repeated', 'addictive', 'Thenak', 'to', 'still',
'sleepiness', 'the', 'Stopped', 'of', 'it', 'scream', 'scremaming',
'until'}

TextBlobNeg: {'horrible'}

Post Polarity: 0.525
```

Although many of the words classed as neutral by both methods are used by posters to indicate a health outcome, they can conceivably be used in either a positive or negative context. It was very noticeable in these examples that each method identified almost entirely sets of positive and negative word. I have highlighted the few that were the same for clarity. Future work could investigate this making further refinements to the lexicons.

In terms of identifying expressions of perceived causality, I used the same key ngram based method on both datasets. In the manual coding of P1 I was able to easily split each post into 'Pre' and 'Post' Modafinil which helped determine the base state, action, consequences for identifying causal text. This was not possible with the P3 dataset, but for future work I plan on using the tense of the POS tags to see if this can be replicated.

6.8 Discussion of P3 results

Both methods of sentiment analysis showed significant positive sentiment from the posts. As with the exploratory study, the posters from the P3 dataset were experiencing a wide range of conditions and symptoms. All the theme identification methods were effective in highlighting words and phrases that mapped directly to the themes and some of the subthemes from P1.

6.8.1 Descriptive Statistics

Although reddit appeared to be by far the largest data source, it comprises over 2.8 million subreddits, or individual communities [329] covering a wide range of subjects. These subreddits can be viewed as individual data sources, each with their own demographic and user community. Having an open access policy to its archives [186] it is a widely used data source for many researchers [330].

6.8.2 Theme Detection

Both the LDA and NMF topic modelling methods require manual tuning of multiple parameters to obtain the most accurate results. The results were interesting, but not as helpful as I had hoped they would be. The majority of the individual words returned were clearly related to the P1 themes, but despite the coherence testing suggesting that the optimum number of topics was 27, it was difficult to get a clear interpretation of the themes of the posts from models run using that parameter with any model. However, running repeated variations of the coherence test returned markedly different suggested optimal numbers of topics for the data, which suggests that coherence testing is also not a reliable indicator of the optimal number of topics.

Although topic modelling can identify the topics contained in text, the bag of words approach used by the LDA and NMF models ignores the order of the words. Therefore it does not pick up on the nuances and deeper understanding contained within them, so cannot explain how they are talked about [331]. Adding ngram features to the simpler BOW method can improve the results [332]. Pre-processing techniques designed to reduce the complexity of the models can also remove information from the text [200]. The LDA models in particular can take many hours to run, so there needs to be a balance between the time taken, the computing resources available and the depth of analysis required.

The interactive visualisations of the LDA method made it easy to see the degree of overlap between the topics. The ability to adjust the relevance metric after generating the topics allows varying the display to show words that are exclusive to the selected topic or more general ones that may appear in other topics. However, the gensim LDA models were very slow to run, especially when increasing the number of iterations over the data.

The sklearn NMF algorithm appeared to return the most interpretable groups of topic words, and from the comparative bar graphs of both sklearn algorithms returned more balanced topics. However, in terms of developing a general methodology the way that the sklearn categorised 94% of the posts into just eight topics might elicit some potentially interesting outliers.

The Top2Vec method had several advantages over the LDA or NMF methods which were very helpful in answering the research question. The results from this method were much easier to interpret as a human, although there was still a large amount of crossover of topics within each of the 367 distinctions the model had made. Although it was not immediately apparent from the lists of topic words how some of the differentiations were made, it did clearly distinguish several of the subcodes that were identified in the qualitative analysis. In also returning the posts that are most representative of each topic, it is also easily possible to go back to read the full text of the post to help get a better understanding. By default, the method returns the top 50 words for each topic it identifies, which greatly aided the interpretation. Overall, the Top2Vec vector based returned the most interpretable topics, representing all the main P1 themes and many of the sub themes.

Corpus linguistics takes a different approach to identifying the distinct key words that represent the corpus being analysed. By comparing the frequencies of words between the target corpus and a reference corpus, it highlights the words and terms that appear most regularly in the target corpus. For keywords these are generally nouns and adjectives that represent the entities described in the data, as other parts of speech tend to be similar across all texts [47]. The keyterms were the easiest to map to the existing themes, while the key ngrams and normal ngrams were the most useful for indicating perceived causation or beliefs.

Overall, all the methods generated words or phrases that were relevant to the research question, but more work needs to be done to optimise the process. It would be interesting to rerun the LDA and NMF models using 367 as the parameter for the number of topics, and also to extend the output to generate 50 words per topic. Both the keyword and keyterm extractions gave results that were highly relevant to the specific question about Modafinil experiences. I only tried to map the top 100 of each but extending this process to the full 1000 of each might also show up some interesting findings. The keyword extraction was particularly useful for identifying other drug names from the posts. The finding that second largest class of drug names mentioned were antidepressants is in line with the suggestion from the exploratory study that Modafinil can be more effective for some people that antidepressants. Acquisition of Modafinil was another major theme from the exploratory study, and virtually all the methods identified this. Although clinicians can prescribe it freely in the US [4], posters reported how it was the insurance companies that were refusing to cover payment for it. Within the UK, current guidelines require a secondary care diagnosis of narcolepsy to be made before it can be prescribed, limiting its use within primary care [333]. Many of the key words that could be ascribed to either effect or outcome require looking at in context such as the adjective collocation diagram (Figure 6-14) or word sketch for 'feel' (Figure 6-15) as they could be describing features that were pre or post intervention. Using the POS tag of the relevant tokens could also help identify causal sequences.

My mapping of the output topics to a single category was itself subjective, and no doubt other interpretations of the categorisation process could be made. In addition to standardising the output from the models to give a fairer comparison, future work could involve a second reviewer of this process.

6.8.3 Sentiment Analysis

All the methods of assessing the sentiment on both datasets showed the majority of posts expressing positive sentiment. In the exploratory P1 study, I calculated the overall agreement at post level of the TextBlob method compared to manual evaluation. Depending on the range used for neutral, the overall agreement was either 59% or 64%. In this study I added the VADER method to the comparison, and also calculated the post level agreement for each category of sentiment; positive, neutral or negative (Table 6-23).

This showed that TextBlob was the most successful at matching positive posts (84%) while VADER was better at identifying negative posts (77%), although VADER was more consistent.

Both TextBlob and VADER are lexicon-based methods, and as such are constrained by the domains upon which they are based. The TextBlob lexicon focuses on adjectives found in customer product reviews. Each adjective has been manually tagged with values for polarity and subjectivity [323]. The VADER lexicon is based on more general language. The examples shown in section 6.7 where the two methods both returned the opposite sentiment to the manual evaluation, show how differently they classify each post.

Although the results showed similar overall percentage figures for the positive sentiment, my comparison of the level of agreement for the individual P1 posts indicated that the methods could be improved. The ability to be able to easily modify the Vader lexicon showed an improvement in agreement levels with only a few changes. Modifying the VADER lexicon is straightforward, and any changes can be easily documented in the code. Further work to finetune the lexicon to an outcomes-based health domain would improve the agreement further.

Effectiveness was evaluated at the post level. Both NLP methods assume that every post can be identified as either positive or negative, but do not allow for the 'mixed' category that was manually evaluated in the exploratory study. This category was defined in Table 3-2 as including posts where both positive and negative effects were reported; and it was unclear as to which sentiment prevailed. Improvements for future work could include evaluation at the sentence level [80], especially for those posts that the libraries classified as neutral, together with greater refinement of the lexicons in terms of the words and phrases that posters use to describe their health state.

The plots of word count and sentiment value demonstrate how VADER seems to give the most extreme ratings, both positive and negative to the posts with the longest word counts, whereas the TextBlob method showed a more normal type of distribution. Other studies have suggested that positive social media comments tend to be shorter in length but less specific, while negative ones are longer but less frequent [65,334]. Future work could explore this further.

Previous studies have commented on how lexicon based tools trained on general language do not perform as well on health related text [145]. Although lexicon-based sentiment analysis can give an accurate assessment of text that contains words that express a strong positive or negative sentiment, posts that do not contain many of these predefined words are harder to evaluate. One of the features of the informal nature of SGOPE is that the writers assume that readers can readily infer the affective reaction they are describing. Descriptive phrases such as "I could go back to work" or "it gave me a headache" infer the effect of the event but would be viewed as neutral statements by most sentiment analysis models.

6.9 General method development

The overall methodology of the main study was designed so that it can be applied to other research questions using unstructured data. The principles of the methods I have used in this study have shown that they can be used inductively on large volumes of unstructured text to extract the themes, sentiment, and expressions of perceived causality.

Unsupervised methods align more to the inductive approach of qualitative studies and are shown to be effective for exploring SGOPE data. Both topic modelling and the extraction of keywords, keyterms and key ngrams identify what is being spoken about, but not how the word or phrase is used in context. Combining NLP with corpus linguistics draws on the strengths of both disciplines [200,282] and allows the researcher to identify the content that is most relevant to the research question [335].

6.10 Strengths and Limitations of P3

Using unsupervised methods allows an inductive approach to the analysis. The use of multiple methods to identify both themes and effectiveness, and the comparison of the findings with each other in addition to those from the exploratory dataset is a strength of this study. The themes generated from the exploratory study were used to guide and evaluate the analysis process. The methods and parameters were explained, addressing one of the criticisms from the review.

The main limitations of this study relate to the original dataset as supplied by Treato. Although it included many more data sources than it would have been feasible for me to identify and collect from, the general data quality was inferior to that of the exploratory study. Posts were included if they included the terms Modafinil, Armodafinil, Provigil or Nuvigil anywhere in the text, but additional work to remove less relevant posts would have improved the quality. In addition to removing any irrelevant posts, further step could also be to extract only first-person experiences.

Dealing with threaded conversations is more complex than the review type posts of the exploratory study. Analysing the responses to a thread as a standalone document can lose the context provided by the thread title but including the title with each post can skew frequency-based analysis by the over representation of the title words. It is also possible in replying to a specific thread about Modafinil that relevant posts that did not explicitly include one of the terms could have been excluded from the collection process. Even though exact duplicate posts were removed, the analysis process also showed that some partial duplicates remained, probably because of the poster quoting a previous post in their response.

Results from topic modelling should not be seen as deterministic. They are dependent on the choices made by the researcher as to the level and methods of pre-processing, the parameters set and the algorithms used [295]. Additional human evaluation of the topic classification would increase the validity of the study [280]. The late addition of Top2Vec to the methods made comparisons between the methods more complicated. Previous evaluation papers have compared the default ten-word output to assess their performance, but the 50 word standard output per topic from Top2Vec meant that it was an unfair comparison. I also only mapped the top 100 of the corpus specific keywords and keyterms to the P1 themes. Doing this for the top 1000 may have shown different results.

Although I only made a few changes to the VADER lexicon, they did noticeably improve the evaluation process. A deeper examination and added refinement of the words it contains would be advantageous. It would be interesting to develop a lexicon that reflected health outcomes.

Chapter 7 Discussion and conclusion

7.1 Outline

This chapter recaps the purpose of the project, brings together and discusses the findings from the previous chapters. It compares the findings with other studies and considers if, how or why SGOPE data could be used within healthcare to augment existing knowledge.

7.2 Recap of project aim

The aim of this study was to explore the use of SGOPE data to contribute to health research, particularly in terms of understanding the patient perspective of effectiveness and the outcomes considered important. Modafinil was chosen as a case study as it is already licensed within the NHS for narcolepsy but the existing conventional RCT based evidence for its wider use is inconclusive. The specific research question asked about the potential of analysis of SGOPE data relating to Modafinil use to become part of the evidence for its effectiveness in practice.

The overall approach to the design

- Aimed to emulate the depth of qualitative research but at scale and more objectively.
- Recognised the value of the posters views as to what works and matters to them in terms of healthcare.
- Developed a general method that can be adapted to other health research questions.

Chapter 3 included the exploratory study, chapter 4 reviewed the purposes and methods that have been used to date on this data source, while chapters 5 and 6 described the main study.

7.3 Summary of findings

Analysis of this data, especially when undertaken using a causal dispositionalist framework, allows for a genuinely patient centred approach capable of augmenting existing evidence generation methods. Combining the findings with a dispositionalist philosophy of knowledge generation allows for plurality of methods in evidence creation thereby addressing some of the current limitations of EBM.

Beginning with a comparison of qualitative and NLP methods on a small sample of SGOPE data relating to experiences of Modafinil, the exploratory P1 study showed how both approaches could demonstrate similar results. Manual and NLP sentiment analysis showed that in contrast to current RCT based evidence, most posters with a wide range of conditions found Modafinil effective. Expressions of perceived causality and effectiveness were identified by both methods demonstrating the potential to augment existing knowledge. For those finding it beneficial, the project showed how the use of Modafinil impacted on both their health and on the context of their wider life. For these posters it was the ability to re-engage with 'normal life' rather than a change in any clinical markers that they valued. Both methods successfully identified the entities and topics contained in the posts.

The umbrella scoping review highlighted the very recent growth of interest into the use of SGOPE as a data source, the consensus being that method development is still at a very early stage with many issues still to be overcome. Mental health, ADR detection, and infectious disease have been the health topics of most focus. Most studies have used data from a single site, usually Twitter. Many studies still used qualitative or mixed methods, but supervised machine learning was the most widely used NLP method. Frequently discussed issues included the limitations of a supervised approach, the lack of explanation of the methods used and the need for ethical debate.

Using an unsupervised approach to allow the data to tell its own story, the main P3 study combined NLP with corpus linguistics discover the topics that posters were writing about, how effective they found Modafinil, and ways of identifying perceived causation. Different methods to identify themes and assess effectiveness were compared.

Findings from both P1 and P3 studies were broadly similar. Although a range of positive and negative experiences were reported, both studies showed strong findings of posters finding Modafinil effective for their symptoms. Effectiveness was assessed as being significantly positive (55%-72%) with an average of 64% in both studies and by all methods. Similar themes were identified by both qualitative and computational analysis, and the strengths and limitations of the methods were explored. Difficulties in being prescribed or acquiring Modafinil were significant themes from both studies. All the topic modelling methods returned topics containing words that clearly related to and could be mapped to the themes and subcodes from the exploratory study. Linguistic analysis identified expressions of causal belief, which can be either be used to supplement existing knowledge or to guide the development of new research questions or study designs.

This study contributes to the development of methods that can be used to analyse large quantities of unstructured data and demonstrates how SGOPE data can be used to augment existing health research. However, it only scratches the surface of what will be possible in terms of unstructured text analysis, and there are many areas that could be refined and improved. I cannot say that I have solved many of the potential issues of NLP analysis that were identified in the exploratory study. Spellings and ambiguity are still a real issue, but there do not seem to be any current methods that can reliably be used to change only those instances that are known to be wrong, without inadvertently changing words that are actually correct. Identifying and correctly interpreting negation is another ongoing research area in NLP [336,337].

7.4 Comparison with Cochrane reviews

These findings of overall effectiveness contrast strongly with the existing current RCT and systematic review evidence used to determine treatment pathway options for clinicians [25]. A search of Cochrane systematic reviews with the term 'Modafinil' in the title, abstract or keyword returned 16 reviews as of May 2021 (Table 7-1). Rather than searching for every review or RCT of Modafinil, I used the Cochrane reviews as a comparison as they subject the individual trials to a critical appraisal, are recognized as providing a high-quality level of both assessment and evidence synthesis and are also used to contribute to clinical guidelines [338]. Some of trials considered multiple interventions for a condition, whereas others looked specifically at Modafinil. The authors of each review present their own conclusion of the included studies in terms of an interpretation of the quality of the evidence and suggestions as to how to address remaining uncertainties [339]. From their conclusions I extracted the comments that were relevant to this project.

Table 7-1: Cochrane reviews with Modafinil in title, abstract or as keyword - May 2021

	Title	Extract from Author's conclusion	Ref
1	Modafinil for people with schizophrenia or related disorders	Due to methodological issues, low sample size, and short duration of the clinical trials as well as high risk of bias for outcome reporting, most of the evidence available for this review is of very low or low quality. For results where quality is low or very low, we are uncertain or very uncertain if the effect estimates are true effects, limiting our conclusions. Specifically, we found that modafinil is no better or worse than placebo at preventing worsening of psychosis; however, we are uncertain about this result. We have more confidence that participants receiving modafinil are no more likely to leave a trial early than participants receiving placebo. However, we are very uncertain about the remaining equivocal results between modafinil and placebo for outcomes such as improvement in global state or cognitive function, incidence of adverse events, and changes in quality of life. More high-quality data are needed before firm conclusions regarding the effects of modafinil for people with schizophrenia or related disorders can be made.	Ortiz-Ordain et al 2019 [340]
2	Amphetamines for Attention Deficit Hyperactivity Disorder (ADHD) in adults.	The short study length and the restrictive inclusion criteria limit the external validity of these findings…. possibility that the results of the included studies were biased was high...	Castells 2018 [341]
3	Pharmacological interventions for apathy in Alzheimer's disease	There was insufficient evidence from one very small study of modafinil to determine the effect of modafinil on apathy assessed with the FrSBe-apathy subscale: MD 0.27, 95% CI -3.51 to 4.05, n = 22, 1 study, low quality of evidence.	Ruthirakuhan 2018 [342]
4	Treatment of fatigue in amyotrophic lateral sclerosis/motor neuron disease	We found very low-quality evidence suggesting possible improvements in fatigue for modafinil treatment versus placebo (MD -11.00, 95% CI -23.08 to 1.08). We cannot be certain about the effects of any of the interventions studied because of imprecision (small numbers of participants, wide CI), and possible study limitations. It is impossible to draw firm conclusions about the effectiveness of interventions to improve fatigue for people with ALS/MND as there are few randomised studies, and the quality of available evidence is very low.	Gibbons 2018 [343]
5	Psychostimulant drugs for cocaine dependence.	Mixed results/deserves further investigation	Castells et al. 2016 [15]
6	Interventions for the management of fatigue in adults with a primary brain tumour.	There was insufficient evidence to draw reliable and generalisable conclusions regarding potential effectiveness or harm of any pharmacological or non-pharmacological treatments for fatigue in people with PBT. More research is needed on how best to treat people with brain tumours with high fatigue.	Day 2016 [344]

	Title	Extract from Author’s conclusion	Ref
7	Pharmacotherapy for chronic cognitive impairment in traumatic brain injury.	There is insufficient evidence to determine whether pharmacological treatment is effective in chronic cognitive impairment in TBI. Whilst there is a positive finding for rivastigmine on one primary measure, all other primary measures were not better than placebo. The positive findings for (−) -OSU6162 are interpreted cautiously as the study was small (n = 6). For modafinil and atomoxetine no positive effects were found. All four drugs appear to be relatively well tolerated, although evidence is sparse.	Dougall et al. 2015 [14]
8	Interventions for fatigue in Parkinson's disease.	On current evidence no clear recommendation / need to develop self-reported fatigue questionnaires	(Elbers et al. 2015) [17]
9	Pharmacological treatments for fatigue associated with palliative care.	Based on limited evidence, we cannot recommend a specific drug for the treatment of fatigue in palliative care patients. Fatigue research in palliative care seems to focus on modafinil and methylphenidate, which may be beneficial for the treatment of fatigue associated with palliative care although further research about their efficacy is needed. Dexamethasone, methylprednisolone, acetylsalicylic acid, armodafinil, amantadine and L-carnitine should be further examined. Consensus is needed regarding fatigue outcome parameters for clinical trials. Further research needed.	(Mücke et al. 2015)
10	Treatment for postpolio syndrome.	Due to insufficient good-quality data and lack of randomised studies, it was impossible to draw definite conclusions about the effectiveness of interventions for PPS.	Koopman 2015 [19]
11	Interventions for preventing and ameliorating cognitive deficits in adults treated with cranial irradiation.	The remaining studies did not have a sufficient number of participants to provide reliable results. The drugs used had few side effects (adverse events), although these were not reported well. Recruitment and retention of trial participants for these medical drug studies is difficult.	Day 2014 [16]
12	Pharmacological interventions for sleepiness and sleep disturbances caused by shift work.	Both modafinil and armodafinil increase alertness and reduce sleepiness to some extent in employees who suffer from shift work sleep disorder, but they are associated with adverse events. Caffeine plus naps reduces sleepiness during the night shift, but the quality of evidence is low. Based on one low quality trial, hypnotics did not improve sleep length and quality after a night shift. We need more and better-quality trials on the beneficial and adverse effects and costs of all pharmacological agents that induce sleep or promote alertness in shift workers both with and without a diagnosis of shift work sleep disorder. We also need systematic reviews of their adverse effects	Liira 2014 [20]

	Title	Extract from Author's conclusion	Ref
13	Efficacy of psychostimulant drugs for amphetamine abuse or dependence.	Inconclusive - numbers of included studies and participants are limited and information on relevant outcomes, such as efficacy according to the severity of dependence or craving, is still missing.	Pérez-Mañá et al. 2013 [345]
	Treatment for postpolio syndrome.	Insufficient data/lack of trials/impossible to draw conclusions...	Koopman 2011 [346]
14	Caffeine for the prevention of injuries and errors in shift workers.	We need more and better-quality trials on the beneficial and adverse effects and costs of all pharmacological agents that induce sleep or promote alertness in shift workers.	Ker 2010 [347]
15	Psychostimulants for depression.	There is some evidence that in the short-term, PS reduce symptoms of depression. Whilst this reduction is statistically significant, the clinical significance is less clear. Larger high-quality trials with longer follow-up and evaluation of tolerance and dependence are needed to test the robustness of these findings and, furthermore, to explore which PS may be more beneficial and in which clinical situations they are optimal.	Candy 2008 [348]
16	Psychostimulants for hypersomnia (excessive daytime sleepiness) in myotonic dystrophy.	There is low quality evidence from two small trials that psychostimulants do not significantly improve the maintenance of wakefulness test in myotonic dystrophy. There is low quality evidence from four studies that modafinil significantly improves the Epworth Sleepiness Scale. More randomized trials are needed to evaluate the efficacy and safety of psychostimulants.	Annane et al. 2006 [349]

As clinical guidelines are largely based on the findings from systematic reviews [23–25], it is understandable that given these conclusions that NICE concludes that there is not enough evidence to support the more widespread use of Modafinil in clinical practice [350].

These Cochrane reviews also demonstrate the lack of 'information gain' that this RCT based research has achieved [117]. Systematic reviews show how trials report either the effects of a single dose or a regular daily dose for a limited time [10,116,351,352], whereas my findings include much greater variety of usage patterns. Similarly, my results also illustrate how some posters have varied dosage patterns and amounts to find the optimal dosage regime for them, with some finding that lower doses than those usually prescribed were more effective. The large number of posts describing its use for extended periods also give insights into its long term usage, which has yet to be evaluated in a trial [21]. Both my studies also demonstrated the existence of a possible tolerance effect but included the suggestion that taking occasional breaks or taking as required appeared to be a viable method of retaining effectiveness over time. Identified side effects generally reflected those already known [209], however the retrospective nature of the posts enabled the discovery of the temporary nature of some common side effects, a factor that will not be reflected in single dose trials.

One of the possible reasons for the inconclusive trial evidence to date is the heterogeneity of effect that can occur within trials [353]. Trials generally exclude participants with multiple comorbidities as these may act as confounders when measuring effectiveness [354] whereas many of the posters have two or more co-existing conditions, and may use combinations of interventions, or react to a single intervention in different ways. Another factor may be due to the wide range of conditions where symptoms of fatigue or cognitive dysfunction may occur.

The inconsistency in findings aligns with the theory that the eligibility criteria used to select trial participants can produce findings that are in contrast to those that occur in real world settings and populations [136], thus supporting the view that we need to include other forms of knowledge in the evidence hierarchy [121,133,355].

7.5 Comparison with other health related studies using SGOPE

Delivering effective and resource efficient healthcare requires finding ways to maximize the knowledge of both what works and what patients consider effective treatments and outcomes. The apparent contradictions also has implications on both on patient care and the efficiency of healthcare provision, either through the patient not receiving an intervention that may be effective, or by receiving one that is ineffective [354,356].

Using patient generated data can also help to guide research to addressing the outcomes that patients deem important, as in addition to the disconnect between research led and patient priorities, other studies demonstrate how clinicians and patients can view aspects of the same intervention very differently [357]. We already know from previously mentioned high profile examples [29,31,358] that patients have much to contribute towards knowledge generation in healthcare and that quality of life can be more important to them than symptoms [359].

SGOPE has already been shown to be an important source of information as to the outcomes and QoL factors that posters value [360]. A comparison of topic modelling to the QoL questions in self-administered QoL standardised questionnaires such as EORTC QLQ-C30 and EORTC QLQ-BR23 for breast cancer patients has also been demonstrated to match

22/23 from a cancer specific forum, and 20/23 from a Facebook corpus. Topic modelling also identified a further 5 topics from the data that were not currently evaluated in the questionnaires [360]. Word level sentiment analysis on Twitter data has been shown to correlate well to Gallup survey measures of wellbeing across the US [361].

Topic modelling can be used both retrospectively as in this study or as an aid to clinical decision making. A recent study used the LDA method to look at MSK referral letters to help with triaging patients onto the most appropriate clinical pathway. Using clinicians to evaluate the returned topics, they found that it was an effective method of reducing bottlenecks that build up in the triage process [280].

Studies have begun to look at the lexical and grammatical features of causal statements in free text [198] and some work has been done using NLP to identify pharmacological adverse events from social media [73,362,363] suggesting that negative effectiveness can be shown from this type of data.

An alternative approach to identifying causal text from Twitter data was tried by Doan et al. They used a rule-based approach, based on six defined patterns that used trigger verbs and nouns, but found that the number of sentences that matched the rules was very small, ranging from 0.8% to 1.8% [364].

SGOPE has already been shown to be filling an unmet need for health information and related emotional support for patients [54,137,145,365]. The review showed how the main health areas that SGOPE data is being used for are mental health, adverse event detection, and infectious disease tracking [187,227,228]. Despite the fears about the impact of 'Dr Google' and that social media is full of misinformation [137], a recent study of 5000 US participants given a list of symptoms found that internet searches were associated with modest but significant improvements in diagnosis [366].

It can also be used to understand more about patient behaviours. It is recognised that outcomes are determined as much by patient behaviour as by interventions [367]. Adherence or non-compliance to prescribed medications is recognized as a problem with significant economic costs [89]. Although an intervention may have been shown by high quality research to be effective, it only results in a good outcome if the patient adheres to the prescribed course of action or dosage, which they may not if it is having a negative impact on some area of their life [91]. Understanding more of why patients behave the way they do from SGOPE can identify opportunities to improve clinician-patient communications and explore alternative interventions [368]. Where posters include discussion of what other medications they have tried or moved onto, SGOPE may also generate knowledge to improve adherence rates. A recent study explored the potential for using matching techniques to match drug names with 'stopped taking' and 'made me' text patterns to identify medication non adherence in Twitter data, concluding that posters were expressing not only the fact that they were not complying, but including their reasons for not doing so [90].

The Covid pandemic has also led to some interesting ethical questions as to whether the principle of not acting until the definitive trial results of a RCT are proved can be justified, either on a scientific or moral ground, while people are being harmed by a condition [369], especially when large volumes of observational data are becoming widely available [133].

7.6 Strengths and Limitations

Taking a staged approach to the project allowed for comparison between methods and datasets, with the findings from the exploratory study acting as a benchmark for the main study. The P1 study compared qualitative and NLP analysis. The P3 study compared various methods both with each other and to the P1 results. The methods and parameters used were explained. Combining NLP with linguistics draws and brings together the strengths from both fields enabling an analysis that tries to go beyond NLP to NLU [200,282].

The use of multiple data sites increased the representativeness of the sample and reduced the potential for demographic bias or emotional contagion. Although Twitter data is available in massive quantities and is extremely useful for certain types of health research questions, I felt it was too limited in both character length and demographics to be able to answer the research question effectively. As with any form of research, SGOPE data can be subject to forms of bias. Data may be subject to health user bias; in that it may be those that are most interested in their heath that contribute. Even though SM usage is so widespread, those writing the posts are a self-selecting subgroup of the overall population of users. Each platform will have different ratios, but one estimate suggests that it is only 10% of users that create posts, while 90% will observe without submitting their own comment [370].

The umbrella scoping review was the first of its type since 2018 and summarised the current state of the art of the field. Including discussion of the ethical issues relating to

SGOPE data use and explaining the analysis methods addresses some of the limitations of existing studies as identified by the review.

In terms of SGOPE data itself, public domain datasets such as that used in the P3 study are likely to be the messiest and noisiest. In retrospect the P1 user review data was generally more relevant and specific to Modafinil experiences than the larger dataset. Posts from condition specific forums tend to be longer and more informative than those from general sites such as Facebook [360]. Data cleaning is crucial to this type of analysis, and although cleaning at this volume of data cannot guarantee that the dataset is 100% clean or reliable data from moderated forums might be easier to clean and more focused on the research topic. However, it is possible that any moderation of the post could lose some of the detail. The structure of the P1 data also included some specific data fields regarding features such as dosage detail, the duration of taking Modafinil or other demographic information. Future work could also include extracting this from the P3 data.

It is impossible to guarantee that the information in the post is accurate or complete. Diagnosis of any condition may have been done by the poster rather than by a clinician [140]. There is often a perception that data from social media is particularly unreliable [371], but in reality no form of communication can be guaranteed to be a transparent window into the mind of the originator. Whether in speech or written communication all authors can filter or frame their output to suit either their own objectives or the perceived expectations of others [47].

7.7 Implications for clinical practice

The current position of NICE is to say that there is no evidence to support the use of Modafinil for patients with cognitive or fatigue symptoms. While being technically accurate, a better approach may be to say there is no 'good' evidence to either support or rule out its use for particular patients and that we should look for ways to augment current knowledge. Any further indication licensing of Modafinil in the UK is unlikely due to lack of evidence as to safety and efficacy [372]. RCTs are extremely expensive, and with little incentive to undertake new trials, pharmaceutical companies are unlikely to invest further in a drug first developed in 1998 and whose main patents expired in 2005. At present the default approach is to not use the intervention unless it is proven to be effective by trials, but I would argue that it would be beneficial to both patients and service providers to explore alternative data sources and study designs that could augment the existing knowledge base. In the specific case of Modafinil, an alternative method of evaluation for future research that may work for such a case as this where the drug is already licensed for use within the NHS, and is judged as having minimal side effects and a low abuse potential [3] may be one based on a 'N of 1' design in conjunction with a symptom tracking app to get a real world perspective of its effectiveness for symptoms of fatigue and cognitive dysfunction other than narcolepsy.

- Allow people specifically requesting it to become part of a large-scale study to evaluate efficacy etc.
- No control group would be needed - just compare findings with standard care.
- If carried out within NHS, checks for contraindications can be carried out before patients started on it.
- Potential to build a body of high-quality real-world evidence across a range of conditions quite quickly.
- Potential for positive patient empowerment/satisfaction if patients feel included in decision making.

Future Research

There are many other avenues for further work in this area. Suggestions for extensions to this project have already been mentioned. There are many other health topics where SGOPE data could offer additional insights to existing knowledge. We know that many of the most active users of online spaces are those with long term conditions or rare diseases [54,137], but as the example of Long Covid showed, health conversations and topics are constantly fluid. Future work will include using these methods on SGOPE data relating to other topics. I also want to apply them on the unstructured free text in clinical notes and those from calls to 111. The methods themselves can be developed further. More open source datasets such as i2b2 [373] and MIMIC [374] are becoming available along with health specific tools such as Med7 [375] and CogStack [376] that will help improve them.

7.8 Conclusion

This study contributes to the development of methods that can be used to analyse large quantities of unstructured health text data and demonstrates how SGOPE can be used to augment existing health research. In contrast to the existing inconclusive systematic review evidence for Modafinil for anything other than narcolepsy, both studies found that most posters find Modafinil to be effective in dealing with fatigue and cognitive symptoms across a wider range of conditions.

Although the two methods are very different, I have demonstrated how computational methods can extract the same main topic areas as qualitative analysis. It also demonstrated how SGOPE shows potential for the identification of perceived causation and evaluation of the effectiveness of Modafinil.

Although much work is needed to refine the techniques and address the challenges identified, this project demonstrate how SGOPE can help address some of the identified issues with a research driven agenda [117] and complement RCTs. The inconclusive evidence base and failure to address the issues that really matter to patients occurs across many health conditions [356] and needs to be addressed if healthcare is to become truly patient centred. Developing reliable methods of SGOPE analysis will allow researchers to assess the content and develop evidence that can be used by policy makers in a timely manner.

References

1. Joint Formulary Committee. Modafinil [Internet]. British National Formulary (online). NICE; 2017 [cited 2017 Jul 4]. Available from: https://bnf.nice.org.uk/drug/modafinil.html

2. Medicines, Healthcare products Regulatory Agency. Modafinil: restricted use recommended [Internet]. GOVUK. GOV.UK; 2014 [cited 2019 Jul 7]. Available from: https://www.gov.uk/drug-safety-update/modafinil-restricted-use-recommended

3. www.medicines.org.uk. Modafinil Provigil 200mg tablets - Summary of Product Characteristics (SPC) [Internet]. medicines.org.uk. electronic Medicines Compendium - eMC; 2016 [cited 2016 Nov 21]. Available from: http://www.webcitation.org/query?url=https%3A%2F%2Fwww.medicines.org.uk%2Femc%2Fproduct%2F4320%2Fsmpc&date=2019-05-29

4. Frost J, Okun S, Vaughan T, Heywood J, Wicks P. Patient-reported outcomes as a source of evidence in off-label prescribing: analysis of data from PatientsLikeMe. J Med Internet Res [Internet] 2011 Jan 21;13(1):e6. PMID:21252034

5. Gajewski M, Weinhouse G. The Use of Modafinil in the Intensive Care Unit. J Intensive Care Med [Internet] 2016 Feb;31(2):142–145. PMID:25716122

6. Castells X, Ramos-Quiroga JA, Bosch R, Nogueira M, Casas M. Amphetamines for Attention Deficit Hyperactivity Disorder (ADHD) in adults. Cochrane Database Syst Rev [Internet] 2011 Jun 15;(6):CD007813. PMID:21678370

7. Möller F, Poettgen J, Broemel F, Neuhaus A, Daumer M, Heesen C. HAGIL (Hamburg Vigil Study): a randomized placebo-controlled double-blind study with modafinil for treatment of fatigue in patients with multiple sclerosis. Mult Scler [Internet] 2011 Aug;17(8):1002–1009. PMID:21561959

8. Stankoff B, Waubant E, Confavreux C, Edan G, Debouverie M, Rumbach L, Moreau T, Pelletier J, Lubetzki C, Clanet M, French Modafinil Study Group. Modafinil for fatigue in MS: a randomized placebo-controlled double-blind study. Neurology [Internet] 2005 Apr 12;64(7):1139–1143. PMID:15824337

9. Tuken M, Kiremit MC, Serefoglu EC. On-demand Modafinil Improves Ejaculation Time and Patient-reported Outcomes in Men With Lifelong Premature Ejaculation. Urology [Internet] 2016 Aug;94:139–142. PMID:27151339

10. Goss AJ, Kaser M, Costafreda SG, Sahakian BJ, Fu CH, Goss AJ, Kaser M, Costafreda SG, Sahakian BJ, Fu CHY. Modafinil augmentation therapy in unipolar and bipolar depression: a systematic review and meta-analysis of randomized controlled trials. J Clin Psychiatry [Internet] Physicians Postgraduate Press, Inc.; 2013 Nov 15;74(11):1101–1107. PMID:24330897

11. Ondo WG, Fayle R, Atassi F, Jankovic J. Modafinil for daytime somnolence in Parkinson's disease: double blind, placebo controlled parallel trial. J Neurol Neurosurg Psychiatry [Internet] BMJ Publishing Group Ltd; 2005 Dec 1 [cited 2017 Jan 18];76(12):1636–1639. PMID:16291885

12. Cooper MR, Bird HM, Steinberg M, Cooper MR, Bird HM, Steinberg M. Efficacy and safety of modafinil in the treatment of cancer-related fatigue. Ann Pharmacother [Internet] 2009;43(4):721–725. PMID:19318599

13. Spathis A, Fife K, Blackhall F, Dutton S, Bahadori R, Wharton R, O'Brien M, Stone P, Benepal T, Bates N, Wee B. Modafinil for the treatment of fatigue in lung cancer: results of a placebo-controlled, double-blind, randomized trial. J Clin Oncol [Internet] Cochrane Central Register of Controlled Trials; 2014 Jun 20;32(18):1882–1888. PMID:24778393

14. Dougall D, Poole N, Agrawal N. Pharmacotherapy for chronic cognitive impairment in traumatic brain injury. Cochrane Database Syst Rev [Internet] 2015; PMID:26624881

15. Castells X, Cunill R, Pérez-Mañá C, Vidal X, Capellà D. Psychostimulant drugs for cocaine dependence. Cochrane Database Syst Rev [Internet] 2016 Sep 27;9:CD007380. PMID:27670244

16. Day J, Zienius K, Gehring K, Grosshans D, Taphoorn M, Grant R, Li J, Brown PD. Interventions for preventing and ameliorating cognitive deficits in adults treated with cranial irradiation. Cochrane Database Syst Rev [Internet] 2014 Dec 18;(12):CD011335. PMID:25519950

17. Elbers RG, Verhoef J, van Wegen EEH, Berendse HW, Kwakkel G. Interventions for fatigue in Parkinson's disease. Cochrane Database of Systematic Reviews [Internet] John Wiley & Sons, Ltd; 2015. p. 0:CD01. PMID:26447539

18. Peuckmann V, Elsner F, Krumm N, Trottenberg P, Radbruch L. Pharmacological treatments for fatigue associated with palliative care. Cochrane Database Syst Rev [Internet] 2010 Nov 10;(11):CD006788. PMID:21069692

19. Koopman FS, Beelen A, Gilhus NE, de Visser M, Nollet F. Treatment for postpolio syndrome. Cochrane Database Syst Rev [Internet] 2015 May 18;(5):CD007818. PMID:25984923

20. Liira J, Verbeek JH, Costa G, Driscoll TR, Sallinen M, Isotalo LK, Ruotsalainen JH. Pharmacological interventions for sleepiness and sleep disturbances caused by shift work. Cochrane Database Syst Rev [Internet] 2014 Aug 12;(8):CD009776. PMID:25113164

21. emc. Modafinil 100mg tablets - Summary of Product Characteristics (SPC) [Internet]. emc. 2017 [cited 2017 Aug 14]. Available from: http://www.medicines.org.uk/emc/medicine/27040

22. Lyon J. Chess Study Revives Debate Over Cognition-Enhancing Drugs. JAMA [Internet] 2017 Aug 16; PMID:28813563

23. Kones R, Rumana U, Merino J. Exclusion of 'nonRCT evidence' in guidelines for chronic diseases - is it always appropriate? The Look AHEAD study. Curr Med Res Opin [Internet] 2014 Oct;30(10):2009–2019. PMID:24841173

24. Ogilvie D, Adams J, Bauman A, Gregg EW, Panter J, Siegel KR, Wareham NJ, White M. Using natural experimental studies to guide public health action: turning the evidence-based medicine paradigm on its head. J Epidemiol Community Health [Internet] 2020 Feb;74(2):203–208. PMID:31744848

25. Schlegl E, Ducournau P, Ruof J. Different Weights of the Evidence-Based Medicine Triad in Regulatory, Health Technology Assessment, and Clinical Decision Making. Pharmaceut Med [Internet] 2017 Jul 11;31(4):213–216. PMID:28824273

26. Frost JH. The case for using social media to aggregate patient experiences with off-label prescriptions. Expert Rev Pharmacoecon Outcomes Res [Internet] Taylor & Francis; 2011 Aug;11(4):371–373. PMID:21831014

27. Lamas M. Essure Side Effects [Internet]. drugwatch. 2020 [cited 2020 Oct 16]. Available from: https://www.drugwatch.com/essure/side-effects/

28. Vercillo K. How Social Media Is Changing the Way We Think About Our Health [Internet]. Blonde & Balanced. 2016 [cited 2019 Dec 17]. Available from: https://www.blondeandbalanced.com/how-social-media-is-changing-the-way-we-think-about-our-health/

29. Cumberledge J. First Do No Harm– The report of the Independent Medicines and Medical Devices Safety Review [Internet]. The Independent Medicine & Medical Devices Safety Review. 2020 [cited 2020 Oct 7]. Available from: https://www.immdsreview.org.uk/Report.html

30. Williamson J. The COVID-19 evidence base | EBM+ [Internet]. EBM+. 2020 [cited 2020 Jul 25]. Available from: http://ebmplus.org/the-covid-19-evidence-base/

31. Callard F, Perego DE. How and Why Patients Made Long Covid. Soc Sci Med [Internet] 2020 Oct 7;113426. [doi: 10.1016/j.socscimed.2020.113426]

32. Pawelek DAA. Prevalence of ongoing symptoms following coronavirus (COVID-19) infection in the UK - Office for National Statistics [Internet]. Office for National Statistics; 2021 [cited 2021 Jun 5]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/prevalenceofongoingsymptomsfollowingcoronaviruscovid19infectionintheuk/4june2021

33. Statista. Social media usage worldwide. Statista ; 2019. Report No.: 12393.

34. Kemp S. The Global State of Digital in October 2019 [Internet]. We Are Social. 2019 [cited 2020 Mar 21]. Available from: https://wearesocial.com/blog/2019/10/the-global-state-of-digital-in-october-2019

35. Gohil S, Vuik S, Darzi A. Sentiment Analysis of Health Care Tweets: Review of the Methods Used. JMIR Public Health Surveill [Internet] publichealth.jmir.org; 2018 Apr 23;4(2):e43. PMID:29685871

36. Lavertu A, Altman RB. RedMed: Extending drug lexicons for social media applications [Internet]. biorvix.org. Stanford University; 2019. [doi: 10.1101/663625]

37. Kemp S. Digital 2019: The United Kingdom — DataReportal – Global Digital Insights [Internet]. DataReportal – Global Digital Insights. DataReportal – Global Digital Insights; 2019 [cited 2019 Dec 31]. Available from: https://datareportal.com/reports/digital-2019-united-kingdom

38. Anderson M, Perrin A. Tech Adoption Climbs Among Older Adults [Internet]. Pew Research Center: Internet, Science & Tech. 2017 [cited 2017 Aug 14]. Available from: http://www.pewinternet.org/2017/05/17/tech-adoption-climbs-among-older-adults/

39. Fox S. The social life of health information [Internet]. Pew Research Center. 2014 [cited 2017 Aug 12]. Available from: http://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/

40. Mueller J, Jay C, Harper S, Davies A, Vega J, Todd C. Web Use for Symptom Appraisal of Physical Health Conditions: A Systematic Review. J Med Internet Res [Internet] jmir.org; 2017 Jun 13 [cited 2017 Jul 29];19(6):e202. PMID:28611017

41. Vicari S, Cappai F. Health activism and the logic of connective action. A case study of rare disease patient organisations. Inf Commun Soc [Internet] Taylor & Francis; 2016 Nov 1;19(11):1653–1671. PMID:27499676

42. Oprescu F, Campo S, Lowe J, Andsager J, Morcuende JA. Online information exchanges for parents of children with a rare health condition: key findings from an online support community. J Med Internet Res [Internet] 2013 Jan 22;15(1):e16. PMID:23470259

43. Dewing C, Jones T. Future of Patient Data. Future Agenda, London; 2018.

44. Brain K, Moriarty Y, Quinn-Scoggins H, Cannings-John R, Grozeva D. Cancer symptom experience and help-seeking behaviour in the UK adult population during the COVID-19 pandemic [Internet]. CABS; 2021 Feb. Report No.: 1. Available from: https://cabs-study.yolasite.com/resources/CABS%20Policy%20briefing%20report%20FINAL_250221.pdf

45. Suresh S. Patient-generated health data can provide value in clinical care, research settings. AAP News [Internet] 2021 May 5 [cited 2021 May 5]; Available from: https://www.aappublications.org/news/2020/07/01/hit070120

46. Smith J, Bartlett J, Buck D, Honeyman M. Online Support: Investigating the role of public online forums in mental health [Internet]. DEMOS /The Kings Fund; 2017 Apr. Available from: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/Online_Support_Investigating_role_public_online_forums_mental_health.pdf

47. Brookes G. Corpus linguistics in illness and healthcare contexts: A case study of diabulimia support groups. Applying Linguistics in Illness and Healthcare Contexts [Internet] Bloomsbury Academic; 2016 [cited 2021 Mar 25]. [doi: 10.5040/9781350057685.0009]

48. Park A, Conway M, Chen AT. Examining thematic similarity, difference, and membership in three online mental health communities from reddit: A text mining and visualization approach. Comput Human Behav [Internet] 2018 Jan;78:98–112. PMID:29456286

49. Kallinikos J, Tempini N. Patient data as medical facts: social media practices as a foundation for medical knowledge creation. Information Systems Research [Internet] INFORMS Institute for Operations Res.and the Management Sciences; 2014 [cited 2016 Apr 15];25(4):817–833. [doi: 10.1287/isre.2014.0544]

50. Vayena E, Brownsword R, Edwards SJ, Greshake B, Kahn JP, Ladher N, Montgomery J, O'Connor D, O'Neill O, Richards MP, Rid A, Sheehan M, Wicks P, Tasioulas J. Research led by participants: a new social contract for a new kind of research. J Med Ethics [Internet] 2016 Apr;42(4):216–219. PMID:25825527

51. Workman TA. Engaging Patients in Information Sharing and Data Collection: The Role of Patient-Powered Registries and Research Networks [Internet]. Rockville (MD): Agency for Healthcare Research and Quality (US); 2013. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24156118

52. Frossard C, Marques Dias M. Interface - Communication, Health, Education. Interface [Internet] 2016 Feb;20(57). [doi: 10.1590/1807-57622014.1334]

53. Mazanderani F, O'Neill B, Powell J. "People power" or "pester power"? YouTube as a forum for the generation of evidence and patient advocacy. Patient Educ Couns [Internet] 2013 Dec;93(3):420–425. [doi: 10.1016/j.pec.2013.06.006]

54. Allen C, Vassilev I, Kennedy A, Rogers A. Long-Term Condition Self-Management Support in Online Communities: A Meta-Synthesis of Qualitative Papers. J Med Internet Res [Internet] eprints.soton.ac.uk; 2016 Mar 10;18(3):e61. PMID:26965990

55. van Berkel JJ, Lambooij MS, Hegger I. Empowerment of patients in online discussions about medicine use. BMC Med Inform Decis Mak [Internet] 2015 Apr 8;15:24. PMID:25888825

56. Mahr D. Self-reporting and participatory health platforms: Empowerment through sharing information about oneself online? | Bill of Health [Internet]. Bill of Health. 2017 [cited 2020 Jan 21]. Available from: https://blog.petrieflom.law.harvard.edu/2017/05/01/self-reporting-and-participatory-health-platforms-empowerment-through-sharing-information-about-oneself-online/

57. Hixson JD, Barnes D, Parko K, Durgin T, Van Bebber S, Graham A, Wicks P. Patients optimizing epilepsy management via an online community: the POEM Study. Neurology [Internet] 2015 Jul 14;85(2):129–136. PMID:26085605

58. McGregor F, Somner JEA, Bourne RR, Munn-Giddings C, Shah P, Cross V. Social media use by patients with glaucoma: what can we learn? Ophthalmic Physiol Opt [Internet] 2014 Jan;34(1):46–52. PMID:24325434

59. Carron-Arthur B, Ali K, Cunningham JA, Griffiths KM. From Help-Seekers to Influential Users: A Systematic Review of Participation Styles in Online Health Communities. JMIR 2015;

60. Ziebland S, Powell J, Briggs P, Jenkinson C, Wyke S, Sillence E, Harris P, Perera R, Mazanderani F, Martin A, Locock L, Kelly L, Booth M, Gann B, Newhouse N, Farmer A. Examining the role of patients' experiences as a resource for choice and decision-making in health care: a creative, interdisciplinary mixed-method study in digital health [Internet]. Southampton (UK): NIHR Journals Library; 2016. [doi: 10.3310/pgfar04170]

61. Vijayan J. Solving the unstructured data challenge. CIO 2015;

62. Abbasi A, Adjeroh D, Dredze M, Paul MJ, Zahedi FM, Zhao H, Walia N, Jain H, Sanvanson P, Shaker R, Huesch MD, Beal R, Zheng W, Abate M, Ross A. Social Media

Analytics for Smart Health. IEEE Intell Syst [Internet] 2014 Mar;29(2):60–64. [doi: 10.1109/MIS.2014.29]

63. Powell J, Boylan A-M, Greaves F. Harnessing patient feedback data: A challenge for policy and service improvement. Digital Health [Internet] SAGE Publications; 2015 Dec 1 [cited 2016 Oct 6];1:2055207615617910. PMID:29942547

64. Brooker P, Barnett J, Cribbin T. Doing social media analytics. Big Data & Society [Internet] SAGE Publications Ltd; 2016 Dec 1;3(2):2053951716658060. [doi: 10.1177/2053951716658060]

65. Khanbhai M, Anyadi P, Symons J, Flott K, Darzi A, Mayer E. Applying natural language processing and machine learning techniques to patient experience feedback: a systematic review. BMJ Health Care Inform [Internet] 2021 Mar;28(1). PMID:33653690

66. Gonzalez-Hernandez G, Sarker A, O'Connor K, Savova G. Capturing the Patient's Perspective: a Review of Advances in Natural Language Processing of Health-Related Text. Yearb Med Inform [Internet] Germany: thieme-connect.com; 2017 Aug;26(1):214–227. PMID:29063568

67. Dol J, Tutelman PR, Chambers CT, Barwick M, Drake EK, Parker JA, Parker R, Benchimol EI, George RB, Witteman HO. Health Researchers' Use of Social Media: Scoping Review. J Med Internet Res [Internet] 2019 Nov 13;21(11):e13687. PMID:31719028

68. Turing AM. I.—COMPUTING MACHINERY AND INTELLIGENCE [Internet]. Mind. 1950. p. 433–460. [doi: 10.1093/mind/lix.236.433]

69. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a Tool for Health Research: A Systematic Review. Am J Public Health [Internet] 2017 Jan;107(1):e1–e8. PMID:27854532

70. Ravoire S, Lang M, Perrin E, participants of Giens XXXII Round Table No. 6. Advantages and limitations of online communities of patients for research on health products. Therapie [Internet] Elsevier; 2017 Feb;72(1):135–143. PMID:28139283

71. Martinez B, Dailey F, Almario CV, Keller MS, Desai M, Dupuy T, Mosadeghi S, Whitman C, Lasch K, Ursos L, Spiegel BMR. Patient Understanding of the Risks and Benefits of Biologic Therapies in Inflammatory Bowel Disease: Insights from a Large-scale Analysis of Social Media Platforms. Inflamm Bowel Dis [Internet] 2017 Apr 13;23(7):1057-1064. doi: 10.1097/MIB.0000000000001110. PMID:28410343

72. Wilson V. Patient use of the internet for diabetes information. Nurs Times [Internet] 2013;109(23):18–20. PMID:23888634

73. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, Upadhaya T, Gonzalez G. Utilizing social media data for pharmacovigilance: A review. J Biomed Inform [Internet] 2015 Apr;54:202–212. PMID:25720841

74. Tricco AC, Zarin W, Lillie E, Jeblee S, Warren R, Khan PA, Robson R, Pham B, Hirst G, Straus SE. Utility of social media and crowd-intelligence data for pharmacovigilance: a scoping review. BMC Med Inform Decis Mak [Internet] 2018 Jun 14;18(1):38. PMID:29898743

75. Kalf RR, Makady A, Ten Ham RM, Meijboom K, Goettsch WG, IMI-GetReal Workpackage 1. Use of Social Media in the Assessment of Relative Effectiveness: Explorative Review With Examples From Oncology. JMIR Cancer [Internet] 2018 Jun 8;4(1):e11. PMID:29884607

76. Risson V, Saini D, Bonzani I, Huisman A, Olson M. Patterns of Treatment Switching in Multiple Sclerosis Therapies in US Patients Active on Social Media: Application of Social Media Content Analysis to Health Outcomes Research. J Med Internet Res [Internet] 2016 Mar 17;18(3):e62. PMID:26987964

77. Al-garadi MA, Khan MS, Varathan KD, Mujtaba G, Al-Kabsi AM. Using online social networks to track a pandemic: A systematic review. J Biomed Inform [Internet] 2016 Aug 1;62:1–11. [doi: 10.1016/j.jbi.2016.05.005]

78. Fung ICH, Duke CH, Finch KC, Snook KR, Tseng PL, Hernandez AC, Gambhir M, Fu KW, Tse ZTH. Ebola virus disease and social media: A systematic review. Am J Infect Control [Internet] 2016 Dec;44(12):1660-1671. doi: 10.1016/j.ajic.2016.05.011. Epub 2016 Jul 15. [doi: 10.1016/j.ajic.2016.05.011]

79. Lopez-Castroman J, Moulahi B, Aze J, Bringay S, Deninotti J, Guillaume S, Baca-Garcia E. Mining social networks to improve suicide prevention: A scoping review. J Neurosci Res [Internet] 2019; [doi: 10.1002/jnr.24404]

80. Nawab K, Ramsey G, Schreiber R. Natural Language Processing to Extract Meaningful Information from Patient Experience Feedback. Appl Clin Inform [Internet] 2020 Mar;11(2):242–252. PMID:32236917

81. Evans MS. A computational approach to qualitative analysis in large textual datasets. PLoS One [Internet] journals.plos.org; 2014 Feb 3;9(2):e87908. PMID:24498398

82. James Lind Alliance. James Lind Alliance [Internet]. James Lind Alliance. 2017 [cited 2017 Aug 7]. Available from: http://www.jla.nihr.ac.uk/

83. NHS. Five year forward view. NHS England; 2014.

84. Academy of Medical, Sciences. Next steps for using real world evidence: Summary report of a FORUM follow-up roundtable held 24 January 2018 [Internet]. AMS FORUM; 2018 p. 28. Available from: https://acmedsci.ac.uk/file-download/7021031

85. FDA. The Voice of the Patient: A Series of Reports from FDA's Patient-Focused Drug Development Initiative [Internet]. FDA. Center for Drug Evaluation and Research; 2017 [cited 2017 Aug 8]. Available from: http://www.fda.gov/ForIndustry/UserFees/PrescriptionDrugUserFee/ucm368342.htm

86. Burt J, Campbell J, Abel G, Aboulghate A, Ahmed F, Asprey A, Barry H, Beckwith J, Benson J, Boiko O, Bower P, Calitri R, Carter M, Davey A, Elliott MN, Elmore N, Farrington C, Haque HW, Henley W, Lattimer V, Llanwarne N, Lloyd C, Lyratzopoulos G, Maramba I, Mounce L, Newbould J, Paddison C, Parker R, Richards S, Roberts M, Setodji C, Silverman J, Warren F, Wilson E, Wright C, Roland M. Improving patient experience in primary care: a multimethod programme of research on the measurement and improvement of patient experience [Internet]. Southampton (UK): NIHR Journals Library; 2017. [doi: 10.3310/pgfar05090]

87. Portier K, Greer GE, Rokach L, Ofek N, Wang Y, Biyani P, Yu M, Banerjee S, Zhao K, Mitra P, Yen J. Understanding topics and sentiment in an online cancer survivor community. J Natl Cancer Inst Monogr [Internet] 2013 Dec;2013(47):195–198. PMID:24395991

88. Greaves F, Laverty AA, Cano DR, Moilanen K, Pulman S, Darzi A, Millett C. Tweets about hospital quality: a mixed methods study. BMJ Qual Saf [Internet] BMJ Publishing Group Ltd; 2014 Apr 19 [cited 2016 Apr 7];23(10):bmjqs-2014-002875. PMID:24748372

89. Brown MT, Bussell JK. Medication adherence: WHO cares? Mayo Clin Proc [Internet] 2011 Apr;86(4):304–314. PMID:21389250

90. Onishi T, Weissenbacher D, Klein A, O'Connor K, Gonzalez-Hernandez G. Dealing with medication non-adherence expressions in Twitter. Proceedings of the 2018 EMNLP Workshop SMM4H: The 3rd Social Media Mining for Health Applications Workshop & Shared Task [Internet] 2018. p. 32–33. Available from: https://www.aclweb.org/anthology/W18-5908.pdf

91. Seabury SA, Gupta CN, Philipson TJ, Henkhaus LE, PhRMA Medication Adherence Advisory Council. Understanding and overcoming barriers to medication adherence: a review of research priorities. J Manag Care Spec Pharm [Internet] 2014 Aug;20(8):775–783. PMID:25062070

92. Bigeard E, Grabar N, Thiessard F. Detection and Analysis of Drug Misuses. A Study Based on Social Media Messages. Front Pharmacol [Internet] 2018;9(JUL):791. [doi: 10.3389/fphar.2018.00791]

93. World Heath Organisation. Adherence to long-term therapies: Evidence for action [Internet]. WHO; 2003. Available from: https://www.who.int/chp/knowledge/publications/adherence_full_report.pdf

94. Burckhardt P, Padman R. Analyzing Self-Help Forums with Ontology-Based Text Mining: An Exploration in Kidney Space. AMIA Annu Symp Proc [Internet] 2015 Nov 5;2015:1821–30. eCollection 2015. PMID:26958281

95. Rastegar-Mojarad M, Ye Z, Wall D, Murali N, Lin S. Collecting and Analyzing Patient Experiences of Health Care From Social Media. JMIR Res Protoc [Internet] 2015 Jul 2;4(3):e78. PMID:26137885

96. Sarker A, DeRoos A, Perrone J. Mining social media for prescription medication abuse monitoring: a review and proposal for a data-centric framework. J Am Med Inform Assoc [Internet] 2019 Oct 4;27(2):315–329. PMID:31584645

97. Royal Society of Medicine. Medicine and Me programme at the RSM - The Royal Society of Medicine [Internet]. Royal Society of Medicine. 2017 [cited 2017 Aug 12]. Available from: https://www.rsm.ac.uk/about-us/public-information/patients-and-carers.aspx

98. FDA. The Voice of the Patient: CFS & ME [Internet]. FDA; 2013. Available from: http://www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM368806.pdf

99. Hackett KL, Lambson RL, Strassheim V, Gotts Z, Deary V, Newton JL. A concept mapping study evaluating the UK's first NHS generic fatigue clinic. Health Expect [Internet] 2015 Aug 1; PMID:26332418

100. Corrie CFA. Expert-patients [Internet]. Reform; 2015 Feb. Available from: http://www.reform.uk/wp-content/uploads/2015/02/Expert-patients.pdf

101. NHS RightCare. NHS RightCare [Internet]. NHS RightCare. 2017 [cited 2017 Aug 16]. Available from: https://www.england.nhs.uk/rightcare/what-is-nhs-rightcare/

102. Wellcome Trust. Accelerated Access Review: Final Report. Wellcome Trust; 2016 Oct.

103. McKee S. End of the road for Off-Patent Drugs Bill. Pharma Times [Internet] PharmaTimes Media Limited; 2015 Nov 9 [cited 2016 Oct 9]; Available from: http://www.pharmatimes.com/news/end_of_the_road_for_off-patent_drugs_bill_972123

104. Moran N. How to buy more health [Internet]. ScienceBusiness. 2016 [cited 2016 Nov 20]. Available from: http://sciencebusiness.net/news/79785/How-to-buy-more-health

105. Greenhalgh T, Snow R, Ryan S, Rees S, Salisbury H. Six `biases' against patients and carers in evidence-based medicine. BMC Med [Internet] 2015 Sep 1;13(1):200. PMID:26324223

106. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ [Internet] 1996 Jan 13;312(7023):71–72. PMID:8555924

107. Rocca E, Anjum RL. Causal Evidence and Dispositions in Medicine and Public Health. Int J Environ Res Public Health [Internet] Multidisciplinary Digital Publishing Institute; 2020 Mar 11 [cited 2020 Mar 13];17(6):1813. [doi: 10.3390/ijerph17061813]

108. Anjum RL, Copeland S, Rocca E. Medical scientists and philosophers worldwide appeal to EBM to expand the notion of 'evidence'. BMJ Evid Based Med [Internet] 2018 Nov 14;25(1):6–8. PMID:30429163

109. Russo F. Evidence, causation, scientific disagreement [Internet]. 2016 [cited 2021 Jun 19]. Available from: https://ebmplus.org/evidence-causation-disagreement/

110. Deaton A, Cartwright N. Understanding and misunderstanding randomized controlled trials. Soc Sci Med [Internet] 2018 Aug;210:2–21. PMID:29331519

111. Kerry R. Causal Dispositionalism and Evidence Based Healthcare. In: Anjum RL, Copeland S, Rocca E, editors. Rethinking Causality, Complexity and Evidence for the Unique Patient: A CauseHealth Resource for Healthcare Professionals and the Clinical Encounter [Internet] Cham: Springer International Publishing; 2020. p. 201–213. [doi: 10.1007/978-3-030-41239-5_13]

112. Brady P, Goldman M. Social media sourced Real World Evidence: New way to collect evidence? [Internet]. European Pharmaceutical Review. 2016 [cited 2017 Mar 16]. Available from: https://www.europeanpharmaceuticalreview.com/44295/news/blog/social-media-sourced-real-world/

113. Goldacre B. Bad Pharma: How Drug Companies Mislead Doctors and Harm Patients [Internet]. Farrar, Straus and Giroux; 2014. Available from: https://market.android.com/details?id=book-444XAwAAQBAJISBN:9780865478060

114. Altman DG. The scandal of poor medical research. BMJ [Internet] 1994 Jan 29;308(6924):283–284. PMID:8124111

115. CEBM. Catalogue of Bias - Catalog of Bias [Internet]. Catalog of Bias. 2017 [cited 2020 Feb 10]. Available from: https://catalogofbias.org/

116. Battleday RM, Brem AK, Battleday RM, Brem AK. Modafinil for cognitive neuroenhancement in healthy non-sleep-deprived subjects: A systematic review. Eur Neuropsychopharmacol [Internet] Elsevier; 2015 Jan 11;25(11):1865–1881. PMID:26708317

117. Ioannidis JPA. Why Most Clinical Research Is Not Useful. PLoS Med [Internet] 2016 Jun;13(6):e1002049. PMID:27328301

118. Wieringa S, Engebretsen E, Heggen K, Greenhalgh T. Has evidence-based medicine ever been modern? A Latour-inspired understanding of a changing EBM. J Eval Clin Pract [Internet] 2017 May 16; PMID:28508440

119. Sloane R, Osanlou O, Lewis D, Bollegala D, Maskell S, Pirmohamed M. Social media and pharmacovigilance: A review of the opportunities and challenges. Br J Clin Pharmacol [Internet] Wiley Online Library; 2015 Oct 1;80(4):910–920. [doi: 10.1111/bcp.12717]

120. Bensing J. Bridging the gap. The separate worlds of evidence-based medicine and patient-centered medicine. Patient Educ Couns [Internet] 2000 Jan;39(1):17–25. PMID:11013544

121. Anjum RL, Copeland S, Rocca E. Conclusion: CauseHealth Recommendations for Making Causal Evidence Clinically Relevant and Informed. In: Anjum RL, Copeland S, Rocca E, editors. Rethinking Causality, Complexity and Evidence for the Unique Patient: A CauseHealth Resource for Healthcare Professionals and the Clinical Encounter [Internet] Cham: Springer International Publishing; 2020. p. 237–241. [doi: 10.1007/978-3-030-41239-5_16]

122. Edwards R. Living with Complexity and Big Data [Internet]. Uppsala Monitoring Centre; 2018 Apr p. 36. Report No.: 78. Available from: https://view.publitas.com/uppsala-monitoring-centre/uppsala-reports-78/page/28

123. Anjum RL. What is the guidelines challenge? The CauseHealth perspective. J Eval Clin Pract [Internet] 2018 Oct;24(5):1127–1131. PMID:29799154

124. Anjum RL. Dispositions and the Unique Patient. In: Anjum RL, Copeland S, Rocca E, editors. Rethinking Causality, Complexity and Evidence for the Unique Patient: A CauseHealth Resource for Healthcare Professionals and the Clinical Encounter [Internet] Cham: Springer International Publishing; 2020. p. 13–36. [doi: 10.1007/978-3-030-41239-5_2]

125. Greenhalgh T. Health Evidence Network synthesis report 49 - Cultural Contexts of Health: The Use of Narrative Research in the Health Sector [Internet]. Copenhagen: WHO; 2016 Nov. Report No.: 49. Available from: https://www.ncbi.nlm.nih.gov/pubmed/27809425

126. Maxwell JA. The Importance of Qualitative Research for Causal Explanation in Education. Qual Inq [Internet] SAGE Publications Inc; 2012 Oct 1;18(8):655–661. [doi: 10.1177/1077800412452856]

127. JAMA. Instructions for Authors [Internet]. JAMA Network. 2021 [cited 2021 Apr 28]. Available from: https://jamanetwork.com/pages/jama-health-forum-instructions-for-authors

128. Williamson J. Establishing Causal Claims in Medicine. Int Stud Philos Sci [Internet] Routledge; 2019 Jan 2;32(1):33–61. [doi: 10.1080/02698595.2019.1630927]

129. Neeleman A, van de Koot H. The Linguistic Expression of Causation. The Theta System [Internet] Oxford: Oxford University Press; 2012. [doi: 10.1093/acprof:oso/9780199602513.003.0002]

130. InformedHealth.org. Depression: How effective are antidepressants? [Internet]. Institute for Quality and Efficiency in Health Care (IQWiG); 2020 [cited 2021 Jul 5]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK361016/

131. Pigott HE, Leventhal AM, Alter GS, Boren JJ. Efficacy and effectiveness of antidepressants: current status of research. Psychother Psychosom [Internet] 2010 Jul 9;79(5):267–279. PMID:20616621

132. Ioannidis JPA. Effectiveness of antidepressants: an evidence myth constructed from a thousand randomized trials? Philos Ethics Humanit Med [Internet] 2008 May 27;3:14. PMID:18505564

133. Greenhalgh T. Will COVID-19 be evidence-based medicine’s nemesis? PLoS Med [Internet] 2020 Jun;17(6):e1003266. PMID:32603323

134. Schilsky RL. Finding the Evidence in Real-World Evidence: Moving from Data to Information to Knowledge. J Am Coll Surg [Internet] 2017 Jan;224(1):1–7. PMID:27989954

135. Miani C, Robin E, Horvath V, Manville C, Cave J, Chataway J. Health and Healthcare: Assessing the Real-World Data Policy Landscape in Europe [Internet]. RAND Europe; 2014 Apr. Report No.: RR-544-PI. Available from: http://www.rand.org/pubs/research_reports/RR544.html

136. Averitt AJ, Weng C, Ryan P, Perotte A. Translating evidence into practice: eligibility criteria fail to eliminate clinically significant differences between real-world and study populations. npj Digital Medicine [Internet] 2020 May 11;3(1):67. [doi: 10.1038/s41746-020-0277-8]

137. Drewniak D, Glassel A, Hodel M, Biller-Andorno N. Risks and Benefits of Web-Based Patient Narratives: Systematic Review. J Med Internet Res [Internet] 2020 Mar;22(3). [doi: 10.2196/15772]

138. McKenna B, Myers MD, Newman M. Social media in qualitative research: Challenges and recommendations. Information and Organization [Internet] 2017 Jun 1;27(2):87–99. [doi: 10.1016/j.infoandorg.2017.03.001]

139. Vilar S, Friedman C, Hripcsak G. Detection of drug-drug interactions through data mining studies using clinical sources, scientific literature and social media. Brief

Bioinform [Internet] 2018 Sep 28;19(5):863-877. doi: 10.1093/bib/bbx010. PMID:28334070

140. Yin Z, Sulieman LM, Malin BA. A systematic literature review of machine learning in online personal health data. J Am Med Inform Assoc [Internet] 2019 Jun 1;26(6):561–576. PMID:30908576

141. Abbe A, Grouin C, Zweigenbaum P, Falissard B. Text mining applications in psychiatry: a systematic literature review. Int J Methods Psychiatr Res [Internet] 2016 Jun;25(2):86–100. PMID:26184780

142. Convertino I, Ferraro S, Blandizzi C, Tuccori M. The usefulness of listening social media for pharmacovigilance purposes: a systematic review. Expert Opin Drug Saf [Internet] 2018 Nov;17(11):1081–1093. PMID:30285501

143. Demner-Fushman D, Elhadad N. Aspiring to Unintended Consequences of Natural Language Processing: A Review of Recent Developments in Clinical and Consumer-Generated Text Processing. Yearb Med Inform [Internet] Germany; 2016 Nov 10;1(1):224-233. doi: 10.15265/IY-2016–017. PMID:27830255

144. Dreisbach C, Koleck TA, Bourne PE, Bakken S. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. Int J Med Inform [Internet] 2019 May;125(pp 37-46):37–46. PMID:30914179

145. Zunic A, Corcoran P, Spasic I. Sentiment Analysis in Health and Well-Being: Systematic Review. Jmir Medical Informatics [Internet] JMIR Medical Informatics; 2020 Jan [cited 2020 Jan 28];8(1):e16023. [doi: 10.2196/16023]

146. Ru B, Yao L. A Literature Review of Social Media-Based Data Mining for Health Outcomes Research. In: Bian J, Guo Y, He Z, Hu X, editors. Social Web and Health Research: Benefits, Limitations, and Best Practices [Internet] Cham: Springer International Publishing; 2019. p. 1–14. [doi: 10.1007/978-3-030-14714-3_1]

147. Kalambe YS, Pratiba D, Shah P. BIG DATA MINING TOOLS FOR UNSTRUCTURED DATA: A REVIEW. IJITR [Internet] 2015 May 6 [cited 2017 Aug 31];3(2):2012–2017. Available from: http://ijitr.com/index.php/ojs/article/view/610

148. Volpp KG, Mohta N. Health care has a lot to learn from consumer-friendly industries [Internet]. NEJM Catalyst; 2019 Jan p. 16. Available from: https://cdn2.hubspot.net/hubfs/558940/Insights%20Council%20Monthly%20-%20Files/Health%20Care%20Has%20a%20Lot%20to%20Learn%20from%20Consumer-Friendly%20Industries%20-%20update.pdf?__hssc=23193637.1.1549460954179&__hstc=23193637.2904168756e0524b3b7bb2995ba3f84c.1546521469842.1548070697066.1549460954179.3&__hsfp=19376341&hsCtaTracking=26afcc5c-e285-4afd-84d5-67872c47a375%7C42a6f9fd-d57d-4c92-909c-1c8b363b886e

149. Mohta NS, Prewitt E. Measuring What Matters and Capturing the Patient Voice. NEJM; 2017 Oct.

150. Demos / Vox Digitas. Social media is transforming how to study society [Internet]. Demos; 2014. Available from: https://www.demos.co.uk/files/Vox_Digitas_-_web.pdf?1408832211

151. Kruse CS, Goswamy R, Raval Y, Marawi S. Challenges and Opportunities of Big Data in Health Care: A Systematic Review. JMIR Med Inform [Internet] 2016 Nov 21;4(4):e38. PMID:27872036

152. Lindstrom M. Small data: the tiny clues that uncover huge trends. St. Martin's Press; 2016.

153. Dobrossy B, Girasek E, Susanszky A, Koncz Z, Gyorffy Z, Bognar VK. 'Clicks, likes, shares and comments' a systematic review of breast cancer screening discourse in social media. PLoS One [Internet] 2020 Apr;15(4). [doi: 10.1371/journal.pone.0231422]

154. Correia RB, Wood IB, Bollen J, Rocha LM. Mining Social Media Data for Biomedical Signals and Health-Related Behavior. Annu Rev Biomed Data Sci [Internet] annualreviews.org; 2020 Jul;3:433–458. PMID:32550337

155. Munson SA, Cavusoglu H, Frisch L, Fels S. Sociotechnical challenges and progress in using social media for health. J Med Internet Res [Internet] 2013 Oct 22;15(10):e226. PMID:24148206

156. Hodgkin P, Horsley L, Metz B. The Emerging World of Online Health Communities [Internet]. Stanford Social Innovation Review. 2018 [cited 2020 Mar 10]. Available from: https://ssir.org/articles/entry/the_emerging_world_of_online_health_communities

157. Househ M, Borycki E, Kushniruk A. Empowering patients through social media: the benefits and challenges. Health Informatics J [Internet] 2014 Mar;20(1):50–58. PMID:24550564

158. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res [Internet] Journal of Medical Internet Research; 2013 Apr 23 [cited 2016 Mar 1];15(4):e85. PMID:23615206

159. Barros JM, Duggan J, Rebholz-Schuhmann D. The Application of Internet-Based Sources for Public Health Surveillance (Infoveillance): Systematic Review. J Med Internet Res [Internet] 2020 Mar;22(3):e13680. [doi: 10.2196/13680]

160. Gianfredi V, Bragazzi NL, Nucci D, Martini M, Rosselli R, Minelli L, Moretti M. Harnessing Big Data for Communicable Tropical and Sub-Tropical Disorders: Implications From a Systematic Review of the Literature. Frontiers in Public Health [Internet] 2018 Mar;6. [doi: 10.3389/fpubh.2018.00090]

161. Olteanu A, Castillo C, Diaz F, Kıcıman E. Social Data: Biases, Methodological Pitfalls, and Ethical Boundaries. Front Big Data [Internet] 2019 Jul 11;2:13. PMID:33693336

162. Guetterman TC, Fetters MD, Creswell JW. Integrating Quantitative and Qualitative Results in Health Science Mixed Methods Research Through Joint Displays. Ann Fam Med [Internet] 2015 Nov 1;13(6):554–561. PMID:26553895

163. Guetterman TC, Chang T, DeJonckheere M, Basu T, Scruggs E, Vydiswaran VGV. Augmenting qualitative text analysis with natural language processing: methodological study. J Med Internet Res [Internet] 2018 Jun 29 [cited 2018 Aug 21];20(6):e231. PMID:29959110

164. Yin RK. Applications of Case Study Research [Internet]. SAGE; 2011. Available from: https://play.google.com/store/books/details?id=FgSV0Y2FleYCISBN:9781452235783

165. Crowe S, Cresswell K, Robertson A, Huby G, Avery A, Sheikh A. The case study approach. BMC Med Res Methodol [Internet] 2011 Jun 27;11:100. PMID:21707982

166. Pokharel R, Haghighi PD, Jayaraman PP, Georgakopoulos D. Analysing Emerging Topics across Multiple Social Media Platforms. Proceedings of the Australasian Computer Science Week Multiconference [Internet] New York, NY, USA: Association for Computing Machinery; 2019 [cited 2020 Jul 10]. p. 1–9. [doi: 10.1145/3290688.3290720]

167. Gupta A, Katarya R. Social media based surveillance systems for healthcare using machine learning: A systematic review. J Biomed Inform [Internet] 2020 Aug;108. [doi: 10.1016/j.jbi.2020.103500]

168. Coviello L, Sohn Y, Kramer ADI, Marlow C, Franceschetti M, Christakis NA, Fowler JH. Detecting emotional contagion in massive social networks. PLoS One [Internet] 2014 Mar 12;9(3):e90315. PMID:24621792

169. Kramer ADI, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. Proc Natl Acad Sci U S A [Internet] United States; 2014 Jun 17;111(24):8788–8790. PMID:24889601

170. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol [Internet] Routledge; 2006 Jan 1;3(2):77–101. [doi: 10.1191/1478088706qp063oa]

171. Green J, Thorogood N. Qualitative methods for health research [Internet]. Sage Publications; 2013. Available from: https://books.google.com/books?hl=en&lr=&id=qg0tAgAAQBAJ&oi=fnd&pg=PP1&dq=green%2Bthorogood%2Bqualitative%2Bmethods%2Bfor%2Bhealth%2Bresearch&ots=JDTZNxwKue&sig=MGmMr7YNT-4t277Dvd-OBd9T8fI

172. Green J, Britten N. Qualitative research and evidence based medicine. BMJ [Internet] British Medical Journal Publishing Group; 1998 Apr 18 [cited 2021 Jun 19];316(7139):1230–1232. PMID:9583929

173. Greaves F, Ramirez-Cano D, Millett C, Darzi A, Donaldson L. Use of sentiment analysis for capturing patient experience from free-text comments posted online. J Med Internet Res [Internet] Journal of Medical Internet Research; 2013 Nov 1 [cited 2016 Apr 7];15(11):e239. doi: 10.2196/jmir.2721. PMID:24184993

174. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. Int J Evid Based Healthc [Internet] 2015 Sep;13(3):132–140. PMID:26360830

175. Calvo RA, Milne DN, Sazzad Hussain M, Christensen H. Natural language processing in mental health applications using non-clinical texts†. Natural Language Engineering [Internet] Cambridge University Press; 2017 Sep [cited 2019 Sep 13];23(5):649–685. [doi: 10.1017/S1351324916000383]

176. Chiauzzi E, Wicks P. Digital Trespass: Ethical and Terms-of-Use Violations by Researchers Accessing Data From an Online Patient Community. J Med Internet Res

[Internet] Journal of Medical Internet Research; 2019 [cited 2019 Feb 22];21(2):e11985. [doi: 10.2196/jmir.11985]

177. Nicholas J, Onie S, Larsen ME. Ethics and Privacy in Social Media Research for Mental Health. Curr Psychiatry Rep [Internet] Springer; 2020 Nov 23;22(12):84. PMID:33225404

178. Ford E, Shepherd S, Jones K, Hassan L. Toward an Ethical Framework for the Text Mining of Social Media for Health Research: A Systematic Review. Frontiers in Digital Health [Internet] 2021;2:62. [doi: 10.3389/fdgth.2020.592237]

179. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA [Internet] 2013 Nov 27;310(20):2191–2194. PMID:24141714

180. Golder S, Scantlebury A, Christmas H. Understanding Public Attitudes Toward Researchers Using Social Media for Detecting and Monitoring Adverse Events Data: Multi Methods Study. Journal of Medical Internet Research [Internet] Journal of Medical Internet Research; 2019 Aug [cited 2019 Aug 30];21(8):e7081. [doi: 10.2196/jmir.7081]

181. Golder S, Ahmed S, Norman G, Booth A. Attitudes Toward the Ethics of Research Using Social Media: A Systematic Review. J Med Internet Res [Internet] jmir.org; 2017 Jun 6;19(6):e195. PMID:28588006

182. Markham A, Buchanan E. Ethical Decison Making and Internet Research 2.0. AOIR; 2012.

183. The British Psychological Society. Ethics Guidelines for Internet-mediated Research [Internet]. British Psychological Society; 2017. Report No.: INF206. Available from: https://www.bps.org.uk/sites/bps.org.uk/files/Policy/Policy%20-%20Files/Ethics%20Guidelines%20for%20Internet-mediated%20Research%20%282017%29.pdf

184. Beninger K. Social media users' views on the ethics of social media research. The SAGE handbook of social media research methods [Internet] books.google.com; 2017;57–73. Available from: https://books.google.co.uk/books?hl=en&lr=&id=9oewDQAAQBAJ&oi=fnd&pg=PA57&dq=social+media+research+ethics&ots=eOINl4xWAP&sig=j8mbM9L9dT-mWgc-iDn4rAFFOE0

185. Garchitorena M. Meet The Women Who Forced The FDA To Listen About This Controversial Birth Control Device [Internet]. 2016 [cited 2017 May 22]. Available from: http://www.refinery29.uk/2016/04/107550/essure-problems-fda-lawsuit

186. reddit. reddit.com: api documentation [Internet]. reddit.com. 2021 [cited 2021 Jun 19]. Available from: https://www.reddit.com/dev/api

187. Edo-Osagie O, De La Iglesia B, Lake I, Edeghere O. A scoping review of the use of Twitter for public health research. Comput Biol Med [Internet] 2020 Jul;122:103770. PMID:32502758

188. Cooper AK, Coetzee S. On the Ethics of Using Publicly-Available Data. In: Hattingh M., Matthee M., Smuts H., Pappas I., Dwivedi Y., Mäntymäki M., editor. Responsible

Design, Implementation and Use of Information and Communication Technology [Internet] Cham: Springer International Publishing; 2020. p. 159–171. [doi: 10.1007/978-3-030-45002-1_14]

189. What is Open Data? [Internet]. Open Knowledge Foundation. 2020 [cited 2021 Jan 24]. Available from: https://opendatahandbook.org/guide/en/what-is-open-data/

190. Taylor J, Pagliari C. Mining social media data: How are research sponsors and researchers addressing the ethical challenges? Research Ethics [Internet] SAGE Publications Ltd; 2018 Apr 1;14(2):1747016117738559. [doi: 10.1177/1747016117738559]

191. Benton A, Coppersmith G, Dredze M. Ethical Research Protocols for Social Media Health Research. Proceedings of the First ACL Workshop on Ethics in Natural Language Processing [Internet] Valencia, Spain: Association for Computational Linguistics; 2017. p. 94–102. [doi: 10.18653/v1/W17-1612]

192. Privacy Policy [Internet]. drugs,com. 2020 [cited 2021 Jan 14]. Available from: https://www.drugs.com/support/privacy.html

193. WebMD Privacy Policy [Internet]. WebMD. 2020 [cited 2021 Jan 14]. Available from: https://www.webmd.com/about-webmd-policies/about-privacy-policy

194. U.S. Copyright Office. More Information on Fair Use [Internet]. copyright.gov. 2020 [cited 2021 Jan 14]. Available from: https://www.copyright.gov/fair-use/more-info.html

195. British Library. What Is Fair Use? Fair Dealing Copyright Explained [Internet]. British Library. The British Library; 2020 [cited 2021 Jan 14]. Available from: https://www.bl.uk/business-and-ip-centre/articles/fair-use-copyright-explained

196. Gräßer F, Kallumadi S, Malberg H, Zaunseder S. UCI Machine Learning Repository: Drug Review Dataset (Drugs.com) Data Set [Internet]. UCI Machine Learning Repository. 2018 [cited 2021 Jun 19]. Available from: https://archive.ics.uci.edu/ml/datasets/Drug+Review+Dataset+%28Drugs.com%29

197. VERBI Software. MAXQDA [Internet]. 2018 [cited 2017 Dec 21]. Available from: www.maxqda,com

198. McAndrew TC, Bongard JC, Danforth CM, Dodds PS, Hines PDH, Bagrow JP. What We Write About when We Write About Causality: Features of Causal Statements Across Large-scale Social Discourse. Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining [Internet] Piscataway, NJ, USA: IEEE Press; 2016. p. 519–524. Available from: http://dl.acm.org/citation.cfm?id=3192424.3192521

199. Kilgarriff A, Baisa V, Bušta J, Jakubíček M, Kovar V, Michelfeit J, Rychlý P, Suchomel V. SketchEngine [Internet]. Sketch Engine. 2020 [cited 2020 May 21]. Available from: https://www.sketchengine.eu/

200. Lee J. Writing Linguistic Rules for Natural Language Processing [Internet]. Medium. Towards Data Science; 2019 [cited 2020 Apr 28]. Available from: https://towardsdatascience.com/linguistic-rule-writing-for-nlp-ml-64d9af824ee8

201. SketchEngine.eu. English Penn Treebank tagset with modifications [Internet]. SketchEngine.eu. 2018 [cited 2018 Apr 20]. Available from: https://www.sketchengine.eu/english-treetagger-pipeline-2/

202. Ameisen E. How to solve 90% of NLP problems: a step-by-step guide. Insight Fellows Program; 2018 Jan 24 [cited 2019 Aug 2]; Available from: https://blog.insightdatascience.com/how-to-solve-90-of-nlp-problems-a-step-by-step-guide-fda605278e4e

203. TenTen Corpus Family | Sketch Engine [Internet]. Sketch Engine. 2015 [cited 2020 May 21]. Available from: https://www.sketchengine.eu/documentation/tenten-corpora/

204. Sateia MJ. International classification of sleep disorders-third edition: highlights and modifications. Chest [Internet] 2014 Nov;146(5):1387–1394. PMID:25367475

205. Kilgarriff A. Getting to Know Your Corpus. Text, Speech and Dialogue [Internet] Springer Berlin Heidelberg; 2012. p. 3–15. [doi: 10.1007/978-3-642-32790-2_1]

206. Manning C, Raghavan P, Schütze H. Introduction to Information Retrieval. Natural Language Engineering [Internet] Cambridge University Press; 2010 [cited 2019 May 29];16(1):100–103. [doi: 10.1017/S1351324909005129]

207. Sinclair J. Corpus, concordance, collocation. Oxford University Press; 1991. ISBN:9780194371445

208. Loria S. TextBlob [Internet]. 2020. Available from: https://textblob.readthedocs.io/en/dev/index.html

209. Joint Formulary Committee. BNF 75: March - September 2018: Modafinil. London: Pharmaceutical Press; 2018. ISBN:9780857113313

210. Baumer EPS, Mimno D, Guha S, Quan E, Gay GK. Comparing grounded theory and topic modeling: Extreme divergence or unlikely convergence? Journal of the Association for Information Science and Technology [Internet] 2017 Jun 28;68(6):1397–1410. [doi: 10.1002/asi.23786]

211. Bollegala D, Maskell S, Sloane R, Hajne J, Pirmohamed M. Causality Patterns for Detecting Adverse Drug Reactions From Social Media: Text Mining Approach. JMIR Public Health Surveill [Internet] 2018 May 9;4(2):e51. PMID:29743155

212. Mayer-Schönberger V, Cukier K. Big Data: A Revolution that Will Transform how We Live, Work, and Think [Internet]. London: Houghton Mifflin Harcourt; 2013. Available from: https://play.google.com/store/books/details?id=uy4lh-WEhhICISBN:9780544002692

213. Prainsack B. Through Thick and Big: Data-Rich Medicine in the Era of Personalisation. In: Vollman J, Sandow V, Wäscher S, Schildmann J, editors. The Ethics of Personalised Medicine [Internet] Farnham: Ashgate; 2015. p. 161–172. Available from: http://www.webcitation.org/query?url=https%3A%2F%2Fs3.amazonaws.com%2Facademia.edu.documents%2F37568485%2FPrainsack_in_Vollmann_et_al._2015.pdf%3FAWSAccessKeyId%3DAKIAIWOWYYGZ2Y53UL3A%26Expires%3D1559133050%26Signature%3DflNibbTnoM5Cr37wyROKr0Y9MTI%253D%26response-content-disposition%3Dinline%253B%2520filename%253DThrough_Thick&date=2019-05-29

214. Sim I. Two Ways of Knowing: Big Data and Evidence-Based Medicine. Ann Intern Med [Internet] 2016 Apr 19;164(8):562–563. PMID:26809201

215. Chalmers I, Oxman AD, Austvoll-Dahlgren A, Ryan-Vig S, Pannell S, Sewankambo N, Semakula D, Nsangi A, Albarqouni L, Glasziou P, Mahtani K, Nunan D, Heneghan C, Badenoch D. Key concepts for informed health choices: a framework for helping people learn how to assess treatment claims and make informed choices. BMJ Evidence-Based Medicine [Internet] Royal Society of Medicine; 2018 Feb;23(1):29–33. PMID:29367324

216. Fedak KM, Bernal A, Capshaw ZA, Gross S. Applying the Bradford Hill criteria in the 21st century: how data integration has changed causal inference in molecular epidemiology. Emerg Themes Epidemiol [Internet] 2015 Sep 30;12:14. PMID:26425136

217. Colagiuri B, Schenk LA, Kessler MD, Dorsey SG, Colloca L. The placebo effect: From concepts to genes. Neuroscience [Internet] Elsevier; 2015 Oct 29;307:171–190. PMID:26272535

218. Beecher HK. The powerful placebo. J Am Med Assoc [Internet] 1955 Dec 24;159(17):1602–1606. PMID:13271123

219. Kienle GS, Kiene H. The powerful placebo effect: fact or fiction? J Clin Epidemiol [Internet] 1997 Dec;50(12):1311–1318. PMID:9449934

220. Dirkson A, Verberne S, Sarker A, Kraaij W. Data-Driven Lexical Normalization for Medical Social Media. Multimodal Technologies and Interaction [Internet] Multidisciplinary Digital Publishing Institute; 2019 Aug 20 [cited 2019 Sep 3];3(3):60. [doi: 10.3390/mti3030060]

221. Thomson P. not all literature 'reviews' are the same [Internet]. patter. 2013 [cited 2017 Aug 5]. Available from: https://patthomson.net/2013/05/23/not-all-literature-reviews-are-the-same/

222. Grant MJ, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J [Internet] 2009 Jun;26(2):91–108. PMID:19490148

223. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol [Internet] 2005;8(1):19–32. [doi: 10.1080/1364557032000119616]

224. Neveol A, Zweigenbaum P, Section Editors for the IMIA Yearbook Section on Clinical Natural Language Processing. Expanding the Diversity of Texts and Applications: Findings from the Section on Clinical Natural Language Processing of the International Medical Informatics Association Yearbook. Yearb Med Inform [Internet] Germany; 2018 Aug;27(1):193-198. doi: 10.1055/s-0038-1667080. Epub 2018 Aug 29. [doi: 10.1055/s-0038-1667080]

225. Rayyan. Rayyan [Internet]. 2017 [cited 2017 Sep 5]. Available from: https://rayyan.qcri.org/

226. Lewin S, Booth A, Glenton C, Munthe-Kaas H, Rashidian A, Wainwright M, Bohren MA, Tunçalp Ö, Colvin CJ, Garside R, Carlsen B, Langlois EV, Noyes J. Applying GRADE-CERQual to qualitative evidence synthesis findings: introduction to the series. Implement Sci [Internet] 2018 Jan 25;13(1):2. [doi: 10.1186/s13012-017-0688-3]

227. dos Santos BS, Steiner MTA, Fenerich AT, Lima RHP. Data mining and machine learning techniques applied to public health problems: A bibliometric analysis from 2009 to 2018. Comput Ind Eng [Internet] 2019 Dec 1;138:106120. [doi: 10.1016/j.cie.2019.106120]

228. Mavragani A. Infodemiology and Infoveillance: Scoping Review. J Med Internet Res [Internet] 2020 Apr;22(4):e16206. [doi: 10.2196/16206]

229. Zhang Y, Cao BL, Wang YF, Peng TQ, Wang XH. When Public Health Research Meets Social Media: Knowledge Mapping From 2000 to 2018. J Med Internet Res [Internet] 2020 Aug;22(8). [doi: 10.2196/17582]

230. Névéol A, Zweigenbaum P. Making Sense of Big Textual Data for Health Care: Findings from the Section on Clinical Natural Language Processing. Yearb Med Inform [Internet] 2017 Aug;26(1):228–234. PMID:29063569

231. Staccini P, Fernandez-Luque L. Secondary Use of Recorded or Self-expressed Personal Data: Consumer Health Informatics and Education in the Era of Social Media and Health Apps. Yearb Med Inform [Internet] 2017 Aug;26(1):172-177. doi: 10.15265/IY-2017–037. Epub 2017 Sep 11. PMID:29063560

232. Kim SJ, Marsch LA, Hancock JT, Das AK. Scaling Up Research on Drug Abuse and Addiction Through Social Media Big Data. J Med Internet Res [Internet] Journal of Medical Internet Research; 2017 Oct 31 [cited 2017 Nov 1];19(10):e353. doi: 10.2196/jmir.6426. PMID:29089287

233. Zhang AJ, Albrecht L, Scott SD. Using Twitter for Data Collection With Health-Care Consumers: A Scoping Review. International Journal of Qualitative Methods [Internet] 2018 Jan;17(1). [doi: 10.1177/1609406917750782]

234. Hamad EO, Savundranayagam MY, Holmes JD, Kinsella EA, Johnson AM. Toward a Mixed-Methods Research Approach to Content Analysis in The Digital Age: The Combined Content-Analysis Model and its Applications to Health Care Twitter Feeds. J Med Internet Res 2016;18(3):e60. doi: 10.2196/jmir.5391.

235. Charles-Smith LE, Reynolds TL, Cameron MA, Conway M, Lau EHY, Olsen JM, Pavlin JA, Shigematsu M, Streichert LC, Suda KJ, Corley CD. Using Social Media for Actionable Disease Surveillance and Outbreak Management: A Systematic Literature Review. PLoS One [Internet] 2015 Oct 5;10(10):e0139701. PMID:26437454

236. Filannino M, Uzuner O. Advancing the State of the Art in Clinical Natural Language Processing through Shared Tasks. Yearb Med Inform [Internet] Germany; 2018 Aug;27(1):184–192. [doi: 10.1055/s-0038-1667079]

237. Abd Rahman R, Omar K, Noah SAM, Danuri M, Al-Garadi MA. Application of Machine Learning Methods in Mental Health Detection: A Systematic Review. Ieee Access [Internet] 2020;8:183952–183964. [doi: 10.1109/ACCESS.2020.3029154]

238. Castillo-Sánchez G, Marques G, Dorronzoro E, Rivera-Romero O, Franco-Martín M, De la Torre-Díez I. Suicide Risk Assessment Using Machine Learning and Social Networks: a Scoping Review. J Med Syst [Internet] 2020 Nov 9;44(12):205. PMID:33165729

239. Karmegam D, Ramamoorthy T, Mappillairajan B. A Systematic Review of Techniques Employed for Determining Mental Health Using Social Media in Psychological

Surveillance During Disasters. Disaster Med Public Health Prep [Internet] 2020 Apr;14(2):265–272. [doi: 10.1017/dmp.2019.40]

240. Sharma C, Whittle S, Haghighi PD, Burstein F, Keen H. Sentiment analysis of social media posts on pharmacotherapy: A scoping review. Pharmacology Research and Perspectives [Internet] 2020;8(5):e00640. [doi: 10.1002/prp2.640]

241. Lardon J, Abdellaoui R, Bellet F, Asfari H, Souvignet J, Texier N, Jaulent M-C, Beyens M-N, Burgun A, Bousquet C. Adverse Drug Reaction Identification and Extraction in Social Media: A Scoping Review. J Med Internet Res [Internet] Journal of Medical Internet Research; 2015 Jul 10;17(7):e171. PMID:26163365

242. Giuntini FT, Cazzolato MT, dos Reis M de JD, Campbell AT, Traina AJM, Ueyama J. A review on recognizing depression in social networks: challenges and opportunities. J Ambient Intell Humaniz Comput [Internet] 2020 Jan 24;11(11):4713–4729. [doi: 10.1007/s12652-020-01726-4]

243. Qiao J. A Systematic Review of Machine Learning Approaches for Mental Disorder Prediction on Social Media. 2020 International Conference on Computing and Data Science (CDS) [Internet] 2020. p. 433–438. [doi: 10.1109/CDS49703.2020.00091]

244. Wilson E, Kenny A, Dickson-Swift V. Using Blogs as a Qualitative Health Research Tool: A Scoping Review. International Journal of Qualitative Methods [Internet] 2015 Dec 1;14(5). [doi: 10.1177/1609406915618049]

245. Wong A, Plasek JM, Montecalvo SP, Zhou L. Natural Language Processing and Its Implications for the Future of Medication Safety: A Narrative Review of Recent Advances and Challenges. Pharmacotherapy [Internet] United States; 2018 Aug;38(8):822-841. doi: 10.1002/phar.2151. Epub 2018 Jul 22. [doi: 10.1002/phar.2151]

246. Cheerkoot-Jalim S, Kumar KK. A systematic review of text mining approaches applied to various application areas in the biomedical domain. Journal of Knowledge Management [Internet] 2020 Jan 1;ahead-of-print(ahead-of-print). [doi: 10.1108/JKM-09-2019-0524]

247. Su C, Xu ZX, Pathak J, Wang F. Deep learning in mental health outcome research: a scoping review. Transl Psychiatry [Internet] 2020 Apr;10(1). [doi: 10.1038/s41398-020-0780-3]

248. Golder S, Norman G, Loke YK. Systematic review on the prevalence, frequency and comparative value of adverse events data in social media. Br J Clin Pharmacol [Internet] Wiley Online Library; 2015 Oct;80(4):878–888. PMID:26271492

249. Injadat M, Salo F, Nassif AB. Data mining techniques in social media: A survey. Neurocomputing [Internet] 2016 Nov 19;214:654–670. [doi: 10.1016/j.neucom.2016.06.045]

250. Lau AYS, Staccini P, Section Editors for the, Imia Yearbook Section on Education, Consumer Health, Informatics. Artificial Intelligence in Health: New Opportunities, Challenges, and Practical Implications. Yearb Med Inform [Internet] 2019 Aug;28(1):174–178. PMID:31419829

251. Lafferty NT, Manca A. Perspectives on social media in and as research: A synthetic review. Int Rev Psychiatry [Internet] Taylor & Francis; 2015;27(2):85–96. [doi: 10.3109/09540261.2015.1009419]

252. Wongkoblap A, Vadillo MA, Curcin V. Researching Mental Health Disorders in the Era of Social Media: Systematic Review. J Med Internet Res [Internet] 2017 Jun 29;19(6):e228. doi: 10.2196/jmir.7215. PMID:28663166

253. Skaik R, Inkpen D. Using Social Media for Mental Health Surveillance: A Review. ACM Comput Surv [Internet] 2020;53(6):Article 129. [doi: 10.1145/3422824]

254. Ho T-B, Le L, Thai DT, Taewijit S. Data-driven Approach to Detect and Predict Adverse Drug Reactions. Curr Pharm Des [Internet] 2016;22(23):3498–526. doi: 10.2174/1381612822666160509125047. PMID:27157416

255. Pourebrahim B, Keyvanpour M. Adverse Drug Reaction Detection Using Data Mining Techniques: A Review Article. 2020. p. 118–123. [doi: 10.1109/ICCKE50421.2020.9303709]

256. Patel R, Chang T, Greysen SR, Chopra V. Social Media Use in Chronic Disease: A Systematic Review and Novel Taxonomy. Am J Med [Internet] 2015 Dec;128(12):1335–1350. PMID:26159633

257. Falisi AL, Wiseman KP, Gaysynsky A, Scheideler JK, Ramin DA, Chou W-YS. Social media for breast cancer survivors: a literature review. J Cancer Surviv [Internet] 2017 Dec;11(6):808–821. PMID:28601981

258. Sharma V, Holmes JH, Sarkar IN. Identifying Complementary and Alternative Medicine Usage Information from Internet Resources. A Systematic Review. Methods Inf Med [Internet] 2016 Aug 5;55(4):322–332. PMID:27352304

259. Cashion AK, Gill J, Hawes R, Henderson WA, Saligan L. National Institutes of Health Symptom Science Model sheds light on patient symptoms. Nurs Outlook [Internet] 2016 Sep;64(5):499–506. PMID:27349632

260. Rosen A. Tweeting Made Easier [Internet]. blog.twitter.com. 2017 [cited 2021 Jul 3]. Available from: https://blog.twitter.com/en_us/topics/product/2017/tweetingmadeeasier

261. Marjanovic S, Ghiga I, Yang M, Knack A. Understanding value in health data ecosystems [Internet]. RAND; 2017. Report No.: RR 1972. [doi: 10.7249/RR1972]

262. Giustini D, Ali SM, Fraser M, Kamel Boulos MN. Effective uses of social media in public health and medicine: a systematic review of systematic reviews. Online J Public Health Inform [Internet] 2018 Sep 21;10(2):e215. PMID:30349633

263. Salian I. NVIDIA Blog: What's the Difference Between Supervised & Unsupervised Learning? [Internet]. 2018 [cited 2020 Dec 17]. Available from: https://blogs.nvidia.com/blog/2018/08/02/supervised-unsupervised-learning/

264. Neves M, Ševa J. An extensive review of tools for manual annotation of documents. Brief Bioinform [Internet] 2019 Dec 15; PMID:31838514

265. Spasic I, Nenadic G. Clinical Text Data in Machine Learning: Systematic Review. JMIR Med Inform [Internet] 2020 Mar 31;8(3):e17984. PMID:32229465

266. Aylin Tokuç A. Splitting a Dataset into Train and Test Sets [Internet]. Baeldung.com. 2021 [cited 2021 Jun 19]. Available from: https://www.baeldung.com/cs/train-test-datasets-ratio

267. Metz C. Who Is Making Sure the A.I. Machines Aren't Racist? The New York Times [Internet] The New York Times; 2021 Mar 15 [cited 2021 Jun 8]; Available from: https://www.nytimes.com/2021/03/15/technology/artificial-intelligence-google-bias.html

268. Rao P. Fine-grained Sentiment Analysis in Python (Part 1) - Towards Data Science [Internet]. Towards Data Science. Towards Data Science; 2019 [cited 2021 Mar 24]. Available from: https://towardsdatascience.com/fine-grained-sentiment-analysis-in-python-part-1-2697bb111ed4

269. Bender EM, Gebru T, McMillan-Major A, Shmitchell S. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency [Internet] New York, NY, USA: Association for Computing Machinery; 2021 [cited 2021 Jun 8]. p. 610–623. [doi: 10.1145/3442188.3445922]

270. Ibrahim H, Liu X, Zariffa N, Morris AD, Denniston AK. Health data poverty: an assailable barrier to equitable digital health care. Lancet Digit Health [Internet] 2021 Apr;3(4):e260–e265. PMID:33678589

271. Skreta M, Arbabi A, Wang J, Brudno M. Training without training data: Improving the generalizability of automated medical abbreviation disambiguation. In: Dalca AV, McDermott MBA, Alsentzer E, Finlayson SG, Oberst M, Falck F, Beaulieu-Jones B, editors. Proceedings of the Machine Learning for Health NeurIPS Workshop [Internet] PMLR; 2020. p. 233–245. Available from: http://proceedings.mlr.press/v116/skreta20a.html

272. Miotto R, Li L, Kidd BA, Dudley JT. Deep Patient: An Unsupervised Representation to Predict the Future of Patients from the Electronic Health Records. Sci Rep [Internet] 2016 May 17;6:26094. PMID:27185194

273. Marcus G. Deep Learning: A Critical Appraisal [Internet]. arXiv [csAI]. 2018. Available from: http://arxiv.org/abs/1801.00631

274. Goldberg Y. The missing elements in NLP [Internet]. Youtube; 2019 [cited 2021 Jun 21]. Available from: https://www.youtube.com/watch?v=e12danHhlic

275. Marcus G. We can't trust AI systems built on deep learning alone [Internet]. MIT Technology Review. 2019 [cited 2020 Sep 11]. Available from: https://www.technologyreview.com/2019/09/27/65250/we-cant-trust-ai-systems-built-on-deep-learning-alone/

276. Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding [Internet]. arXiv [csCL]. 2018. Available from: http://arxiv.org/abs/1810.04805

277. Strubell E, Ganesh A, McCallum A. Energy and policy considerations for deep learning in NLP [Internet]. arXiv [csCL]. 2019. Available from: http://arxiv.org/abs/1906.02243

278. van der Scharr M, Maxfield N. Making machine learning interpretable: a dialog with clinicians [Internet]. van der Scharr Lab. 2021 [cited 2021 Jun 10]. Available from: https://www.vanderschaar-lab.com/making-machine-learning-interpretable-a-dialog-with-clinicians/

279. Wójcik R. Unsupervised Sentiment Analysis - Towards Data Science [Internet]. Towards Data Science. 2019 [cited 2020 Sep 22]. Available from: https://towardsdatascience.com/unsupervised-sentiment-analysis-a38bf1906483

280. Spasic I, Button K. Patient Triage by Topic Modeling of Referral Letters: Feasibility Study. JMIR Medical Informatics [Internet] JMIR Medical Informatics; 2020 [cited 2020 Nov 6];8(11):e21252. [doi: 10.2196/21252]

281. Bengio Y, LeCun Y. The Future is Self-Supervised [Internet]. 2020. Available from: https://iclr.cc/virtual_2020/speaker_7.html

282. Bender EM, Koller A. Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. ACL 2020; 2020.

283. Oxford Languages. Oxford Languages and Google - English [Internet]. Oxford Languages. 2020 [cited 2021 Jul 3]. Available from: https://languages.oup.com/google-dictionary-en/

284. Treato. Treato [Internet]. Treato. 2017 [cited 2017 Aug 17]. Available from: http://corp.treato.com/story

285. Post M. A Call for Clarity in Reporting BLEU Scores. Proceedings of the Third Conference on Machine Translation: Research Papers [Internet] Stroudsburg, PA, USA: Association for Computational Linguistics; 2018. [doi: 10.18653/v1/w18-6319]

286. Kumar V, Khosla C. Data Cleaning-A Thorough Analysis and Survey on Unstructured Data. 2018 8th International Conference on Cloud Computing, Data Science Engineering (Confluence) [Internet] ieeexplore.ieee.org; 2018. p. 305–309. [doi: 10.1109/CONFLUENCE.2018.8442950]

287. Abdellaoui R, Foulquie P, Texier N, Faviez C, Burgun A, Schuck S. Detection of Cases of Noncompliance to Drug Treatment in Patient Forum Posts: Topic Model Approach. J Med Internet Res [Internet] 2018;20(3). [doi: 10.2196/jmir.9222]

288. OpenRefine [Internet]. 2020. Available from: https://openrefine.org/

289. Schofield A, Thompson L, Mimno D. Quantifying the effects of text duplication on semantic models. Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing [Internet] 2017. p. 2737–2747. Available from: http://www.cs.cornell.edu/~laurejt/papers/text-duplication-2017.pdf

290. Duque T. Text Normalization - Towards Data Science [Internet]. Towards Data Science. 2020 [cited 2021 Apr 12]. Available from: https://towardsdatascience.com/text-normalization-7ecc8e084e31

291. Jurafsky D, Martin J. Spelling Correction and the Noisy Channel [Internet]. Speech and Language Processing. 2020. Available from: https://web.stanford.edu/~jurafsky/slp3/B.pdf

292. Lai KH, Topaz M, Goss FR, Zhou L. Automated misspelling detection and correction in clinical free-text records. J Biomed Inform [Internet] Elsevier BV; 2015 Jun;55:188–195. PMID:25917057

293. Effrosynidis D, Symeonidis S, Arampatzis A. A Comparison of Pre-processing Techniques for Twitter Sentiment Analysis. Research and Advanced Technology for Digital Libraries [Internet] Springer International Publishing; 2017. p. 394–406. [doi: 10.1007/978-3-319-67008-9_31]

294. Valdez D, Picket AC, Young B-R, Golden S. On Mining Words: The Utility of Topic Models in Health Education Research and Practice. Health Promot Pract [Internet] 2021 Mar 24;1524839921999050. PMID:33759597

295. Maier D, Waldherr A, Miltner P, Wiedemann G, Niekler A, Keinert A, Pfetsch B, Heyer G, Reber U, Häussler T, Schmid-Petri H, Adam S. Applying LDA Topic Modeling in Communication Research: Toward a Valid and Reliable Methodology. Commun Methods Meas [Internet] Routledge; 2018 Apr 3;12(2–3):93–118. [doi: 10.1080/19312458.2018.1430754]

296. Albalawi R, Yeap TH, Benyoucef M. Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis. Frontiers Artificial Intelligence Appl [Internet] 2020;3:42. [doi: 10.3389/frai.2020.00042]

297. Mimno D. Grounded theory and topic modelling: extreme divergence or unlikely convergence [Internet]. Turing Institute; 2017 [cited 2020 Jul 26]. Available from: https://www.youtube.com/watch?v=fkzKfK2eEqs&list=PLuD_SqLtxSdV1ns23klRRLJzyN5q5l91U&index=11

298. Greene D. Topic Modelling with Scikit-learn [Internet]. 2017. Available from: https://github.com/derekgreene/topic-model-tutorial/blob/master/topic-modelling-with-scikitlearn.pdf

299. Blei DM. Probabilistic topic models. Commun ACM [Internet] Association for Computing Machinery (ACM); 2012 Apr;55(4):77–84. [doi: 10.1145/2133806.2133826]

300. Chen Y, Zhang H, Liu R, Ye Z, Lin J. Experimental explorations on short text topic mining between LDA and NMF based Schemes. Knowledge-Based Systems [Internet] 2019 Jan 1;163:1–13. [doi: 10.1016/j.knosys.2018.08.011]

301. Jang H, Rempel E, Roth D, Carenini G, Janjua NZ. Tracking COVID-19 Discourse on Twitter in North America: Infodemiology Study Using Topic Modeling and Aspect-Based Sentiment Analysis. J Med Internet Res [Internet] 2021 Feb 10;23(2):e25431. PMID:33497352

302. Suri P, Roy NR. Comparison between LDA NMF for event-detection from large text stream data. 2017 3rd International Conference on Computational Intelligence Communication Technology (CICT) [Internet] ieeexplore.ieee.org; 2017. p. 1–5. [doi: 10.1109/CIACT.2017.7977281]

303. Birks D, Coleman A, Jackson D. Unsupervised identification of crime problems from police free-text data. Crime Science [Internet] 2020 Oct 7;9(1):18. [doi: 10.1186/s40163-020-00127-4]

304. Ghosh D, Guha R. What are we 'tweeting' about obesity? Mapping tweets with topic modeling and Geographic Information System. Cartogr Geogr Inf Sci [Internet] 2013;40(2):90–102. [doi: 10.1080/15230406.2013.776210]

305. Schofield A, Magnusson M, Mimno D. Pulling out the stops: Rethinking stopword removal for topic models. Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers [Internet] 2017. p. 432–436. Available from: https://www.aclweb.org/anthology/E17-2069.pdf

306. Liu Y, Yin Z. Understanding weight loss via online discussions: Content analysis of Reddit posts using topic modeling and word clustering techniques. Journal of Medical Internet Research Vol 22(6), 2020, ArtID e13745 [Internet] Journal of Medical Internet Research; 2020 Jun;22(6):e13745. [doi: 10.2196/13745]

307. Rehurek R. gensim: topic modelling for humans [Internet]. gensim. 2019 [cited 2020 Jun 23]. Available from: https://radimrehurek.com/gensim/auto_examples/tutorials/run_lda.html

308. Prabhakaran S. Topic Modeling in Python with Gensim [Internet]. Machine Learning Plus. 2018 [cited 2020 Jun 14]. Available from: https://www.machinelearningplus.com/nlp/topic-modeling-gensim-python/

309. Rehurek R, Sojka P. Gensim-statistical semantics in python. NLP Centre, Faculty of Informatics, Masaryk University 2011;

310. Mabey B. pyLDAvis [Internet]. Github; 2018 [cited 2021 May 27]. Available from: https://github.com/bmabey/pyLDAvis/issues/79

311. Sievert C, Shirley K. LDAvis: A method for visualizing and interpreting topics. Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces [Internet] Baltimore, Maryland, USA: Association for Computational Linguistics; 2014. p. 63–70. [doi: 10.3115/v1/W14-3110]

312. 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. Scikit-learn: Machine Learning in Python. J Mach Learn Res [Internet] 2011 [cited 2020 Jul 13];12(85):2825–2830. Available from: http://jmlr.csail.mit.edu/papers/v12/pedregosa11a.html

313. Bakharia A. Topic Modeling with Scikit Learn [Internet]. Medium. ML Review; 2016 [cited 2020 Mar 7]. Available from: https://medium.com/mlreview/topic-modeling-with-scikit-learn-e80d33668730

314. Alexander E, Gleicher M. Task-Driven Comparison of Topic Models. IEEE Trans Vis Comput Graph [Internet] 2016 Jan;22(1):320–329. PMID:26340780

315. Chen X, Vorvoreanu M, Madhavan K. Mining social media data for understanding students' learning experiences. IEEE Trans Learn Technol [Internet] IEEE;

2014;7(3):246–259. Available from: https://ieeexplore.ieee.org/abstract/document/6697807/

316. Soltoff B. Topic modeling | Computing for the Social Sciences [Internet]. Computing for the Social Sciences. 2019 [cited 2020 Jun 21]. Available from: https://cfss.uchicago.edu/notes/topic-modeling/

317. O'Callaghan D, Greene D, Carthy J, Cunningham P. An analysis of the coherence of descriptors in topic modeling. Expert Syst Appl [Internet] 2015 Aug 1;42(13):5645–5657. [doi: 10.1016/j.eswa.2015.02.055]

318. Angelov D. Top2Vec: Distributed Representations of Topics [Internet]. arXiv [csCL]. 2020. Available from: http://arxiv.org/abs/2008.09470

319. Angelov D. Top2Vec documentation [Internet]. 2021 Jan. Available from: https://top2vec.readthedocs.io/_/downloads/en/stable/pdf/

320. enTenTen — English corpus from the web | Sketch Engine [Internet]. 2021 [cited 2018 Feb 27]. Available from: https://www.sketchengine.co.uk/ententen-english-corpus/

321. Hutto CJ, Gilbert E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. Eighth international AAAI conference on weblogs and social media [Internet] 2014. Available from: https://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/viewPaper/8109

322. Social Media Research, Group. Using social media for social research: An introduction [Internet]. Government Social Research; 2016 May p. 25. Available from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/524750/GSR_Social_Media_Research_Guidance_-_Using_social_media_for_social_research.pdf

323. Soma J. Comparing sentiment analysis tools [Internet]. investigate.ai. 2021 [cited 2021 Jun 30]. Available from: https://investigate.ai/investigating-sentiment-analysis/comparing-sentiment-analysis-tools/

324. Carrillo-de-Albornoz J, Rodríguez Vidal J, Plaza L. Feature engineering for sentiment analysis in e-health forums. PLoS One [Internet] United States; 2018 Nov 29;13(11):e0207996. PMID:30496232

325. Bonta V, Janardhan NKN. A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis. Asian Journal of Computer Science and Technology [Internet] researchgate.net; 2019;8(S2):1–6. Available from: https://www.researchgate.net/profile/Janardhan_n/publication/333602124_A_Comprehensive_Study_on_Lexicon_Based_Approaches_for_Sentiment_Analysis/links/5d1346ce299bf1547c7f931a/A-Comprehensive-Study-on-Lexicon-Based-Approaches-for-Sentiment-Analysis.pdf

326. Caren N. Word Lists and Sentiment Analysis [Internet]. nealcaren.org. 2019 [cited 2021 Jul 1]. Available from: https://nealcaren.org/lessons/wordlists/

327. Yang J, Han SC, Poon J. A Survey on Extraction of Causal Relations from Natural Language Text [Internet]. arXiv [csIR]. 2021. [doi: 10.18653/v1/P19-1024]

328. Sorgente A, Vettigli G, Mele F. A Hybrid Approach for the Automatic Extraction of Causal Relations from Text. In: Lai C, Giuliani A, Semeraro G, editors. Emerging Ideas on Information Filtering and Retrieval: DART 2013: Revised and Invited Papers [Internet] Cham: Springer International Publishing; 2018. p. 15–29. [doi: 10.1007/978-3-319-68392-8_2]

329. Metrics For Reddit. Complete List Of Subreddits [Internet]. Reddit Metrics. 2021 [cited 2021 May 3]. Available from: https://frontpagemetrics.com/list-all-subreddits

330. Amaya A, Bach R, Keusch F, Kreuter F. New Data Sources in Social Science Research: Things to Know Before Working With Reddit Data. Soc Sci Comput Rev [Internet] SAGE Publications Inc; 2019 Dec 18;0894439319893305. [doi: 10.1177/0894439319893305]

331. Tangherlini TR, Roychowdhury V, Glenn B, Crespi CM, Bandari R, Wadia A, Falahi M, Ebrahimzadeh E, Bastani R. 'Mommy Blogs' and the Vaccination Exemption Narrative: Results From A Machine-Learning Approach for Story Aggregation on Parenting Social Media Sites. JMIR Public Health Surveill [Internet] 2016 Nov 22;2(2):e166. PMID:27876690

332. Nokel M, Loukachevitch N. Accounting ngrams and multi-word terms can improve topic models. Proceedings of the 12th Workshop on Multiword Expressions [Internet] Berlin, Germany: Association for Computational Linguistics; 2016. p. 44–49. [doi: 10.18653/v1/W16-1806]

333. NICE. NICE Evidence Search | modafinil [Internet]. 2016 [cited 2016 Nov 21]. Available from: https://bnf.nice.org.uk/drug/modafinil.html

334. Wagland R, Recio-Saucedo A, Simon M, Bracher M, Hunt K, Foster C, Downing A, Glaser A, Corner J. Development and testing of a text-mining approach to analyse patients' comments on their experiences of colorectal cancer care. BMJ Qual Saf [Internet] qualitysafety.bmj.com; 2015 Oct 28; PMID:26512131

335. Isoaho K, Gritsenko D, Mäkelä E. Topic modeling and text analysis for qualitative policy research. Policy Stud J [Internet] Wiley; 2021 Feb;49(1):300–324. [doi: 10.1111/psj.12343]

336. Lin C, Bethard S, Dligach D, Sadeque F, Savova G, Miller TA. Does BERT need domain adaptation for clinical negation detection? J Am Med Inform Assoc [Internet] 2020 Apr 1;27(4):584–591. PMID:32044989

337. Pröllochs N, Feuerriegel S, Lutz B, Neumann D. Negation scope detection for sentiment analysis: A reinforcement learning framework for replicating human interpretations. Inf Sci [Internet] 2020 Oct 1;536:205–221. [doi: 10.1016/j.ins.2020.05.022]

338. Alper BS, Fedorowicz Z, van Zuuren EJ. Limitations in conduct and reporting of cochrane reviews rarely inhibit the determination of the validity of evidence for clinical decision-making. J Evid Based Med [Internet] 2015 Aug 1;8(3):154–160. [doi: 10.1111/jebm.12161]

339. Cochrane Collaboration. MECIR Manual: Authors' conclusions (R102-R103) [Internet]. Cochrane. 2021 [cited 2021 Jun 14]. Available from: https://community.cochrane.org/mecir-manual/standards-reporting-new-cochrane-intervention-reviews-r1-r109/results-r56-r109/authors-conclusions-r102-r103

340. Ortiz-Orendain J, Covarrubias-Castillo SA, Vazquez-Alvarez AO, Castiello-de Obeso S, Arias Quiñones GE, Seegers M, Colunga-Lozano LE. Modafinil for people with schizophrenia or related disorders. Cochrane Database Syst Rev [Internet] 2019 Dec 12;12:CD008661. PMID:31828767

341. Castells X, Blanco-Silvente L, Cunill R. Amphetamines for attention deficit hyperactivity disorder (ADHD) in adults. Cochrane Database Syst Rev [Internet] 2018 Aug 9;8:CD007813. PMID:30091808

342. Ruthirakuhan MT, Herrmann N, Abraham EH, Chan S, Lanctôt KL. Pharmacological interventions for apathy in Alzheimer's disease. Cochrane Database Syst Rev [Internet] John Wiley & Sons, Ltd; 2018 May 4 [cited 2021 May 22];5(5):CD012197. PMID:29727467

343. Gibbons C, Pagnini F, Friede T, Young CA. Treatment of fatigue in amyotrophic lateral sclerosis/motor neuron disease. Cochrane Database Syst Rev [Internet] 2018 Jan 2;1:CD011005. PMID:29293261

344. Day J, Yust-Katz S, Cachia D, Wefel J, Katz LH, Tremont I, Bulbeck H, Armstrong T, Rooney AG. Interventions for the management of fatigue in adults with a primary brain tumour. Cochrane Database Syst Rev [Internet] 2016 Apr 13;4:CD011376. PMID:27074263

345. Pérez-Mañá C, Castells X, Torrens M, Capellà D, Farre M. Efficacy of psychostimulant drugs for amphetamine abuse or dependence. Cochrane Database Syst Rev [Internet] 2013 Sep 2;(9):CD009695. PMID:23996457

346. Koopman FS, Uegaki K, Gilhus NE, Beelen A, de Visser M, Nollet F. Treatment for postpolio syndrome. Cochrane Database Syst Rev [Internet] 2011 Feb 16;(2):CD007818. PMID:21328301

347. Ker K, Edwards PJ, Felix LM, Blackhall K, Roberts I. Caffeine for the prevention of injuries and errors in shift workers. Cochrane Database Syst Rev [Internet] 2010 May 12;(5):CD008508. PMID:20464765

348. Candy M, Jones L, Williams R, Tookman A, King M. Psychostimulants for depression. Cochrane Database Syst Rev [Internet] 2008 Apr 16;(2):CD006722. PMID:18425966

349. Annane D, Moore DH, Barnes PRJ, Miller RG. Psychostimulants for hypersomnia (excessive daytime sleepiness) in myotonic dystrophy. Cochrane Database Syst Rev [Internet] 2006 Jul 19;(3):CD003218. PMID:16855999

350. NICE. modafinil - UK Database of Uncertainties about the Effects of Treatments - UK DUETS | Evidence search | NICE [Internet]. 2014 [cited 2019 Jan 23]. Available from: https://www.evidence.nhs.uk/search?om=[{%22ety%22:[%22Evidence%20Uncertainties%22]},{%22srn%22:[%22UK%20Database%20of%20Uncertainties%20about%20the%20Effects%20of%20Treatments%20-%20UK%20DUETS%22]}]&q=modafinil&sp=on

351. Ballon JS, Feifel D, Ballon JS, Feifel D. A systematic review of modafinil: Potential clinical uses and mechanisms of action. Psychiatry: Interpersonal and Biological Processes [Internet] 2006 Apr;67(4):554–566. PMID:16669720

352. Sheng P, Hou L, Wang X, Wang X, Huang C, Yu M, Han X, Dong Y, Sheng P, Hou L, Wang X, Wang X, Huang C, Yu M, Han X, Dong Y. Efficacy of modafinil on fatigue and

excessive daytime sleepiness associated with neurological disorders: a systematic review and meta-analysis. PLoS ONE [Electronic Resource] [Internet] Public Library of Science; 2013 Jan 3;8(12):e81802. PMID:24312590

353. Shaffer JA, Falzon L, Cheung K, Davidson KW. N-of-1 randomized trials for psychological and health behavior outcomes: a systematic review protocol. Syst Rev [Internet] 2015 Jun 17;4:87. PMID:26081256

354. Sheridan DJ, Julian DG. Achievements and Limitations of Evidence-Based Medicine. J Am Coll Cardiol [Internet] 2016 Jul 12;68(2):204–213. PMID:27386775

355. Greenhalgh T, Thorne S, Malterud K. Time to challenge the spurious hierarchy of systematic over narrative reviews? Eur J Clin Invest [Internet] 2018 Jun;48(6):e12931. PMID:29578574

356. Chapman S. The problem with sex: is our reluctance to talk about it harming patients? [Internet]. Evidently Cochrane. 2017 [cited 2017 Apr 6]. Available from: http://www.evidentlycochrane.net/problem-with-sex-reluctance-talk-harming-patients/

357. Lapsley P. Lessons from patients' journeys. BMJ [Internet] 2013 Apr 3;346:f1988. PMID:23553972

358. Karmakar D, Hayward L. What can we learn from the vaginal mesh story? Climacteric [Internet] Taylor & Francis; 2019 May 4;22(3):277–282. PMID:30829077

359. Hengartner MP, Plöderl M. Estimates of the minimal important difference to evaluate the clinical significance of antidepressants in the acute treatment of moderate-to-severe depression. BMJ Evid Based Med [Internet] 2021 Feb 16; PMID:33593736

360. Tapi Nzali MD, Bringay S, Lavergne C, Mollevi C, Opitz T. What Patients Can Tell Us: Topic Analysis for Social Media on Breast Cancer. JMIR Med Inform [Internet] 2017 Jul 31;5(3):e23. PMID:28760725

361. Jaidka K, Giorgi S, Schwartz HA, Kern ML, Ungar LH, Eichstaedt JC. Estimating geographic subjective well-being from Twitter: A comparison of dictionary and data-driven language methods. Proc Natl Acad Sci U S A [Internet] 2020 May 12;117(19):10165–10171. PMID:32341156

362. Segura-Bedmar I, Martínez P. Pharmacovigilance through the development of text mining and natural language processing techniques. J Biomed Inform [Internet] dl.acm.org; 2015 Dec;58:288–291. PMID:26547007

363. Yang M, Kiang M, Shang W. Filtering big data from social media – Building an early warning system for adverse drug reactions. J Biomed Inform [Internet] 2015 Apr;54:230–240. PMID:25688695

364. Doan S, Yang EW, Tilak SS, Li PW, Zisook DS, Torii M. Extracting health-related causality from twitter messages using natural language processing. BMC Med Inform Decis Mak [Internet] 2019 Apr 4;19(Suppl 3):79. PMID:30943954

365. Wicks P, Mack Thorley E, Simacek K, Curran C, Emmas C. Scaling PatientsLikeMe via a 'Generalized Platform' for Members with Chronic Illness: Web-Based Survey Study of

Benefits Arising. J Med Internet Res [Internet] 2018 May 7;20(5):e175. PMID:29735472

366. Levine DM, Mehrotra A. Assessment of Diagnosis and Triage in Validated Case Vignettes Among Nonphysicians Before and After Internet Search. JAMA Netw Open [Internet] 2021 Mar 1;4(3):e213287. PMID:33779741

367. Hibbard JH, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. Health Aff [Internet] Health Affairs; 2013 Feb;32(2):207–214. PMID:23381511

368. Whitman CB, Reid MW, Arnold C, Patel H, Ursos L, Sa'adon R, Pourmorady J, Spiegel BMR. Balancing opioid-induced gastrointestinal side effects with pain management: Insights from the online community. J Opioid Manag [Internet] europepmc.org; 2015 Sep 1 [cited 2017 May 3];11(5):383–391. PMID:26535966

369. Greenhalgh T, Knight M, A'Court C, Buxton M, Husain L. Management of post-acute covid-19 in primary care. BMJ [Internet] 2020 Aug 11;370:m3026. PMID:32784198

370. van Mierlo T. The 1% rule in four digital health social networks: an observational study. J Med Internet Res [Internet] jmir.org; 2014 Feb 4;16(2):e33. PMID:24496109

371. Dalmer NK. Questioning reliability assessments of health information on social media. J Med Libr Assoc [Internet] 2017 Jan;105(1):61–68. PMID:28096748

372. NHS PrescQIPP. Modafinil: Restrictions on indications and evidence for off-label uses. 2015 Oct. Report No.: Bulletin 113.

373. i2b2: Informatics for Integrating Biology & the Bedside [Internet]. [cited 2021 Jul 6]. Available from: https://www.i2b2.org/

374. MIMIC II Databases [Internet]. [cited 2021 Jul 6]. Available from: https://archive.physionet.org/mimic2/

375. Kormilitzin A, Vaci N, Liu Q, Nevado-Holgado A. Med7: a transferable clinical natural language processing model for electronic health records [Internet]. arXiv [csCL]. 2020. Available from: http://arxiv.org/abs/2003.01271

376. Jackson R, Kartoglu I, Stringer C, Gorrell G, Roberts A, Song X, Wu H, Agrawal A, Lui K, Groza T, Lewsley D, Northwood D, Folarin A, Stewart R, Dobson R. CogStack - experiences of deploying integrated information retrieval and extraction services in a large National Health Service Foundation Trust hospital. BMC Med Inform Decis Mak [Internet] 2018 Jun 25;18(1):47. PMID:29941004

Appendices

Appendix A Ethics Approval

WARWICK
THE UNIVERSITY OF WARWICK

Biomedical and Scientific Research Ethics Committee
Kirby Corner Road
Coventry
CV4 8UW

Tuesday, 29 October 2019

Julia Walsh
WMS
University of Warwick
Coventry
CV4 7AL

Dear Julia,

Ethical Application Reference: BSREC 11/19-20
Title: Using spontaneously generated online patient experiences to improve healthcare

Thank you for submitting your ethics application to the Biomedical and Scientific Research Ethics Committee (BSREC) for consideration. We are pleased to advise you that, under the authority delegated to us by the University of Warwick Research Governance and Ethics Committee, **full approval for your project is hereby granted.**

Before conducting your research it is strongly recommended that you complete the on-line Research Integrity training:
www.warwick.ac.uk/ritraining. Support is available from the BSREC Secretary.

In undertaking your study, you are required to comply with the University of Warwick's Research Code of Practice:
https://warwick.ac.uk/services/ris/research_integrity/code_of_practice_and_policies/research_code_of_practice/

You are also required to familiarise yourself with the University of Warwick's Code of Practice for the Investigation of Research Misconduct:
https://warwick.ac.uk/services/ris/research_integrity/research_misconduct/codeofpractice_researchmisconduct/

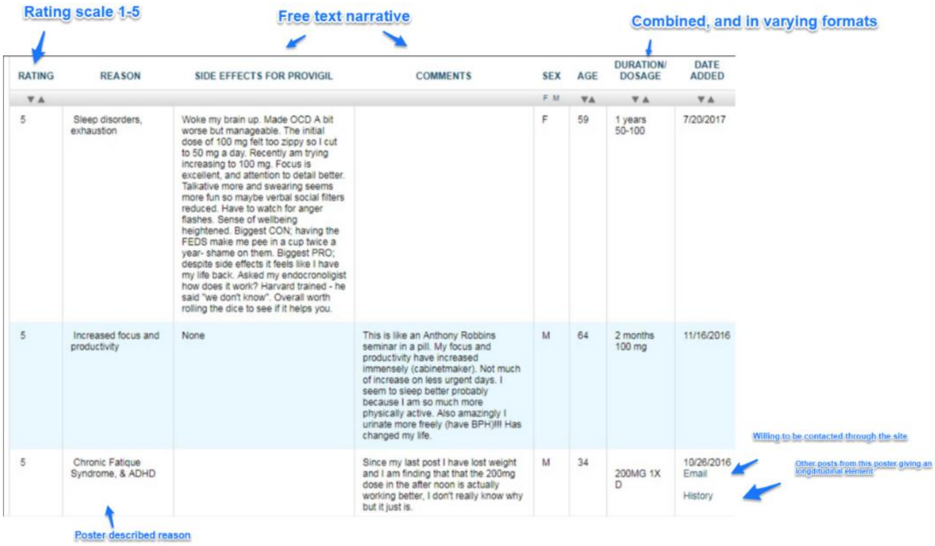
You must ensure that you are compliant with all necessary data protection regulations:
https://warwick.ac.uk/services/idc

Please ensure that evidence of all necessary local permissions is provided to BSREC prior to commencing your study.

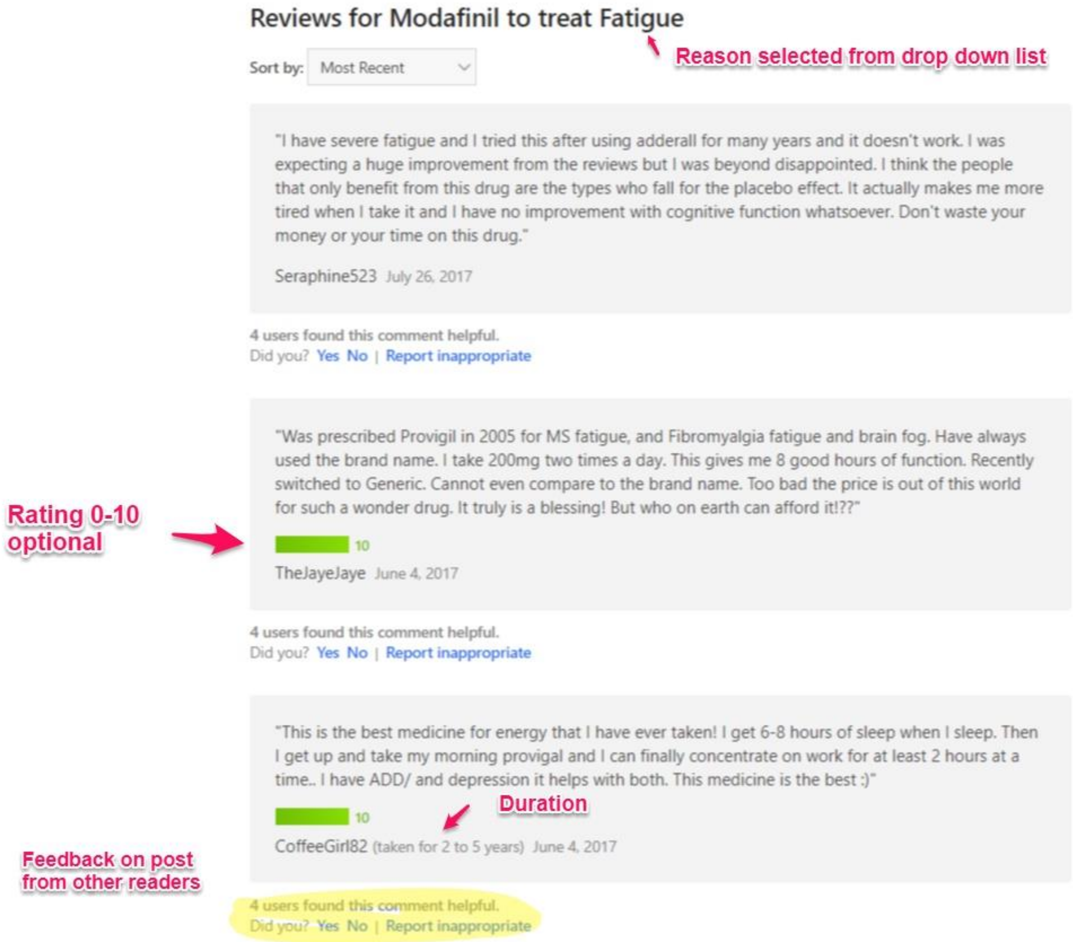
Please also be aware that BSREC grants **ethical approval** for studies. The seeking and obtaining of all other necessary approvals is the responsibility of the investigator.

www.warwick.ac.uk

Appendix B SGOPE example screenshots



RATING	REASON	SIDE EFFECTS FOR PROVIGIL	COMMENTS	SEX	AGE	DURATION/ DOSAGE	DATE ADDED
5	Sleep disorders, exhaustion	Woke my brain up. Made OCD A bit worse but manageable. The initial dose of 100 mg felt too zippy so I cut to 50 mg a day. Recently am trying increasing to 100 mg. Focus is excellent, and attention to detail better. Talkative more and swearing seems more fun so maybe verbal social filters reduced. Have to watch for anger flashes. Sense of wellbeing heightened. Biggest CON; having the FEDS make me pee in a cup twice a year- shame on them. Biggest PRO; despite side effects it feels like I have my life back. Asked my endocronoligist how does it work? Harvard trained - he said "we don't know". Overall worth rolling the dice to see if it helps you.		F	59	1 years 50-100	7/20/2017
5	Increased focus and productivity	None	This is like an Anthony Robbins seminar in a pill. My focus and productivity have increased immensely (cabinetmaker). Not much of increase on less urgent days. I seem to sleep better probably because I am so much more physically active. Also amazingly I urinate more freely (have BPH)!!! Has changed my life.	M	64	2 months 100 mg	11/16/2016
5	Chronic Fatique Syndrome, & ADHD		Since my last post I have lost weight and I am finding that that the 200mg dose in the after noon is actually working better, I don't really know why but it just is.	M	34	200MG 1X D	10/26/2016 Email History



Appendix C P1: Cleaned Data Structure

Field Name	Field Type	Description
DocumentName	Text	Allocated post ID
ConditionPostedAs	Text	Thread or forum that post was made on
ConditionRefined	Text	Condition/s mentioned in post
Author	Text	Author user handle – deleted after duplication checks
AgeGroup	Text	Age standardised to decades
Gender	Text	
Duration	Text	How long the poster had been taking Modafinil for
Final10	Integer	Standardised numeric rating as entered by poster
Text	Text	Narrative content of post
PostDate	Date	Date post made – deleted after duplication checks
PostYear	Integer	Year that post was made
DosageText	Text	Any text relating to dosage
DosageAmt	Text	Amount of dosage if stated
DosageFreq	Text	Frequency of dosage if stated
OtherDrugsTaken	Text	Any other drugs or interventions tried

Appendix D ICSD3 Sleep Disorder Classifications

Classified As	Name
Insomnia	Chronic insomnia disorder
	Short-term insomnia disorder
	Other insomnia disorder
Sleeping related breathing disorders	OSA disorders
	OSA, adult
	OSA, pediatric
	Central sleep apnea syndromes
	Central sleep apnea with Cheyne-Stokes breathing
	Central sleep apnea due to a medical disorder without Cheyne-Stokes breathing
	Central sleep apnea due to high altitude periodic breathing
	Central sleep apnea due to a medication or substance
	Primary central sleep apnea
	Primary central sleep apnea of infancy
	Primary central sleep apnea of prematurity
	Treatment-emergent central sleep apnea
	Sleep-related hypoventilation disorders
	Obesity hypoventilation syndrome
	Congenital central alveolar hypoventilation syndrome
	Late-onset central hypoventilation with hypothalamic dysfunction
	Idiopathic central alveolar hypoventilation
	Sleep-related hypoventilation due to a medication or substance
	Sleep-related hypoventilation due to a medical disorder
	Sleep-related hypoxemia disorder
Central Disorders of Hypersomnolence	Narcolepsy type 1
	Narcolepsy type 2
	Idiopathic hypersomnia
	Kleine-Levin syndrome
	Hypersomnia due to a medical disorder
	Hypersomnia due to a medication or substance
	Hypersomnia associated with a psychiatric disorder
	Insufficient sleep syndrome

Circadian Rhythm Sleep-Wake Disorders	Delayed sleep-wake phase disorder
	Advanced sleep-wake phase disorder
	Irregular sleep-wake rhythm disorder
	Non-24-h sleep-wake rhythm disorder
	Shift work disorder
	Jet lag disorder
	Circadian sleep-wake disorder not otherwise specified
Parasomnias	NREM-related parasomnias
	Confusional arousals
	Sleepwalking
	Sleep terrors
	Sleep-related eating disorder
	REM-related parasomnias
	REM sleep behavior disorder
	Recurrent isolated sleep paralysis
	Nightmare disorder
	Other parasomnias
	Exploding head syndrome
	Sleep-related hallucinations
	Sleep enuresis
	Parasomnia due to a medical disorder
	Parasomnia due to a medication or substance
	Parasomnia, unspecified
Sleep related movement disorders	Restless legs syndrome
	Periodic limb movement disorder
	Sleep-related leg cramps
	Sleep-related bruxism
	Sleep-related rhythmic movement disorder
	Benign sleep myoclonus of infancy
	Propriospinal myoclonus at sleep onset
	Sleep-related movement disorder due to a medical disorder
	Sleep-related movement disorder due to a medication or substance
	Sleep-related movement disorder, unspecified

Appendix E Sketch POS tag list

POS Tag	Description	Example
CC	coordinating conjunction	and
CD	cardinal number	1, one
CDZ	possesive pronoun	one's
DT	determiner	the
EX	existential there	there is
FW	foreign word	d'hoevre
IN	preposition, subordinating conjunction	in, of, like
IN/that	that as subordinator	that
JJ	adjective	green
JJR	adjective, comparative	greener
JJS	adjective, superlative	greenest
LS	list marker	1)
MD	modal	could, will
NN	noun, singular or mass	table
NNS	noun plural	tables
NNSZ	possessive noun plural	people's, women's
NNZ	possessive noun, singular or mass	year's, world's
NP	proper noun, singular	John
NPS	proper noun, plural	Vikings
NPSZ	possesive proper noun, plural	Boys', Workers'
NPZ	possesive noun, singular	Britain's, God's
PDT	predeterminer	both the boys
PP	personal pronoun	I, he, it
PPZ	possessive pronoun	my, his
RB	adverb	however, usually, naturally, here, good
RBR	adverb, comparative	better
RBS	adverb, superlative	best
RP	particle	give up
SENT	Sentence-break punctuation	. ! ?
SYM	Symbol	/ [= *
TO	infinitive 'to'	togo
UH	interjection	uhhuhhuhh
VB	verb be, base form	be
VBD	verb be, past tense	was, were
VBG	verb be, gerund/present participle	being
VBN	verb be, past participle	been
VBP	verb be, present, non-3d person	am, are
VBZ	verb be, 3rd person sing. present	is
VH	verb have, base form	have
VHD	verb have, past tense	had
VHG	verb have, gerund/present participle	having
VHN	verb have, past participle	had

POS Tag	Description	Example
VHP	verb have, sing. present, non-3d	have
VHZ	verb have, 3rd person sing. present	has
VV	verb, base form	take
VVD	verb, past tense	took
VVG	verb, gerund/present participle	taking
VVN	verb, past participle	taken
VVP	verb, present, not 3rd person	take
VVZ	verb, 3rd person sing. present	takes
WDT	wh-determiner	which
WP	wh-pronoun	who, what
WPZ	possessive wh-pronoun	whose
WRB	wh-abverb	where, when
Z	possessive ending	‘s

Appendix F P1: Sentiment evaluation agreement

DocName	Manual		TextBlob	VaderExt
AAPNuv006	Positive		TBpositive	Vnegative
AAPNuv009	Mixed		TBpositive	Vpositive
AAPNuv035	Positive		TBnegative	Vnegative
AAPNuv045	Negative		TBpositive	Vnegative
AAPNuv050	Neutral		TBneutral	Vnegative
AAPNuv063	Positive		TBpositive	Vpositive
AAPNuv086	Positive		TBpositive	Vnegative
AAPNuv090	Positive		TBpositive	Vnegative
AAPNuv095	Positive		TBnegative	Vnegative
AAPNuv100	Positive		TBnegative	Vnegative
AAPNuv102	Positive		TBnegative	Vpositive
AAPNuv107	Positive		TBpositive	Vnegative
AAPNuv123	Negative		TBnegative	Vnegative
AAPNuv133	Negative		TBpositive	Vpositive
AAPNuv138	Negative		TBpositive	Vnegative
AAPNuv146	Negative		TBpositive	Vpositive
AAPNuv153	Mixed		TBpositive	Vnegative
AAPNuv157	Negative		TBnegative	Vnegative
AAPProv011	Positive		TBnegative	Vnegative
AAPProv014	Positive		TBpositive	Vpositive
AAPProv022	Positive		TBpositive	Vnegative
AAPProv034	Positive		TBnegative	Vnegative
AAPProv038	Mixed		TBnegative	Vnegative
AAPProv065	Positive		TBpositive	Vpositive
AAPProv078	Positive		TBnegative	Vnegative
AAPProv102	Positive		TBpositive	Vpositive
AAPProv117	Mixed		TBpositive	Vnegative

DocName	Manual		TextBlob	VaderExt
AAPProv118	Neutral		TBnegative	Vnegative
AAPProv136	Positive		TBpositive	Vpositive
AAPProv142	Positive		TBnegative	Vnegative
AAPProv143	Positive		TBpositive	Vpositive
AAPProv149	Mixed		TBnegative	Vnegative
AAPProv151	Mixed		TBpositive	Vpositive
AAPProv164	Positive		TBnegative	Vpositive
AAPProv167	Mixed		TBpositive	Vnegative
AAPProv170	Positive		TBpositive	Vpositive
AAPProv179	Negative		TBneutral	Vnegative
AAPProv195	Positive		TBpositive	Vnegative
AAPProv207	Positive		TBpositive	Vpositive
AAPProv218	Mixed		TBpositive	Vnegative
AAPProv222	Mixed		TBpositive	Vnegative
AAPProv241	Positive		TBnegative	Vpositive
AAPProv245	Mixed		TBnegative	Vnegative
AAPProv260	Negative		TBpositive	Vpositive
AAPProv291	Negative		TBnegative	Vnegative
AAPProv293	Mixed		TBpositive	Vnegative
AAPProv297	Positive		TBpositive	Vpositive
AAPProv311	Mixed		TBpositive	Vnegative
AAPProv318	Mixed		TBpositive	Vpositive
AAPProv349	Negative		TBnegative	Vnegative
AAPProv353	Negative		TBnegative	Vnegative
AAPProv363	Negative		TBpositive	Vnegative
AAPProv380	Positive		TBpositive	Vpositive
AAPProv388	Positive		TBpositive	Vpositive

DocName	Manual		TextBlob	VaderExt
AAPProv392	Positive		TBpositive	Vpositive
AAPProv395	Positive		TBpositive	Vpositive
AAPProv402	Negative		TBpositive	Vpositive
AAPProv404	Negative		TBnegative	Vnegative
AAPProv421	Positive		TBpositive	Vpositive
AAPProv441	Positive		TBpositive	Vpositive
AAPProv449	Negative		TBnegative	Vnegative
AAPProv460	Negative		TBnegative	Vnegative
AAPProv479	Mixed		TBpositive	Vnegative
AAPProv480	Mixed		TBpositive	Vpositive
AAPProv483	Negative		TBpositive	Vnegative
AAPProv491	Mixed		TBnegative	Vnegative
AAPProv502	Positive		TBpositive	Vnegative
AAPProv506	Mixed		TBnegative	Vnegative
AAPProv507	Mixed		TBpositive	Vnegative
AAPProv514	Positive		TBpositive	Vpositive
AAPProv516	Mixed		TBnegative	Vnegative
AAPProv522	Mixed		TBpositive	Vnegative
AAPProv524	Negative		TBpositive	Vpositive
AAPProv539	Positive		TBpositive	Vpositive
AAPProv567	Mixed		TBnegative	Vnegative
AAPProv573	Negative		TBnegative	Vnegative
AAPProv582	Positive		TBpositive	Vpositive
AAPProv597	Positive		TBpositive	Vpositive
AAPProv604	Negative		TBnegative	Vnegative
DCUR001	Positive		TBpositive	Vpositive
DCUR015	Positive		TBpositive	Vpositive
DCUR016	Positive		TBpositive	Vpositive
DCUR020	Positive		TBpositive	Vpositive

DocName	Manual		TextBlob	VaderExt
DCUR021	Mixed		TBnegative	Vnegative
DCUR028	Mixed		TBnegative	Vnegative
DCUR031	Positive		TBpositive	Vpositive
DCUR034	Positive		TBpositive	Vpositive
DCUR043	Positive		TBpositive	Vnegative
DCUR048	Positive		TBpositive	Vpositive
DCUR051	Positive		TBpositive	Vpositive
DCUR053	Positive		TBpositive	Vnegative
DCUR061	Negative		TBnegative	Vnegative
DCUR066	Positive		TBpositive	Vpositive
DCUR078	Positive		TBpositive	Vnegative
DCUR085	Mixed		TBnegative	Vnegative
DCUR090	Negative		TBnegative	Vnegative
DCUR093	Positive		TBpositive	Vpositive
DCUR094	Mixed		TBpositive	Vnegative
DCUR098	Positive		TBpositive	Vnegative
DCUR100	Positive		TBpositive	Vpositive
DCUR102	Negative		TBpositive	Vnegative
DCUR103	Positive		TBpositive	Vpositive
DCUR113	Positive		TBpositive	Vpositive
DCUR116	Positive		TBpositive	Vpositive
DCUR121	Positive		TBpositive	Vpositive
DCUR130	Positive		TBpositive	Vpositive
DCUR135	Negative		TBpositive	Vnegative
DCUR145	Positive		TBpositive	Vnegative
DCUR149	Positive		TBpositive	Vpositive
DCUR261	Neutral		TBpositive	Vnegative
DCUR265	Positive		TBnegative	Vnegative
DCUR266	Positive		TBpositive	Vnegative

DocName	Manual		TextBlob	VaderExt
DCUR268	Mixed		TBpositive	Vpositive
DCUR278	Positive		TBpositive	Vpositive
DCUR294	Mixed		TBpositive	Vpositive
DCUR295	Positive		TBpositive	Vpositive
DCUR305	Positive		TBnegative	Vpositive
DCUR310	Positive		TBpositive	Vpositive
DCUR324	Positive		TBpositive	Vnegative
DCUR329	Positive		TBpositive	Vnegative
DCUR330	Positive		TBnegative	Vnegative
DCUR337	Positive		TBpositive	Vpositive
DCUR344	Positive		TBpositive	Vpositive
DCUR348	Positive		TBpositive	Vpositive
DCUR358	Positive		TBpositive	Vneutral
DCUR366	Negative		TBnegative	Vnegative
DCUR368	Mixed		TBneutral	Vnegative
DCUR383	Positive		TBnegative	Vpositive
DCUR394	Negative		TBnegative	Vnegative
DCUR400	Positive		TBpositive	Vpositive
DCUR402	Negative		TBnegative	Vpositive
DCUR411	Positive		TBpositive	Vpositive
E01	Negative		TBpositive	Vnegative
E02	Positive		TBpositive	Vpositive
E03	Positive		TBpositive	Vpositive
E04	Positive		TBpositive	Vnegative
E05	Positive		TBpositive	Vpositive
E06	Positive		TBpositive	Vpositive
E07	Positive		TBpositive	Vpositive
E08	Positive		TBpositive	Vnegative
E09	Positive		TBpositive	Vnegative

DocName	Manual		TextBlob	VaderExt
E10	Positive		TBpositive	Vpositive
E11	Mixed		TBpositive	Vpositive
E12	Positive		TBpositive	Vpositive
E13	Positive		TBpositive	Vnegative
E14	Positive		TBpositive	Vnegative
E15	Positive		TBpositive	Vpositive
E16	Positive		TBpositive	Vpositive
E17	Positive		TBnegative	Vpositive
E18	Positive		TBpositive	Vpositive
E19	Positive		TBpositive	Vpositive
E20	Positive		TBpositive	Vpositive
E21	Positive		TBpositive	Vnegative
E22	Positive		TBpositive	Vpositive
E23	Positive		TBpositive	Vpositive
E24	Positive		TBpositive	Vpositive
E25	Positive		TBpositive	Vnegative
E26	Mixed		TBpositive	Vpositive
E27	Mixed		TBpositive	Vnegative
E28	Mixed		TBpositive	Vnegative
E29	Mixed		TBnegative	Vnegative
E30	Neutral		TBnegative	Vnegative
E31	Positive		TBpositive	Vnegative
E32	Negative		TBnegative	Vnegative
E33	Positive		TBpositive	Vpositive
E34	Positive		TBpositive	Vpositive
E35	Positive		TBpositive	Vpositive
E36	Negative		TBpositive	Vnegative
E37	Positive		TBpositive	Vpositive
E38	Positive		TBpositive	Vpositive

DocName	Manual		TextBlob	VaderExt
E39	Positive		TBpositive	Vpositive
E40	Positive		TBpositive	Vpositive
E41	Positive		TBpositive	Vpositive
MUE01	Positive		TBpositive	Vpositive
MUE02	Positive		TBpositive	Vpositive
MUE03	Positive		TBpositive	Vpositive
MUE04	Positive		TBpositive	Vpositive
MUE05	Positive		TBpositive	Vpositive
MUE06	Positive		TBpositive	Vpositive
MUE07	Positive		TBpositive	Vpositive
MUE08	Positive		TBpositive	Vpositive
MUE09	Positive		TBpositive	Vpositive
MUE10	Positive		TBpositive	Vpositive
MUE11	Positive		TBpositive	Vpositive
MUE12	Positive		TBpositive	Vpositive
MUE13	Positive		TBpositive	Vpositive
MUE14	Positive		TBpositive	Vnegative
MUE15	Positive		TBpositive	Vpositive
MUE16	Positive		TBpositive	Vpositive
MUE17	Positive		TBpositive	Vpositive
MUE18	Positive		TBnegative	Vpositive
MUE19	Positive		TBpositive	Vpositive
MUE20	Positive		TBpositive	Vpositive
MUE21	Positive		TBnegative	Vpositive
MUE22	Positive		TBnegative	Vnegative
MUE23	Positive		TBpositive	Vpositive
MUE24	Positive		TBpositive	Vpositive
MUE25	Positive		TBpositive	Vpositive
MUE26	Positive		TBpositive	Vpositive
MUE27	Positive		TBpositive	Vpositive
MUE28	Positive		TBpositive	Vpositive
MUE29	Positive		TBpositive	Vnegative
MUE30	Positive		TBpositive	Vpositive
MUE31	Positive		TBpositive	Vpositive
MUE32	Positive		TBpositive	Vpositive
MUE33	Positive		TBpositive	Vpositive
MUE34	Positive		TBpositive	Vpositive
MUE35	Positive		TBpositive	Vpositive
MUE36	Positive		TBpositive	Vpositive
MUE37	Positive		TBpositive	Vpositive
MUE38	Positive		TBpositive	Vpositive
WM001	Positive		TBnegative	Vpositive
WM002	Positive		TBpositive	Vpositive
WM003	Positive		TBpositive	Vnegative
WM004	Mixed		TBnegative	Vnegative
WM005	Negative		TBneutral	Vnegative
WM006	Mixed		TBpositive	Vpositive
WM007	Positive		TBpositive	Vpositive
WM008	Neutral		TBnegative	Vpositive
WM009	Positive		TBpositive	Vpositive
WM010	Mixed		TBnegative	Vnegative
WM011	Positive		TBpositive	Vpositive
WM012	Mixed		TBpositive	Vnegative
WM014	Positive		TBpositive	Vpositive
WM015	Positive		TBpositive	Vpositive
WM016	Positive		TBnegative	Vnegative
WM017	Positive		TBpositive	Vpositive
WM018	Mixed		TBpositive	Vnegative

DocName	Manual		TextBlob	VaderExt
WM019	Positive		TBpositive	Vpositive
WM020	Mixed		TBnegative	Vpositive
WM021	Mixed		TBnegative	Vpositive
WM022	Mixed		TBpositive	Vnegative
WM023	Positive		TBpositive	Vpositive
WM024	Positive		TBpositive	Vpositive
WM025	Positive		TBpositive	Vnegative
WM027	Negative		TBnegative	Vnegative
WM028	Positive		TBneutral	Vpositive
WM029	Positive		TBnegative	Vpositive
WM030	Positive		TBneutral	Vnegative
WM031	Positive		TBnegative	Vnegative
WM032	Mixed		TBneutral	Vpositive
WM033	Positive		TBpositive	Vnegative
WM034	Positive		TBnegative	Vnegative
WM035	Mixed		TBpositive	Vnegative
WM036	Positive		TBpositive	Vnegative
WM037	Positive		TBnegative	Vnegative
WM038	Mixed		TBpositive	Vpositive
WM039	Negative		TBnegative	Vnegative
WM040	Positive		TBpositive	Vnegative
WM041	Positive		TBneutral	Vpositive
WM042	Neutral		TBneutral	Vnegative
WM043	Positive		TBpositive	Vpositive
WM044	Positive		TBpositive	Vnegative
WM046	Positive		TBpositive	Vnegative
WM047	Negative		TBnegative	Vpositive
WM049	Negative		TBnegative	Vpositive
WM050	Mixed		TBpositive	Vnegative

DocName	Manual		TextBlob	VaderExt
WM051	Positive		TBneutral	Vneutral
WM052	Positive		TBpositive	Vpositive
WM101	Positive		TBpositive	Vpositive
Positive	174		188	147
Mixed	45		10	2
Negative	35		62	111
Unclear	6			
Undefined	0		0	0
	260		260	260

Appendix G Umbrella Review search terms

Setting (general)	Setting (specific)	Analysis	Content	Usage	Technique (general)	Technique (specific)	Tools	Ontologies	Causality	Subject
social media	Facebook	social media analysis	patient experience	symptom	natural language processing	word*embedding	GATE	UMLS	causality	health
online space	twitter	data mining	patient report*	side effect	NLP	supervised	CTakes	SNOMED	causal link	illness
social network*	patientslikeme	text mining	self reported	side-effect	natural language understanding	unsupervised	MetaMap	RxNorm	causal relationship	disease
online network	askapatient	infoveillance	patient generated	adverse drug reaction	machine learning	semi*supervised	word2vec	SIDER	health related causality	
blog*	healthunlocked	crowdsourc*	user generated	ADR	artifiical intelligence	annotation	doc2vec	MedDRA	causation	
microblog	webmd	crowd sourc*	post	effective*	AI	dictionar*	GloVe	lexicon	causal relation	
discussion forum	reddit	social listening	message	outcome	deep learning	rule*based	BERT	ontology		
online community		dataveillance	thread	quality of life	reinforcement learning	SVM	XLNet			
internet		social media listening	conversation	QoL	transfer learning	decision*tree	BRAT			
online		social media mining	ePAT	HRQoL	neural network	knowledge*graph	tagtog			
health forum		signal detection	UGC	impact	algorithm*	CRF	LightTag			

Setting (general)	**Setting (specific)**	**Analysis**	**Content**	**Usage**	**Technique (general)**	**Technique (specific)**	**Tools**	**Ontologies**	**Causality**	**Subject**
patient forum		content analysis	patient authored	causality	corpus linguistics	dictionary*based	StanfordNLP			
patient community			knowledge sharing	causation	network analysis	Latent*Dirichlet	NLTK			
online patient community			self-management	causal link	sentiment analysis	LDA	spaCy			
message board				causal relationship	classifi*	named entity recognition	gensim			
online health community				drug	cluster*	NER				
forum				intervention	topic model*	entity detection				
forum (near health/patient/online)				pharmocovigilance	semantic	entity relation*				
community (" ")				adverse effects	data science					
				adverse events	big data					
					supervised learning					
					unsupervised learning					

Appendix H Included papers by journal

Review	Journal
Allen 2016	Journal of Medical Internet Research
Barros 2020	Journal of Medical Internet Research
Dol 2019	Journal of Medical Internet Research
Drewniak 2020	Journal of Medical Internet Research
Hamad 2016	Journal of Medical Internet Research
Kim 2017	Journal of Medical Internet Research
Mavragani 2020	Journal of Medical Internet Research
Zhang 2020	Journal of Medical Internet Research
Lardon 2015	Journal of Medical Internet Research
Wongkoblap 2017	Journal of Medical Internet Research
Lopez-Castroman 2019	Journal of Medical Internet Research
Demner-Fushman 2016	Yearbook of Medical Informatics
Filannino 2018	Yearbook of Medical Informatics
Gonzalez-Hernandez 2017	Yearbook of Medical Informatics
Lau 2019	Yearbook of Medical Informatics
Neveol 2017	Yearbook of Medical Informatics
Neveol 2018	Yearbook of Medical Informatics
Staccini 2017	Yearbook of Medical Informatics
Pourebrahim 2020	IEEE
Qiao 2020	IEEE
Abd Rahman 2020	IEEE
Dobrossy 2020	PLoSone
Charles-Smith 2015	PLoSone
Wilson 2015	International Journal of Qualitative Methods
Zhang 2018	International Journal of Qualitative Methods
Al-Garadi 2016	Journal of Biomedical Informatics

Review	Journal
Gupta 2020	Journal of Biomedical Informatics
Skaik 2020	ACM Computer Survey
Fung 2016	American Journal of Infection Control
Sinnenberg 2017	American Journal of Public Health
Patel 2015	American Journal of Medicine
Tricco 2018	BMC Medical Informatics & Decision Making
Golder 2015	British Journal of Clinical Pharmacology
Vilar 2018	Briefings in Bioinformatics
Edo-Osagie 2020	Computers in Biology & Medicine
Santos 2019	Computers & Industrial Engineering
Ho 2019	Current Pharmaceutical Design
Karmegan 2020	Disaster medicine and public health preparedness
Convertino 2018	Expert Opinion on Drug Safety
Gianfredi 2018	Frontiers in Public Health
Dreisbach 2019	International Journal of Medical Informatics
Lafferty 2015	International Review of Psychiatry
Abbe 2016	International Journal of Methods in Psychiatrtric Research
Sarker 2019	Journal of American Medical Informatics
Falisi 2017	Journal of Cancer Survivorship
CastiiloSanchez 2020	Journal of Medical Systems
Cheerkoot-Jalim 2020	Journal of Knowledge Management
Yin 2019	JAMA Medical Informatics
Zunic 2019	JMIR Medical Informatics
Gohil 2018	JMIR Public Health Surveillance
Giuntini 2020	Journal of Ambient Intelligence and Humanized Computing
Sharma 2016	Methods of Information in Medicine
Calvo 2017	Natural Language Engineering
Injadat 2016	Neurocomputing

Review	Journal
Sharma 2020	Pharmacology Research and Perspectives
Wong 2018	Pharmocotherapy
Ru & Yao 2019	Social Web and Health Research (book)
Su 2020	Translational Psychiatry

Appendix I Dataset cleaning

On entire file:

TimeStamp:

Parse on “ “ > reduce to PostDate.

Create PostYear from PostDate.

Delete unwanted bits.

URL:

Parse on “//:” to lose http part from front of field.

Parse on “/” to separate other subfields. Rename first subfield as SiteName

Create SubSite field: if site = reddit > SubSite = subreddit name

if([SiteName]=="www.reddit.com",[url_2_3],””)

Author:

Split author field (remove siteID)

Content:

Replace ‘_x000D_’ with space in content & title fields

Generate > replaceAll([Content], “\\n”, “ “)

Other:

Remove Treato generated fields

Try to identify thread titles where possible

Remove empty columns

Appendix J OpenRefine

Text transform on 58750 cells in column SiteName: value.replace("www.","")

Remove exact duplicates (262 posts)

Remove WebMd posts with truncated duplicates (39 post)

Create new column PossSpam based on column Text by filling 68721 rows with grel:if(value.contains("http"),"http", if(value.contains("www"),"www",""))

www = 156, http = 3207, blank = 65359

Manually inspect the 156 ‘www’ posts – 36 ‘obvious spam’, 120 not. No deletions made at this point – decision re spam classifier pending…

Dimension reduction of remaining columns of url parsing. Columns copied to maintain original detail then reduction carried out on the new variants. Aim is to only keep for analysis conditions or titles, ignore thread identification. Can always go back to original data if this becomes important in the future.

URL22: 3707 values

Value.replace([/0-9]/,””) reduced no of values to 636

Cluster & Edit (key collision / fingerprint) plus delete irrelevant = 621 values

Cluster & edit (Nearest neighbor / levenshtein / radius 1.0 / block Chars 6) plus delete irrelevant clusters = 604 values

Cluster & edit (Nearest neighbor / ppm / radius 1.0 / block chars 6) plus delete irrelevant clusters = 576 values

Cluster & edit (Nearest neighbor / levenshtein / radius 1.0 / block chars 5) plus delete irrelevant clusters = 572 values

Cluster & edit (Nearest neighbor / levenshtein / radius 2.0 / block chars 4) plus delete irrelevant clusters = 568 values

Cluster & edit (Nearest neighbor / ppm / radius 2.0 / block chars 4) plus delete irrelevant clusters = 546 values

Replace all generic / non health values with “” = 503 values

value.replace("index.asp?forumID=&subject=","") = 502 values

value.replace("-t.html#p","") = 501 values

value.replace("showthread.php?","") = 498

blastmytwitter.com = obvious spam 32 records all by 1 author wrhel DELETE

replace “&p=#post” / etc = **469**

URL_2_3: 10,468 values

Cluster & Edit (key collision / fingerprint) plus delete irrelevant = 10,403 values

Cluster & Edit (key collision / ngram-fingerprint/ ngram size 2) = 10,385 values

Cluster & edit (Nearest neighbor / ppm / radius 1.0 / block chars 6) plus delete irrelevant clusters = 10,059 values

Cluster & edit (Nearest neighbor / levenshtein / radius 1.0 / block chars 6) plus delete irrelevant clusters = 10,039 values replace(value,/\d/,"") = 3416 values - too blunt > lost some from mg /yrs/ years etc

Cluster & Edit (key collision / fingerprint) plus delete irrelevant = 3321 values

Cluster & edit (Nearest neighbor / ppm / radius 1.0 / block chars 10) plus delete irrelevant clusters = 3086 values

Cluster & Edit (key collision / fingerprint) plus delete irrelevant = 3012 values

Filter on reddit sites. If SubSite=url23 > delete url23.

URL24: 17,811 values

Cluster & Edit (key collision / fingerprint) plus delete irrelevant = 17705 values

value.replace(/[\d]+[-]/,"") = 17572

Cluster & edit (Nearest neighbor / levenshtein / radius 1.0 / block chars 6) plus delete irrelevant clusters = 17140

Cluster & edit (Nearest neighbor / levenshtein / radius 1.0 / block chars 6) plus delete irrelevant clusters = 17132

value.replace(/[\d]+[-]/,"") = 17005

Split column on “.” > Delete unwanted columns = 15607

Cluster & Edit (key collision / fingerprint) plus delete irrelevant = 15582 values

Cluster & edit (Nearest neighbor / ppm / radius 1.0 / block chars 6) plus delete irrelevant clusters =15353

if(value.contains("?page="),"", value) = 15319

if(value.contains("comment_id"),"", value) =10873

Cluster and edit: (Key collision / metaphone3) plus delete irrelevant = 6570

Cluster & Edit (key collision / fingerprint) plus delete irrelevant = 6564values

Cluster and edit: (Key collision / cologne phonetic) plus delete irrelevant = 6446

Cluster and edit: (Nearest neighbour / levenshtein 2.0 /8) plus delete irrelevant = 6424

Split column on “?” > get rid of excess content

Cluster & edit (Nearest neighbour / ppm / radius 1.0 / block chars 6) plus delete irrelevant clusters =**6329**

url2_5: 17,575 values

Cluster & Edit (key collision / fingerprint) plus delete irrelevant =17567

Cluster and edit: (Nearest neighbour / levenshtein 1.0 /6) plus delete irrelevant (x3) =15602

Cluster & edit (Nearest neighbour / ppm / radius 1.0 / block chars 10) plus delete irrelevant clusters (x5) = 13866

Cluster & edit (Nearest neighbour / ppm / radius 2.0 / block chars 10) plus delete irrelevant clusters (x2) = 13489

Cluster & edit (Nearest neighbour / ppm / radius 3.0 / block chars 10) plus delete irrelevant clusters (x2) = 13296

Cluster & Edit (key collision / ngram fingerprint/2) plus delete irrelevant = 13284

Cluster & Edit (key collision / ngram fingerprint/3) plus delete irrelevant = 13274

Cluster & Edit (key collision / ngram fingerprint/4) plus delete irrelevant = 13261

Cluster & Edit (key collision / ngram fingerprint/5) plus delete irrelevant = 13248

Cluster & Edit (key collision / ngram fingerprint/6) plus delete irrelevant = 13183

Cluster & Edit (key collision / ngram fingerprint/7) plus delete irrelevant = 5466

Cluster and edit: (Key collision / metaphone3) plus delete irrelevant (x2) =2672

Cluster and edit: (Key collision / cologne phonetic) plus delete irrelevant (x2) = **2616**

url2_6: 11,400 values

Cluster and edit: (Nearest neighbour / levenshtein 1.0 /6) plus delete irrelevant =10,793

Cluster and edit: (Nearest neighbour / levenshtein 1.0 /6) plus delete irrelevant =10,210

Cluster & edit (Nearest neighbour / ppm / radius 1.0 / block chars 10) plus delete irrelevant clusters = 9296

Cluster and edit: (Key collision / metaphone3) plus delete irrelevant = 7427

url2_7: 22,296 values

Cluster and edit: (Key collision / metaphone3) plus delete irrelevant = 5348

Cluster and edit: (Key collision / cologne phonetic) plus delete irrelevant = 3786

Cluster and edit: (Key collision / Daitch -Mokotof) plus delete irrelevant = 2645

Cluster and edit: (Key collision / Beider-Morse) plus delete irrelevant = 2615

Cluster and edit: (Nearest neighbour / levenshtein 1.0 /6) plus delete irrelevant = 2611

Cluster and edit: (Nearest neighbour /ngram fingerprint/ 10) plus delete irrelevant = 20

Delete irrelevant = **17**

Get values into ThreadTitle

If >1 possible value – need to sort out if values are either, subsite, title or irrelevant.

Extract all SiteConditions from manual assessment of SiteName and SubSite.

This allowed for extraction of SiteCondition from sites like reddit with subreddits for each topic and general health forums such as HealthUnlocked with separate subforums for different conditions. Will not be an exhaustive list – (use NER on Text field to extract full list) but will give indications of the range and where posters are choosing to comment.

Spam Removal

Appendix K P3 Clean Data Structure

Field Name	Field Type	Description
P3ID	Integer	Allocated post ID
SiteName	Text	Main site
SubSite	Text	Sub reddit /forum or thread or forum that post was made on
SiteCondition	Text	Condition extracted from site URL
TextOnly	Text	Narrative content of post
Title	Text	Title of thread
TextWithTitle	Text	Combined narrative and thread title
PostYear	Integer	Year that post was made
Author	Text	Author user handle – deleted after duplication checks

Appendix L Stopword Lists

NLTK

```
['i', 'me', 'my', 'myself', 'we', 'our', 'ours', 'ourselves', 'yo
u', "you're", "you've", "you'll", "you'd", 'your', 'yours', 'your
self', 'yourselves', 'he', 'him', 'his', 'himself', 'she', "she's
", 'her', 'hers', 'herself', 'it', "it's", 'its', 'itself', 'they
', 'them', 'their', 'theirs', 'themselves', 'what', 'which', 'who
', 'whom', 'this', 'that', "that'll", 'these', 'those', 'am', 'is
', 'are', 'was', 'were', 'be', 'been', 'being', 'have', 'has', 'h
ad', 'having', 'do', 'does', 'did', 'doing', 'a', 'an', 'the', 'a
nd', 'but', 'if', 'or', 'because', 'as', 'until', 'while', 'of',
'at', 'by', 'for', 'with', 'about', 'against', 'between', 'into',
'through', 'during', 'before', 'after', 'above', 'below', 'to', '
from', 'up', 'down', 'in', 'out', 'on', 'off', 'over', 'under', '
again', 'further', 'then', 'once', 'here', 'there', 'when', 'wher
e', 'why', 'how', 'all', 'any', 'both', 'each', 'few', 'more', 'm
ost', 'other', 'some', 'such', 'no', 'nor', 'not', 'only', 'own',
'same', 'so', 'than', 'too', 'very', 's', 't', 'can', 'will', 'ju
st', 'don', "don't", 'should', "should've", 'now', 'd', 'll', 'm'
, 'o', 're', 've', 'y', 'ain', 'aren', "aren't", 'couldn', "could
n't", 'didn', "didn't", 'doesn', "doesn't", 'hadn', "hadn't", 'ha
sn', "hasn't", 'haven', "haven't", 'isn', "isn't", 'ma', 'mightn'
, "mightn't", 'mustn', "mustn't", 'needn', "needn't", 'shan', "sh
an't", 'shouldn', "shouldn't", 'wasn', "wasn't", 'weren', "weren'
t", 'won', "won't", 'wouldn', "wouldn't"]
```

Additional stopwords added to nltk – topic modelling

'modafinil', 'nuvigil', 'armodafinil','provigil', 'modafanil','nootropic', 'nootropics', 'stack', 'sideeffects'

'is', 'it', 'the', 'and', 'but', 'also', 'then'

Appendix M Top2Vec DeepLearn Topics and words

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
0	2017	insurance	Acquisition		Insurance	insurance coverage cover authorization medicare covered pocket copay denied approve refills company pay walgreens pays supply deductible cost paying costs afford covering savings refill insurances medicaid companies costco approved formulary tier filled assistance cvs hoops fill appeal covers ins goodrx coupon auth income dexone pricing samples expensive appeals card paid
1	1614	husband	Symptom Impact	Outcome	Family	husband kids son house mom couch mother laundry hh dishes daughter asleep teachers nap family park teacher her home falling cooking bathroom naps dad school fall friends automatic dog ready room supportive paralysis alarm bed car cataplexy dinner grocery him shower driving floor apartment kid she alarms wife parents stairs
2	1485	gpc	Other Interventions			gpc oxiracetam aniracetam pramiracetam noopept cdp bacopa alcar stack mane choline piracetam lion ashwagandha racetams alpha epa monnieri dha rhodiola fish racetam dmae centrophenoxine creatine theanine ginkgo biloba phenylpiracetam oxi uridine picamilon omega ani sulbutiamine huperzine gingko oil multivitamin carnitine taurine nefiracetam bitartrate acetyl inositol extract phenylalanine vinpocetine multivitamins supplements
3	801	rash	SideEffects		Rash	rash sjs allergic skin rashes johnson steven hives stevens itchy sores reaction itch ulcers swelling itching throat reactions syndrome sore allergies rare allergy fever lips spots blisters threatening antihistamines fatal swollen benadryl flu discontinue mouth arms nose freaked painful coincidence dry antihistamine subsided died serious histamine infection red cleared infections
4	790	reuptake	MoA			reuptake dopamine adenosine dat inhibiting norepinephrine inhibitor receptors histaminergic agonism serotonin inhibition noradrenaline orexin neurons action nucleus histamine adrenergic dri receptor mechanism hypothalamus agonist indirectly gaba acts transporter activation uptake neuron agonists stimulates weak selective mechanisms affinity glutamate dopaminergic inhibit systems releasing da inhibitors antagonist actions directly psychostimulant promoting inhibits

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
5	661	build	Dosage		Tolerance	build tolerance builds cycling buildup built consecutive cycle building alternate weekends reset holidays break cycled resistance cumulative bromantane rapidly alternating usage aggrenox flucloxacillin cycles redose daily wed fri weekdays gradual dependency returns sparingly pea stabilize moduretic ideally phenibut depleted week skipping spironolactone everyday dosing mon downs tolerant develop growth chasing
6	660	waking	Symptom Impact	Effect		waking asleep slept nap sleeping wake alarms woke alarm night groggy bedtime bed restful falling sleepy refreshed nighttime clock fall yawning limitstart noon naps fragmented napping pm tired awake worn rested mornings exhausted fell laying nights cereal drove dozing woken rem nightmares sleep afternoons spells pattern paralysis foggy dream lunchtime
7	655	bipolar	Conditions		Mental Health	bipolar mania manic stabilizer hypomanic antipsychotic lithium tegretol pdoc hypomania seroquel abilify antipsychotics trileptal stabilizers lamictal zyprexa depakote neutral latuda aap geodon depressive psychosis episode atypical saphris psychotic hypo sedating depression pdocs mood episodes risperdal stable ii cocktail moods rapid activating sedation depressed blown lamotrigine fetzima bp psychiatrists olanzapine mdd
8	653	amantadine	Other Interventions			amantadine karen lipoic msers adderrall neuros neuro ms constipation carnitine xx tecfidera relapses viral acetyl symmetrel amantidine ldn lisa ampyra onemedstore clorazepate rrms ulcers rebif parcel fatigue hugs artvigil perk employer wheelchair ala dmd vibramycin wakalert avonex advertised alex dollar slightest plz goodness prednisolone sounded inexpensive laxative inconsistent fakes tremor
9	649	drinking	Alcohol	Side Effects Dosage		drinking hangover drink drank water alcohol beers dehydration hydrated dehydrated drunk cups drinks exertion thirsty wine soda decaf hungover shake hydration litres cup drinker coffee beer coconut oz ice caffeinated ache headache protein breakfast booze floor fluid intake bathroom gross hungry headaches bowel brutal tea coke rehab espresso toilet poop
10	611	mslt	Conditions		Sleep	mslt latency psg rem ih arousals naps diagnosis eds cataplexy idiopathic narcolepsy onset apneas hypersomnia diagnostic hypersomnolence hh showed fragmented abnormal apnea study stage nap specialist min osa criteria sleep diagnosed paralysis slept fell snoring xyrem ahi pulmonologist cpap respiratory diagnose apnoea stages daytime minutes dx treated results test neurologist

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
11	591	eat	Food		Appetite	eat carbs eating protein hungry calories bread snack meal veggies meals carb meat vegetables fat appetite calorie fruit food fats sugar fruits hunger diet breakfast dinner foods nuts chocolate chicken suppressant weight dairy diets cheese coconut keto eggs lunch ate intake cook lbs egg decaf shakes butter cream fasting ice
12	524	adderrall	Other Interventions			adderrall karen amantadine hugs lipoic luck carnitine inexpensive jitters lisa foods caffine cranky helps lifesaver alex vitamins fatigue works acetyl perk helped norco ala error energy concerta magnesium root bless saver goodness pace sublingual crash overdo unpredictable remedies mo acid natural flare worked dollar buproprion adrafanil aldactazide blah jittery wishing
13	522	physics	Education			physics math studying concepts meditation exams exam programming tasks learning learn skills creativity distractions discipline boring material task gpa semester language playing creative sentences genius concentrating intelligence engineering focus focused grades remembering project assignments video distracted lectures distraction completing topics lecture projects games english dependant mindset paragraph retain thinking connections
14	507	hab	Modafinil		Name variations	hab artvigil waklert modvigil wakalert sun modalert pharma batch modafinilcat brands armod batches sunpharma pack modafil nrx modup armo packaging strips fillers arrived fake blister rxrex afinil modadropship medsforbitcoin fakes onemedstore foil cat counterfeit responder sunmodalert realised ordered reddit retailer vendor sample strip pharmaceuticals legit armodafinil prescript disappointed compares imgur
15	507	tunnel	Symptoms		Fatigue	tunnel yawning felt focused woke lunch distracted vision pm yawn usual hours mins stayed drank hour normally bored revision slight later productive ate feel focussed procrastinating hrs drove popped tired desk noon sleepy focus distractions tasks eyes finished evening hungry took feeling groggy mentally essay slept today spaced motivated morning

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
16	503	tysabri	Conditions		Multiple Sclerosis Cancer	tysabri copaxone relapses rebif avonex rrms tecfidera dmd relapse ms mri lesions ampyra gilenya progression ppms infusion neuro injection spasticity bladder wheelchair amantadine spinal injections flares neurologists flare ldn spasms shots fatigue steroids msers neuros ccsvi spine progressive baclofen numbness chemo amantidine dx therapies lisa yrs neurologist treatments sclerosis vertigo
17	491	tracking	Acquisition		Online purchasing	tracking received medsforbitcoin package confirmation arrived payment email customer placed reship btc shipment modadropship sent emailed service number transaction usps refund envelope ordered packages ems parcel arrive bitcoin packaging skypax receive orders modup emails bitcoins discreet customs recieved communication ship onemedstore modafinilcat tpe contacted receiving rxrex customers shipped address vendor
18	471	cpap	Conditions		OSA	cpap machine ahi asv bipap mask leaks apneas pap osa apnea apap obstructive compliant arousals settings titration masks machines respiratory breathing pressure central oxygen nasal numbers daytime auto snoring events data surgery nightly sleepiness residual periodic sleep air breathe pulmonologist fragmented untreated therapy cm pillow adjusted ai restorative average software
19	461	liver	MoA			liver prodrug converted metabolized adrafinil converts metabolizes conversion metabolite enzymes olmifon toxicity unscheduled metabolize adra precursor metabolites enzyme bloodstream thistle toxic identical inactive active broken damage processes resulting elderly analogue importing primarily equal discontinued possess crl processed metabolism metabolic regulated pseudo systems roughly legal eugeroics pro approximately equivalent powdercity searched
20	451	appreciated	Information sharing			appreciated hi anyone advance insight thank thanks hello appreciate input posts share wondering welcome greatly comments asks curious hey grateful chime sharing wondered ebv lurking apologies anybody tecfidera replies shared please post rambling feedback member joined recommendations guys tysabri experiences insights topics site nubrain xx wakalert subreddit flucloxacillin amantadine tips

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
21	435	concerta	Other Interventions			concerta ir vyvanse focalin xr straterra ritalin extended gp desoxyn dexedrine acting dex release prescribe aldactazide combivir nuvigl rheumatologist microdose prescribing appointment strattera grad refills psychologist adderall adhd overkill amantadine psychiatrist asks referral ciprofloxacin ranexa shortage norvasc adderrall cozaar gps prescribes alesse neuro hoops microdosing ot levlite methylphenidate floxin clorazepate
22	420	importing	Acquisition		Online purchasing	importing import customs seized illegal imported laws letter possession legal legality law caught dea package country border packages reship eu unscheduled controlled prohibited overseas registered iv parcel police australian dexone possess goods countries shipment intent states japan licensed substance lawyer pbs custom shipments federal grey stating china united foreign destroy
23	409	msers	Symptoms			msers karen neuros amantadine ms relapse xx relapses fatigue gps dmd sue progression gp neuro nurse symptom copaxone rrms ppms rebif licensed trust debilitating remission lifesaver neurologists treatments adjustments avonex variable overdoing managing spasticity manage tysabri nhs neurologist ourselves hugs lisa wheelchair prescribe ldn luck cope tecfidera supportive resting ccsvi
24	408	ra	Conditions		Autoimmune	ra fibro rheumatologist fibromyalgia rheumy cfs chronic savella lupus fatigue plaquenil fm restorative ciltep autoimmune collective debilitating esterom curing flares tremendously adderrall cymbalta ltp arthritis prednisone fog aches amantadine forskolin wheel remedies acupuncture flare joint drugreview saver limitstart acetylcholine artichoke illnesses libido helped oxycocet answers denial ebv hugs interferes lifesaver
25	405	lay	Symptom Impact		Fatigue	lay asleep fall nap falling couch sleepy bed slept noon naps alarm driving laying drugreview exhausted fell tired sleeper refreshed woke yawning wake pm napping bedtime limitstart tv fallen alarms wheel worn chair sleeping dozing afternoons nuvigl teachers afternoon zombie nighttime desipramine sleeps ready dishes awake ritilan ra pillow olanzapine

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
26	403	worn	Symptom Impact		Fatigue	worn yawning limitstart exhausted constipation sleepy nap laying awake zombie crash cranky tired adderrall bathroom lay pregnant dishes dozing crashed nodding bed vape drugreview asleep chocolate dragging barely slept perk eyes noon couch drank crashing assure spacey constantly drowsy chores sinemet fatigued fall sleeping spaced falling woke cook lipoic nod
27	394	obstructive	Conditions		OSA	obstructive excessive apnea shift sleepiness daytime shifts disorder osa disorders treat narcolepsy medically cpap rotating adults apnoea associated approved workers compliant conditions residual label sleep central breathing patients syndrome diagnoses eds promoting treating suffering indications wakefulness drowsiness idiopathic frequently commercial sdmb marketed treatment es designed students prescribed condition commonly continuous
28	382	parnate	Unclear			parnate nardil augmenting maois maoi desipramine tca nri augment augmentation poop snri mao hypotension abilify nortriptyline antidepressant hypertensive reboxetine sa pdoc aap contraindicated scott energizing titrate sedation moclobemide imipramine sedating snris atypical ssri bupropion tricyclic agomelatine ads lyrica anhedonia fetzima jim crisis ssris ad dexamphetamine pdocs titrating monitoring mr neutral
29	368	weight	SideEffects		Weight	weight lbs pounds gained gain lost calories loss lose gaining overweight losing appetite eating eat hungry weigh suppressant fat calorie lb topamax carbs metformin hunger food carb savella kg bumped pound bread meals shakes meal nausea struggled nauseated intensely diet healthier portion exercising force metabolism zyprexa neutral june nauseous fruit
30	356	teva	Modafinil		Manufacturers	teva cephalon patent mylan manufacturer inc manufacturers generics expired companies pharmaceuticals par patents maker versions makers pharma suit company manufactured rights marketing brand market version delay selling pharmaceutical april manufacture generic name agreement fda hab watson european sun million costco licensed counterfeit news manufacturing branded purchased sold supply trade indian

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
31	346	fma	Recreational		Recreational	fma fa eph rc ethylphenidate mpa megathread rcs fpm stim recreational stims functional analogues comedown euphoria euphoric fucked meo redose meth amps mph methyl analogue mdma medium party speed drugged edited amp smooth psychedelics duration stimulation binge benign eugeroics clean dri coke cos etizolam amphetamine ime compulsive subtle lasting shit
32	345	maximize	Unclear			maximize cereal protein flucloxacillin beans heh ai breakfast tech zopiclone bowl beast upgraded humor tonight gear loosing lorazepam chasing overkill tapered round sinemet suprax homework ativan bupe taper gross olanzapine mum stabilize apple intend organized tap glycine decaf sigh measured klonopin reflux buspar tulip cd stage tms blunting mini jpg
33	343	randomized	RCTs			randomized placebo systematic blind clinical conclusions patients subjects pmid participants efficacy cognition studies double abstract measures psychiatry published trials adults improvements cognitive observed shown associated review trial reported treatment significantly executive functions study findings improved enhancement objective improving schizophrenia evidence suggests pharmacological literature outcomes individuals data behavioral deficits dependence controlled
34	337	flmodafinil	Other Interventions			flmodafinil crl hydrafinil fluorenol fladrafinil eugeroic analog methyl eugeroics affinity promise ceretropic analogues metacam wakefulness duration aka petcam prl stronger enhancer compound adrafinil dat potent notably olmifon sticking compounds nsi opportunity euthyrox smoother promotion molecule meanwhile enhancing lacks pitolisant structure introduce potency suspension ring novel speedy closely cleaner boosting range
35	322	smell	SideEffects		Urine	smell urine smelling smells odor pee sulfur taste piss bitter tastes fake batch breath sweat counterfeit geoff distinct fakes plastic water dehydrated nmr methylene odd saliva toilet jpg passed gross tests dissolve attached bladder thirsty kidneys reminded bowl ferret likelihood imgur modafnil cat nauseated tongue swallow responder ache blister kidney

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
36	321	stomach	Dosage		Food	stomach empty breakfast meal eat hungry food meals protein eaten dinner snack eating lunch cereal carbs absorption shake bread carb nauseous ate chicken apple upset ache nutrients water beans fats nausea mins thirsty calorie veggies chocolate decaf bowel cup fruit mayo sickness bowl shower fat hunger gross vegetables balanced ice
37	317	yea	Unclear			yea model addicts da tinnitus amph ghb hydrocodone army solutions sobriety cured knocks horny ignorant nitrofurantoin gun repeated teenager lexotanil amps ama meth ranexa failing altered fma dat sibutramine wouldnt cocaine ccsvi abusing neurotoxicity methamphetamine um transporter tarka euphoric hardcore bromazepam knocked fa baclofen dates fasoracetam yawn ruined noprescription disappointing
38	316	venlafaxine	Spam??			venlafaxine agomelatine snri ssri mirtazapine reboxetine escitalopram augment nortriptyline antidepressant augmenting ssris imipramine effexor tca amitriptyline tricyclic snris buspirone augmentation remeron xl bupropion fluoxetine sertraline pristiq xeloda zoloft fuel apathy nri abilify moclobemide parnate desipramine stablon anhedonia sexual ad serotonin citalopram prozac advertisement atypical cymbalta maoi gad trazodone buproprion wellbutrin
39	315	vit	Vitamins			vit vitamin vitamins iu complex iron magnesium deficient multi anemia zinc deficiencies multivitamin deficiency minerals thyroid lipoic fish calcium omega oil tsh supplement checked cog sunlight potassium carnitine mcg panel hypo msers bloodwork supplements units synthroid ra citrate supplementing acid epa ala essential nutrients sublingual absorb pep amantadine ginseng lifesaver
40	311	skypax	Acquisition		Online purchasing	skypax germany eu reship ship packages parcel netherlands customs countries ships italy shipments package seized india importing ireland country goods located shipment packaging europe send european seller refund modiodal sweden fee medsforbitcoin arrived tracking sending import address border canada mail japan custom sunmodalert america australian worldwide uk checks australia strict

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
41	307	prozac	Other Interventions		Antidepressants	prozac snri ssri psychiatrists ssris depressants lexapro effexor antidepressants snris depressant celexa antidepressant viibryd ads abilify zoloft augmentation buproprion pdoc hycodan suppressant tca ativan remeron depression augment noprescription pdocs wellbutrin pristiq ad xl stabilizers augmenting jayster cymbalta anhedonia recieved fetzima overkill aldactazide lamictal welbutrin advertisement anti tricyclic existed parnate energizing
42	302	feedbacks	Acquisition		Online purchasing - feedback	feedbacks edandmore spier spierx sun modalert modapro brand sunpharma blister customer received envelope iop counterfeit packaging reviews ordered branded merged fake alertec batch tpe responsive ed smell judged emailed brands product bitter ems arrives vendor complaints onemedstore sample shipment taste tracking legit fakes awaiting modavigil visa placed urine tags pleased
43	302	exam	Education			exam exams studying revision assignments grades concentrating final preparing courses upcoming intensive finals physics blah classes nighter student retain engineering graduate math distractions engineer completing session workload school programming finished frisium uni attend semester lectures sessions nighters fma marathon season chess topics concepts occurs finish project university practicing grad finishing
44	296	commenttext	Unclear			commenttext div itemprop entry content class post severity dumb tag restore phendimetrazine augment exclusively combating brazil augmentation escitalopram risperidone desipramine tricky waited facts drugsgear zoloft blunting seizure legitimate wakalert zyban reboxetine scammed bupropion lamictal xeloda cried shelf strangely orlistat alternatively ymmv availability patent intuniv norvir neurology ad gad olanzapine smoother
45	292	narcoleptic	Conditions	Other Interventions	Narcolepsy ADHD Asthma	narcoleptic treats husband hubby complained race son narcolepsy daughter adhd married exelon prosteride addy nuvigl hyperactivity cranky he kids sleeps aricept smoother him sct pemoline inattentive narcoleptics asthma compliant ranexa tadalafil adderrall children parent child aldactazide supeudol pap laxative diagnosed messages his japan coluracetam armod sunmodalert joined teachers born sdmb

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
46	290	zyprexa	Emotions	Conditions Other Interventions	Mental Health	zyprexa wierd resistance im yea prozac vary kinda sh mellow doesnt faith moods ripped alot thats didnt depressed theres lamictal everytime created jesus anyways stream praying mgs paranoid boring abusing calm anymore paranoia thoughts couldnt totally psych sedated sadness agitation wasnt silly cloud sedation cant hyperactivity its gotta mood manic
47	288	birth	Pregnancy /Birth			birth hormonal iud contraception control ladies pregnant pregnancy interferes nu women hormones bc hormone interacts effectiveness endo interact baby breakthrough estrogen interfere estradiol method antibiotics protection breastfeeding stilnox interaction decreases period nursing painful ineffective tecfidera acne avonex methadone barrier bloodstream interacting pharmacist interactions pricey hmm heavier married cycles newly skip
48	287	legs	SideEffects		Anatomy	legs shoulder neck arm joint spasms leg arms muscle cramps shoulders tingling tightness knees painful pains pain knee numbness joints muscles feet chest aches spine stiffness swelling weakness swollen sore tension discomfort tight numb surgeries nauseated rash toes sensations hives twitching ache sensation skin itching itch wheelchair itchy ulcers bladder
49	286	idiopathic	Conditions		Hypersomnia	idiopathic hypersomnia ih hypersomnolence apnea daytime mslt narcolepsy diagnosed obstructive osa diagnosis cataplexy latency sleepiness eds naps specialist excessive psg cpap study rem abnormal newly indications nighttime diagnostic perscribed sleep apnoea teenager residual disorders narcoleptic asv snoring diagnoses fragmented determined paralysis treated untreated criteria believes nap apneas hypothyroidism conditions slept
50	286	microdosing	Dosage		Recreational	microdosing lsd microdose microdoses ug alternating micro mushrooms psychedelics shrooms psychedelic tripping combos mdma dosing trip accurately creativity creative dmt milligrams floxin cos combining fpm rcs experiment redose rebound phenibut peoples experimenting holy neurofeedback meo modaf experiments surprisingly cerebrolysin htp anxiolytic exclusively trips moments planned negatively lostfalcos tool methylene marathon

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
51	285	comparison	Comparison			comparison vs comparing modafnil brings bike animals collective sober amp erection stim meridia emotion curing amps aa horrid coaster amph focalin lacks milder intense ride context binge childhood walls realise enjoyed discussing recreational gut awareness prescript sobriety meth meetings visual tournaments esports software bang clonazolam bursts hardcore hppd killed likelihood
52	285	appreciate	Information sharing			appreciate thanks thank appreciated insight reply advance feedback input hi suggestions welcome posts flucloxacillin replies wondered chris poster levlite info merged hey thread links carvedilol posted post replying anyone hello updates responses opinions update frisium comments blurry gmail sharing please recommendations guys site chime detailed darvocet diethylpropion images member searching
53	284	disability	Symptom Impact			disability employer accommodations job fired jobs understanding supportive boss reasonable adjustments career federal ssdi education paperwork employment teachers employee services department classes disabled school narcolepsy graduate assignments diagnosis newly teacher visits hypersomnolence applying cataplexy income children network desk guilty bills parents naps manage colleagues dealing college diagnosed eds commute pwn
54	275	splitting	Dosage			splitting split yawning afternoon horrid shortness alarm crashed scored washing lunch aldactazide noon kicked feedbacks thirsty doubled clock pounding stairs drank worn weather cereal limitstart beast half hydrated quarters lunchtime mistake depleted subsided tryptophan litres apple bus gilenya drove buildup gross kitchen dozing walks tablet morning frisium midnight levlite reminder
55	274	prednisone	Conditions	Other Interventions		prednisone flares plaquenil rheumy lupus rheumatologist ra inflammation flare fibro autoimmune fatigue joint swelling debilitating fibromyalgia inflammatory savella tremendously cfs joints chronic yo infections disease exhaustion bid restorative chemo anemia flu masks tapering ulcers arthritis ms protocol immune enzymes celebrex dmd fm encouraged couch ironically radiation appt adderrall pain wiped

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
56	274	limitless	Enhancement			limitless nzt movie film closest smart watched enhancer iq intelligence character watching smarter hype genius movies solving reality tv molecule cognitive exist enhancement mystery increasingly claim youtube magical enhancing creatively enhances intellectual named fluorenol enhancers boosting miracle documentary needles enhance ultimate valley world tim executive evil introduction nootropic growing upgraded
57	266	monday	Temporal			monday friday wednesday thursday tuesday sunday weekend mon fri holiday weekends reset wed tarka tomorrow premarin holidays paxil week limitstart insufficiency skipped chores slept christmas sat appointment wound pm crashed endep nightly mri excitotoxicity interviews evista skipping bystolic paperwork lunchtime coaster roller knee pea routine august evening grapefruit tonight benefited
58	263	smoke	Smoking			smoke smoking smoked weed cigarettes cigarette smoker vape marijuana pot tobacco cannabis cessation nicotine stoned mushrooms quit alcohol thc sober beers meo cravings mdma apartment bowl gum loop quitting psychedelics joint heroin relaxing booze massively shrooms cb coke drank chain cocaine craving tears beer king addicted wine cereal crystal opiates
59	255	isomer	MoA			isomer isomers racemic enantiomer molecule mixture molecules patent active mirror inactive armoda isolated converted contains handed chemically ingredient nuvigl afaik identical prodrug conversion armodafinil solely bio unwanted created lasting meaning named formulation compound aggrenox differ pure fluorenol waklert patents crl lasts sulfur manufacturers metabolizes stays chemist shorter claim suggests versions
60	254	tightness	SideEffects		Cardiac	tightness chest palpitations heart pressure pains tight pounding cardiovascular pulse cardiac shortness tachycardia blood rate heartbeat bp painful arm cardiologist attack freaked bpm cramps ringing shaking jaw hypertension breathing tension ears irregular beating breath muscles joint muscle beat shoulder beats sensations throat ache asthma forehead hypertensive dying neck tingling rapid

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
61	254	asleep	Symptoms		Impact	asleep falling fall driving fell wheel bathroom desk tv bed alarm nodding nap couch stairs sleepy lay awake woke wake nod noon shower napping car drugreview alarms dozing slept truck drivers husband randomly road doze watching sleeps foot sitting fallen hour ritilan chair barely drive home woken bedtime room limitstart
62	253	pregnancy	Pregnancy /Birth			pregnancy pregnant breastfeeding baby nursing born child xyrem daughter birth medschat breast son women iud cataplexy pwn married wife teacher vomiting struggled newest woman she contraception her narcolepsy cried nurse getelementbyid children caring worried infections harmful constipation teachers hycodan husband risks category discussed continued difficulties reporting hypersomnolence warnings circles pantoprazole
63	253	seized	Acquisition		Online Purchasing Customs	seized customs package reship letter skypax packages importing shipment parcel sent tracking opened shipments import arrived received refund address envelope held french mail stating registered letters send australian ships custom ship arrive destroy caught ems seizure number usps email goods confirmation ordered country eu seizures item receive imgur thailand nov
64	250	esports	Ethics			esports tournaments banned wada doping players athletes athlete sports prohibited player league agency sport games competitive performance ethical chess poker enhancing advantage unfair game race competition playing ban elite cheating testing play endurance pro professional season steroids debate tested policy world incorrect australian sum gold dilantin named video statement media
65	250	bulk	Acquisition			bulk powder rechem purity retailer china powders trustworthy legit city grams plastic ships spelling pbs qhi frisium protocol powdercity sketchy nmr darvon hypersomnolence worldwide latency caps england lupus sunmodalert sr ship packaged diagnose ordering edandmore alabama thailand conversion gel buying awaiting prices sell cilostazol progressive locked zopiclone shelf shipping branded

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
66	248	suboxone	Addiction			suboxone opiate subutex opiates subs heroin wd withdrawals methadone addicted addict opioids morphine opioid kratom withdrawal taper benzos bupe detox benzo oxy withdrawl oxycontin quit sub hooked turkey fentanyl addicts rehab abusing addiction addictive quitting tapering pregabalin oxycodone dxm intolerable snort meth tapered abused dependency mgs habit clean wean film
67	243	focalin	Spam??			focalin concerta eric straterra vyvanse ritalin dexedrine extended traditional acting adhd ir dex release desoxyn methylphenidate stim stimulants smoother clorazepate nuvigl inattentive adderall strattera prefer marketed vigils moody modaf label adults dydrogesterone xr isomer newly abused hoops eds desirable sdmb prescribes pwn japanese reboxetine titrating effecting dexamphetamine exclusively varies treating
68	242	payment	Spam??			payment btc bitcoin bitcoins debit card transaction visa account credit bank cards paypal fraud confirmation transfer tracking medsforbitcoin cc circle sent customer mastercard email modafinilcat hassle refund geoff sunmodalert scam received checks onemedstore nrx cancel ems reship wire seller placed qhi number paid accept mymodafinil vendor worldwide express discounts recieved
69	238	hair	SideEffects		Hair	hair loss thick growth accutane finasteride losing thin growing spots grow weight officially drastic contfzzza progs propecia itching sudden inducing male litezennopost perception lost deficiencies getstring mirror txt hide boat threatening prosteride gain phendimetrazine pounds existence shed medically sheet married acne yo gaining darvon cook slowed men stabilized prevacid powders
70	237	responder	Outcome		Non responders	responder responders percentage batch responsive genetics gene genes genotype genetic armo snp andme rs cat respond likelihood val expecting aas survey log tulip october awaiting shout epa cm estradiol bpc variables engaged horrid benefited buddy mct horny tweaked negatives metabolizer aa ani threshold talkative patience disappointed tdcs rhodiola FALSE tren

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
71	237	clenching	SideEffects		Clenching	clenching jaw grinding teeth gum mouth tight dry tense chewing remind chew lips bite sore dehydration tension saliva sores tongue painful grind tightness aches muscles throat neck ears thirsty solved ache discomfort adjusting ringing hurt asthma twitching ear dull swollen swallowing noise anxious yawn dehydrated gradual relax chest uncomfortable needing
72	235	alertec	Modafinil		Name variations	alertec modapro modavigil modalert names brand brands modiodal sun generics spier indian versions modvigil branded modafil cephalon name pharmaceuticals sunpharma edandmore counterfeit pilagan mylan manufacturers spierx teva pharma feedbacks tpe authoritative dexone hab prosteride nubrain worldwide qhi iop par cozaar packaging fillers vendor european maker india retailer countries keflex waklert
73	235	mdma	Other Interventions		Recreational	mdma mushrooms roll neurotoxicity viramune cb lsd booze psychedelics psychedelic asshole dmt smoked redose meet microdoses ketamine rolling meo shrooms fancy beers coke trip stoned recreational weed urgent death february fa selank ug messed cheers neurotoxic tripping fun alcohol occasions snp benzodiazepines binge htp tonight tears damage kratom dumb comedown
74	233	nih	Information Sources			nih nlm ncbi pmc gov pubmed articles www http abstract html pdf pmid systematic notably suggesting effected endep according table fwiw ingest enhances restricted seizures incident sertraline aswell pharmacological sorted xeloda htm tremors disorders published blockers studies diagnostic review impairment link unknown criteria ss depressants adults monoamine thc findings euthyrox
75	231	podcast	Biohacking			podcast rogan dave asprey bulletproof joe tim butter biohacking testosterone mct his diets claims paleo talks guy upgraded lostfalcos diet hack skeptical marketing silicon synthetic youtube aging neurofeedback fat david hill bullet athlete he thyroid grass fats valley himself hacking athletes video carb anastrozole rat impressed introduction doubts executive likes
76	230	html	Web pages			html http www com php eriacta page url comments vpxl org reddit showthread blog prescript cycrin pages https link messages click brugada net tadalafil fedex ru google blisters nrx posted drugsgear faq vb jpg strips nih googled reviews worldwide cod browser aspx tid site creatively device article flonase search forums

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
77	230	pilots	Military			pilots missions military fighter soldiers air force forces army jet combat war flying mission scientists flight amphetamines duty students fly tarka solely special enhancing drivers performance uses focussed media speed revealed forming extensively enhancers modern levlite freely ritilan lag doping race athletes australian darvocet ciprofloxacin molecules continuous narcoleptics rates ball
78	229	tremendously	Conditions	Other Interventions	Autoimmune	tremendously fibro lyrica rheumatologist fibromyalgia rheumy savella lupus percocet ra fatigue pain gmail moduretic lisa hugs oxycontin debilitating cymbalta joints flares helped drugreview prednisone aldactazide makers norco hubby morphine narcotic illnesses methadone fm nerve aphthasol understands exercising esterom reps oxycodone darvocet adrafanil cfs ritilan chronic thru weights tysabri vigil adderrall
79	221	ugh	Mixed			ugh coq sadly nmda sucks stromectol deserve adderrall ritilan wikipedia instructions meh smells motilium pqq view smell accupril messes sulfur ketamine antagonists lifesaver desipramine cried item arent posts pissed perscribed nortriptyline agonists cilostazol bloody alesse biaxin urine phendimetrazine deny pregnenolone annoyed pulmicort images joy dizzy laying odor tms wasted links
80	221	didnt	Unclear			didnt amantadine sheffield wasnt havent wouldnt perk dont plz ive doesnt amantidine couldnt cuz karen ins bupe im pregnenolone fakes film horror inexpensive david cant youre laxative wanna church ulcers wakalert cousin rebif bc complained walls zembrin waited dilantin refills clears stories essay ur replying meo pd megathread omg id
81	214	kidneys	SideEffects			kidneys wtf kidney girlfriend mri yellow spinal dehydrated endep gi bowel lesions gift beans vertigo tap sweet fluid settings briefly jump poison scored remembered screaming aches fever progressive swollen subsided herbal mile skipped gas tryptophan periodic vomiting pee swelling eriacta pa freaked remedies intolerable bladder cried ice neck pounding rehab
82	212	newer	Unclear			newer dexone version nuvigl patent carnitine makers modafanil acetyl selank concept eulexin personality aggrenox cousin versions pilots formula older gut hoops provigal liking parcel missions clorazepate modanifil companies soldiers formulation adrafanil jet copay talkative magazine pays prosteride amantadine modaf xeloda lipoic keflex medicaid encouraging floxin prescript downsides spain anastrozole molecules

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
83	211	prn	Unclear			prn flonase risperdal latuda weekly klonopin lunesta lamictal seroquel abilify individual lithium fetzima pdoc group geodon aap neutral depakote hurting cocktail mania therapy zyprexa blown sedating titrate bipolar manic jayster viibryd hunger drugsgear gad turkey asv cr quetiapine trileptal tegretol xanax hmmm ads divided stabilizer downer activating remeron zoloft imitrex
84	210	forming	Recreational		Addiction	forming habit addicted addictive addict recovering abused addiction dependency narcotic abuse addicts keen becomes narcotics supervision dependent ponstan aa dependence hooked license surely adderrall warning perscribed abusing require logic euphoria psychological amphetamines damaging substance motivate medicated shouldn cardiologist grades opiates prone pilots soldiers practitioner missions hyperactive mess productivity label sdmb
85	207	panic	SideEffects			panic attacks attack memories psychosis anxiety episodes desperately episode automatic paranoia fear unmedicated angry hh cataplexy psychotic antipsychotics anxiolytic bipap ly suicidal hallucinations scary blown ptsd progressively disaster crying benzodiazepine hide terrified depakote recipe signature keflex messes rage deals connection crashed gad heart hopeless tears exacerbate pissed impairment fears moments
86	207	cpap	Conditions		OSA	cpap pap machine mask apnea osa ahi obstructive apneas asv bipap leaks compliant masks bi central oxygen nasal machines breathing snoring arousals surgeries daytime apap surgery nightly restorative air sleepiness untreated respiratory device settings husband nose pressure debt sleep titration miserable events fragmented sleeps breathe latency overweight capable narcoleptic residual
87	205	wiki	Unclear			wiki wikipedia en org https paxil url stevens http erowid eugeroics page showthread anxiogenic vpxl endep erection pots darvon tweak tms exacerbate relaxation eriacta prescript losartan html euthyrox com sensations ring php ibs itraconazole eugeroic acupuncture brugada apathy clorazepate fails frisium tid zaps levlite tinnitus carvedilol sulfur beginner andme culprit

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
88	204	inattentive	Symptom Impact	Condition Symptom Impact Other Interventions	Cognitive	inattentive hyperactive adhd sct vyvanse pi impulsive concerta dexedrine add adult ritalin sluggish strattera bipolar dex focalin hyperactivity straterra ocd diagnoses ir hated exacerbated organized confusing quetiapine dexamphetamine perscribed reboxetine adderall afternoons patents hypomanic crashes heh label modafanil unmedicated burnout pdoc eulexin officially teachers official intuniv alter dp productivity adults
89	203	pressure	SideEffects	Symptom Impact	Cardiac	pressure blood pulse heart bp rate bpm raise elevated cardiologist monitor ranges pounding chest checked palpitations hypertension stroke tachycardia cholesterol cardiovascular blocker oxygen aa resting lowers bystolic checking raised heartbeat adrenaline leaks alesse mask laying bipap erection potassium checks monitoring tightness sensations kit eph adjusted measuring collective spells curing rapid
90	199	paralysis	Conditions	Other Interventions	Sleep	paralysis hh hallucinations dreams dream dreaming cataplexy vivid eds rem sp naps laughing automatic xyrem falling asleep episodes mslt attacks diagnosis nightmares lucid ih psg visual narcolepsy voices waking nighttime diagnosed napping fell specialist sudden slept nap newly woke scary fragmented latency sleeps snoring behavior behaviors narcoleptic sleep appointment during
91	199	circadian	Circadian			circadian rhythm melatonin bright clock reset delayed alarm wake light sleep bedtime adapt cycles disorder messed waking nights insomnia shift hygiene sleeping lag phase jet mornings pattern shifts therapy schedule regulate glycine ject cortisol cycle lunesta hypersomnia night awakening laziness earlier late maintaining patterns midday rls chronically lights restful bed
92	199	dmt	Recreational			dmt cb shrooms mxe meo ketamine mdma weed psychedelics lsd mushrooms rcs psychedelic trip tripping etizolam stash dxm trips fpm smoked benzos clonazolam pcp alcohol edited beers rc cocaine occasions sober kratom microdoses benzo isomers booze cannabis ug smoking load coke eph microdose recreational thc stoned mpa fma phenibut fa
93	198	alarm	Outcome		Fatigue	alarm alarms clock wake set waking bed bedtime mornings refreshed slept setting wakes groggy ready noon early woke sleeper morning sleeping downer night loud rested nap restful woken pillow tap shower app late ambien circadian hour pm hypersomnolence minutes drag knocks gabapentin sleep asleep earlier toss bathroom melatonin turn cycles

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
94	196	iron	Tests		Blood Tests	iron anemia deficiency thyroid vitamin tsh deficiencies checked bloodwork hypothyroidism vit units blood panel synthroid cortisol levels deficient chemo iu tested tumor rheumatologist hypo diabetes zinc apnea tests supplementation ruled adrenal labs adrenals cardiologist yay insufficiency potassium fatigue undiagnosed hormone test bone testosterone mri borderline autoimmune vitamins lab apneas fatigued
95	196	bless	Unclear			bless god praying prayers blessed pray laurie levlite gad plaquenil limitstart rambling ya ritilan nuvigl pristiq lisa abilify buspar ty cymbalta ssdi yrs aldactazide wonderful wishing appointments constipation prednisone hycodan understands hugs cranky zolpimist drugreview jim scott geodon vyvance tx lord bout acetazolamide resistant virus edited pramipexole sharing zyprexa fibro
96	195	workout	Exercise		Exercise	workout preworkout workouts pre gym craze scoop lifting cardio creatine dmaa weights fitness bodybuilding endurance fat training eca aas raising athlete taurine peptides butter viracept protein hgh green bpc insulin concentrated supps beta ephedrine stim calorie kid lostfalcos fasting meal gains lift fats blend glucose ec cm crushing mass pair
97	195	bro	Unclear			bro degrees yo shout prednisone buddy collective formulary txt progs libido litezennopost getstring varying contfzzza temperature inpatient sh casual summer accomplish file texas estradiol tamoxifen sooo resources psg porn commit horny heat suicide lexapro reduces FALSE pics walked girl hab adderral imitrex sex darvon anhedonia migraines tapered ama ice augmentation
98	195	images	Unclear			images count greater posts links view must currently forum post purity poster merged knowledgeable lurking levlite onemedstore dedicated edandmore whilst nmr qualified exelon thread offering contribute split tms pics kemadrin informative authoritative urine modifinil legality valerian naltrexone itraconazole trustworthy spier query searching feedbacks posted join essay impulsive mentioning splitting incorrect
99	194	yep	Unclear			yep addicts sobriety retail meth dying crystal remission nope gad dp cured bid abusing meh bowel curing commenttext bout withdrawl comfort intention butalbital overload emsam itemprop march financially meetings ketamine ly excuse coke div burnt teens lives sex joined attractive wow illnesses heroin timestamp ying aa sake boy ltp mdd

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
100	194	tongue	Dosage			tongue sublingually dissolve swallow sublingual swallowing taste saliva bitter orally crushed bloodstream tastes ingest mouth absorption gi water absorbed gross bioavailability liquid chew crush soluble crushing seconds teeth flush glass under sores collapse ulcers nasty throat administration sore quicker sinemet biology holding empty gradual absorb stomach powder chaos snorting bite
101	194	celexa	Other Interventions		Antidepressants	celexa lexapro escitalopram welbutrin citalopram quitting prozac ssri buproprion xl effexor buspar cymbalta zoloft fetzima metacam petcam wellbutrin depressants closely antidepressant vast viibryd latuda buspirone snri hesitant abilify augment megathread trazadone tricyclic zolpimist antidepressants darvocet lamictal mono genius battling signature remeron scares olanzapine pristiq sexual ciprofloxacin taper pdocs pdoc hopeless
102	193	bpm	SideEffects		Cardiac	bpm rate resting heart pulse bp pressure palpitations blood cardiovascular heartbeat hr monitor cardiac hypertension cardiologist tachycardia chest pounding stairs oxygen monitoring irregular breath standing beat shortness rapid average exercising cardio acceptable racing endurance elevated measured raises raise adrenaline tightness sitting minute beats blocker blockers laying chair forehead pots session
103	191	agent	MoA			agent promoting wakefulness eugeroic wakefullness hoops anecdotal notably promise psychostimulant disorders distractions teva fluorenol glycine application mindfulness enhancer vigilance statements agents eugeroics cd jump supporting promote lacks gentle inc benefited exacerbate milder promotion subjective promotes snp sustain pitolisant techniques clorazepate modafinal relaxation cognition bio aggrenox pi enhancers unheard mystery ephedra
104	190	groups	Acquisition			groups canadian delivery fedex cheapest cheap cod buy online membership internet price worldwide overnight motilium mastercard cozaar pharmacies pharmacy discount shipping alabama generic perscription prescript feedbacks indian frisium kemadrin vpxl saturday nextday lamisil union dexone packaging visa prescription counterfeit branded sale discreet iop purchase presciption nubain canada shipped pricing india

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
105	190	turkey	Withdrawal			turkey taper cold quitting withdrawal tapering withdrawals quit stopping wean withdrawl rebound gradually weaning zaps tapered kemadrin cessation withdraw dropping transition hardest shooting klonopin numb flu latuda getelementbyid opiate benzodiazepine hurting lunesta smoking prn limitstart coaster luvox refill eliminated replacing bitch pdoc wd suboxone emotionally effexor display smallest fever attempted
106	190	fear	Emotions			fear er stacks visit panic et tachycardia accidentally dying hooked truth bp od evil fma someday excited lord attack euthyrox cups dangers ai flucloxacillin cardiac loosing tren exception noob fa overdose john ignorant unknown died mct recreationally cardiovascular alters explore eriacta holy cried dizzy fasoracetam mixing complications meo neurotoxicity tears
107	188	wonders	Enhancement			wonders movie epa limitless aldactazide acetazolamide gentle uridine dha ranexa adderrall propecia poster immensely restorative nzt citrate gel fibro flare kamagra incredible darvocet thankful fatique videos tarka contfzzza organization combivir carnitine loved biaxin vigil watched understanding fibromyalgia minded bike litezennopost wiped progs getstring supplementing savella medicating rheumatologist yoga creatively cycrin
108	187	creatively	Outcome			creatively solving increasingly concluded decision possibly improve world making healthy smart safe narcolepsy problem prescribed motion people drug smarter taken solved hooked adults genuinely persons conscious traumatic bouts happiness subsided field memories remembering rare interacting suicidal medication brutal enhancers make movie creative darvocet projects creativity elderly limitless hppd incident agitation
109	186	forskolin	Other Interventions			forskolin artichoke zembrin ltp ciltep extract camp chemically pde tulip induced phenylalanine synergistic alcar ginkgo inhibition lostfalcos stack potentiates executive dlpa biloba dependent learning pregnenolone huperzine serotonergic uridine mechanism interfere consistently dha impair chapter fluency induction ache galantamine diminishing nalt sufficient forgive inducing mechanisms synthesis expression alpha selective gpc choline

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
110	186	reddit	Information Sources			reddit comments afinil https faq com subreddit posted www ptsd comment moderator dual acetazolamide pics longecity beginner beginners http thread video exelon mods brugada posting tracking link aricept guide va vb lurking bouts improvements skeptical aggrenox comprehensive reviews reply modapharma hcg ceretropic tren commented monnieri date sub definitive explained august
111	185	inattentive	Symptoms			inattentive organized coping heh anna skills strategies hyperactive adhd iq adult communicate um projects tasks motivate add organization distraction impulsive effort growing jobs accomplish purely role grades contrary behavioral therapist anecdotally nope semester hypo bipolar living endep tx firm manic teachers buddy accidentally burnout accomplished relationships bored goals improved social
112	184	rs	Effect		Non responders	rs snp genotype andme val gene comt met genes genetic genetics respond aa metabolizer pmid responder responders individuals responsive executive response responded data percentage raw threshold distinct deprived findings population whereas hypothesis prefrontal subjects respect functioning suggests testing responses abstract frequency efficient improving baseline conclusion survey variation environment expression skeptical
113	184	amantidine	Mixed			amantidine amantadine lisa tysabri sinus adderrall cough narcotic ampyra ebv rebif moderation flares relapses norco fatigue ms infection advertised viral swelling rrms neuro lesions flu sue bladder hives flare infusion savella tecfidera cog spasticity msers spray swollen shots orange mri relapse avonex surprising dmd copaxone virus ear upset fought neuros
114	184	evaluation	Conditions		Mental Health	evaluation sct psychiatrist sertraline relaxation depressive depression inattentive medicines drowsiness atomoxetine lack query depressant treating ssris treatment disorders addressing underlying antidepressant atypical hello opinion proper symptoms commonly relieved psychiatrists medicinal diagnoses affective diagnosis motivate adhd patients tiredness bupropion advertisement nhs suffering poor tobacco discuss hence treat depressants fluoxetine ssri anhedonia

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
115	183	newly	Conditions		Narcolepsy	newly diagnosed narcoleptic cataplexy xyrem treated eds hh tips ih idiopathic wreck paralysis pwn beginner napping falling naps noob supportive nighttime yay et iud unmedicated teacher train allergies hallucinations rough diagnosis portion wheel son medicated narcolepsy teaching al fragmented vigils pregnant grad husband lortab exploring nap medically frustrated boyfriend birth
116	181	bull	Other Interventions			bull red oz drinks sized decaf monster drank shots drink drinking beers removed shake caffeinated sugar coffee subreddit thyroid xeloda cup drinker espresso birthday meanwhile valerian hardcore cilostazol bright shakes pseudoephedrine rechem drunk afterwards tumor drugreview marathon drivers tactics degrees commercial tsh benign lips tea div hypothyroidism container hangover spots
117	181	cardio	Exercise			cardio workouts lifting weights workout gym fitness endurance training exercising strength exercises exercise intensity bodybuilding running aas bike lift calorie preworkout sports marathon fats physical yoga exertion athlete dual stamina reps fat craze muscular bpc mass pre cm creatine tren weight calories cardiovascular tournaments units mile walking lean evenings train
118	180	lipoic	Other Interventions			lipoic carnitine acetyl ala acid alpha lisa supplements amantadine taurine alcar msers coq aging vitamins amino focalin vitamin supplement vit hugs tyrosine phenylalanine september gpc chest picamilon bone acids fatty neuros biloba huperzine fatigue ginkgo anytime karen lion mane complex shakes nac pains omega oil gingko dexone walmart bang vinpocetine
119	179	cyp	MoA			cyp enzyme enzymes induces metabolism metabolized inhibits induction metabolize plasma metabolites grapefruit concentrations induce breakdown pathway liver responsible metabolizes toxicity metabolizer inhibit metabolite finasteride moderate conversion inhibiting interactions inhibitors pharmacology inhibition inhibitor interaction interact bupe theoretically pharmacist absorption blocking regard suggests potentiate uptake omeprazole clinically broken inactive opioids mao metabolic

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
120	177	adrenal	Conditions		Endocrine	adrenal adrenals cortisol thyroid hormones insufficiency hormone dhea burnout levels saliva testosterone deficiencies stress induction lyme blood tsh synthroid healing endo deficiency iron adequate toll curve inflammation tumor borderline immune supplementation presumably panel bloodwork tested intolerable connected production exhaustion hormonal digestive gut vitamin girlfriend rhythm scott checked evista hypothalamus test
121	177	spasticity	Conditions		Multiple Sclerosis	spasticity spasms ap baclofen ampyra amantadine muscle gabapentin nerve stiffness legs ms pump bladder boards walking tysabri relapses progression ccsvi leg ldn sclerosis fatigue karen avonex twitching spinal tremors walk msers constipation amantidine wheelchair neuro tremor copaxone spine weakness rebif swear stretching mri diazepam videos numbness lesions muscles booze dmd
122	174	cleaner	Effect			cleaner smoother flmodafinil smooth locked washed metabolize converted stim effecting bodybuilding aggrenox supeudol tunnel pitolisant conversion dirty redose binge kicks eugeroic vote corrected wouldnt empty rates marathon balanced cog modifinil crl veggies endep hardcore prodrug drs endurance discomfort sized hydrafinil prefer carb tim workouts exertion introduced adrafinil meat haha afinils
123	173	modifinil	Acquisition		Online purchasing	modifinil modafil lupus researched ppl fee edandmore purity blister packaging rechem virus iop phenergan section reviews shed italy xx par legit doubts responsive hab onemedstore feedbacks scam legitimate checks modapharma prednisolone shipments pharma cleocin modern package site desoxyn darvon received shrooms refund boards recieved packaged ripped mymodafinil shape sunmodalert confirm
124	172	blister	Information sharing			blister foil packs packaging packaged envelope blisters strips plastic box discreet pack strip bottle package received container inside gel manufactured counterfeit tabs imgur arrived outside tpe scratch jpg indian geoff air expired marked batch pics customs pictures fake original legitimate store bottles onemedstore packages looks grab india reship modafil shelf

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
125	171	euphoric	Mixed			euphoric kratom euphoria strains wd opiate gf bitter neuroprotective himself altering discussions chasing enjoyed withdrawl opiates laugh peripheral document fever ur fatal recreational sympathetic sees estrace beers links mellow display depressant dunno dance getelementbyid stamina progressive chase withdrawals morphine withdrawal pregabalin bout bupe sober attitude count viral simultaneously relaxation posts
126	170	deprivation	Symptom Impact		Effect	deprivation deprived rested deficits impaired polyphasic sleep nights pmid executive functioning finals maintaining melatonin impair contfzzza sleeping improving wakes patterns memory severely peptides night getstring cognition recovered performance baseline function progs tryptophan litezennopost burnout hungover analytical nighter restful retention efficient fragmented impairment systematic architecture ideal hgh restore chronically habits cognitive
127	169	butter	Biohacking		Food	butter mct bulletproof grass coconut dave asprey fats fat bullet oil upgraded coffee meat recipe diet organic veggies chocolate beans bp carb protein egg diets paleo carbs changer proof milk podcast bpc hacking eggs tim fed cheese fish grain decaf rogan processed ice gross gold combine chicken nutrients nuts bio
128	169	nhs	Acquisition		HCP	nhs gp referral private gps consultant specialist prescribe appointment neurologist dmd clinic neurologists hospitals pbs referred prescribing neuro diagnostic diagnosis refer neuros asked nurse medicaid mri hospital awaiting ldn psychiatrist licensed psychiatrists rheumatologist expense ms apnoea narcolepsy relapses scan reluctant procedure local referring xx specializes minded refused appointments metacam sue
129	169	saver	Outcome			saver ritilan deny life cleaning ground miracle everywhere accupril narcotic drugreview project procrastinating goodness evenings career absolute pulmicort enantiomer dishes aldactazide timestamp induces stromectol importing crippling racemic burnt lunchtime appreciated ug flares reasoning understands unpredictable cloud strain ultimate analog laundry ty chew connections thc ppms methadone hydergine dozing acetazolamide nuvigl

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
130	166	intuniv	Spam??			intuniv guanfacine clonidine straterra strattera atomoxetine focalin adhd development reboxetine adjunct vyvanse contraindicated expired inattentive shrink costly voices add nri adults sct tricyclic dexedrine concerta dlpa hyperactive desoxyn titrating dex doxycycline odds ought patents display hyperactivity gluten patent alternatives bet pics mph stablon stims treating pi commenttext refuses obtain tricky
131	166	united	Acquisition		Online purchasing	united pharmacies scam retailer states geoff nrx uk sell estrace pulmicort trustworthy sheffield india ordered onemedstore blister legit packaging tarka england edandmore selling websites indian fraud pharma seller sold modafil site reviews consultation eu qhi sun iop avandia modifinil ship cheapest purchasing frisium worldwide prescript sales ly ject flucloxacillin sorta
132	165	interactions	Interactions			interactions interaction antibiotics interact interacts infections methadone maois kidney cough pharmacist mixing lyme hypertensive birth celexa contraception sinus responding depressants types flucloxacillin oxycontin warnings pharmacists cautious addressed toxicity rage crisis panic contraindicated itching clarify medications wort recommending established correlation alternate absorption googling simultaneously sertraline infection dizziness opioids injections tachycardia suffers
133	165	provigal	Mixed			provigal modafanil cloud lorazepam ativan chaos barely radiation ladies vigil unbelievable vice metacam petcam aas biology lifted peptides inattentive biohacking wks modern levlite climb chemo crossed accomplished hgh assure tumor practitioner understood king cheating gf unmedicated aggrenox aderall poorly bupe magazine ritilan nu teacher experiments frisium player profit autism mission
134	163	savella	Conditions			savella cymbalta fibro pace fibromyalgia overdo aches rheumy expand pain hurt wiped rheumatologist fm laurie retired contfzzza armoda limitstart hugs chronic sick gabapentin getstring progs cfs litezennopost lyrica itchy stamina remedies disappointment kamagra blessed nausea hives pounds sickness clothes aldactazide txt ranexa illness fatique ra nerve dizziness sweat dizzy spine

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
135	162	rechem	Spam??			rechem nmr lab purity ca chemist compound synthesis powder forehead seller pure mass send chemical testing anonymous batch equipment labs powders sketchy sent phenergan counterfeit netherlands trustworthy floxin smells peaks accurately tested canadian evista across eulexin refund analogue dydrogesterone conversion sellers reship vendors disclaimer legit nurses sunpharma cardiologist unsafe modafnil
136	161	sustainable	Spam??			sustainable alprazolam lorazepam xanax gut bromazepam newbie nutritional screen beers ativan clonazepam onemedstore downer stressful selank hydrocodone lexotanil utterly iron htp incorrect provigal soma cried fixing heroin benign stilnox spell comfort clonazolam cigarette downers borderline carnitine lunesta vicodin zaleplon pains cured gabapentin modanifil virus gad statement nitrofurantoin death wd tryptophan
137	161	survey	Mixed			survey users gwern gather div rs itemprop commenttext snp responses pmid participants content trazodone msg grain accidentally andme researcher sclerosis alternatively usage matters tms display contribute analysis getelementbyid gene impacts gluten entry units topic introduction ethics genotype cleaned burn ugh frequency notably aa anecdotal discussing discussions val remission enter itch
138	161	analogue	Acquisition		Legality	analogue iv dea law narcotic imported possess controlled import illegal legality country scheduled substance legal prohibited named unscheduled importing requires regulated classified laws schedule possession technically banned opioids iii dependence category federal analog lawyer established od abuse unheard ban former wada recreational problematic relative grey FALSE traveling states countries withdrawl
139	160	lyme	Conditions		Infectious Autoimmune	lyme ebv infection infections viral cfs antibiotics virus chronic mono undiagnosed autoimmune disease immune diseases kidney treated adrenal shots flu fibromyalgia cd dying sinus allergies correlation rheumatologist insufficiency iron illness born par treatments symptoms sickness exposure allergy ear ccsvi fatigue adrenals illnesses fibro cough lupus fever fl tested unrelated flonase

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
140	159	sweet	Unclear			sweet spot lovely teva favourite quarter beans darvon quarters fever tastes skipped insane inc internal beers authorization tongue estradiol reads litres rhythm pqq tense curb ranexa egg chicken caffine paperwork preparing lostfalcos completing aldactazide flying chocolate lingering ug mono cephalon september campus interacting mb tarka context darn unbearable dissolve redose
141	159	fingers	Emotions			fingers crossed climb crossing toes handed chaos augment refuse sees tingling wks refreshing hassle retail bupe remedies thailand happier boyfriend tag fail bottom desipramine sometime essay levlite mon vigil keen swallow advil augmenting flucloxacillin pool metacam outrageous frisium petcam tricks hopeful terrified cordarone accommodations excited roller crying teachers extensive pricey
142	157	curing	Mixed			curing collective libido anhedonia apathy agomelatine answers moclobemide numbness augmenting emotions augment mirtazapine brainfog reboxetine fluoxetine ssri erection experiences eric gad ssris anyhow parnate augmentation aswell horny depressed emotional sexual realized believed nri somewhat huperzine chat lexapro decreasing bupropion snris apathetic adjunct buddy phenylalanine sex galantamine antidepressant motivation noradrenaline stablon
143	157	norco	Other Interventions			norco oxycontin percocet drugreview nubain gmail aspx yahoo methadone oxycodone morphine narcotic oral ritilan hives xanax vicodin subutex fibro cordarone wd lortab bars injections klonopin benzo opiates wondered itching topamax rheumy tysabri adderrall addict flares es metacam thru pain motilium petcam bladder email savella tylenol spasms lupus lisa oxy addicted
144	156	dude	Mixed			dude idiot bowel itching wtf movements tone lesions fucking masters bullshit fuck batch poison movement partial trt showed forgive relationships waited abnormal english display ignorant tears scared olanzapine seizures rant ss rrms yo darvon indicate officially male opened joy ass junk killers truth posts tren scratch mice counterfeit teenager pics

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
145	156	patches	Smoking			patches nicotine patch gum smoker cigarettes tobacco smoking chewing chew vape smoke addictive applied smoked analytical quit cigarette habit occasion coconut addicted pieces fentanyl spray overdo dlpa quarters flush lifesaver enhancer cessation bromantane suprax dydrogesterone relaxing dad exploring sublingual forming bang maoi occasional nefiracetam coluracetam selegiline bpc ferret aim classical
146	156	tolerant	Tolerance			tolerant becoming builds nurses tolerance building stores cycled breastfeeding build sinemet eph grow built concepts creates timestamp mpa pea rapidly consecutive buildup chess hypertensive espresso cycling tightness beans cos downer skipping butalbital forehead oz path yohimbine divided rc stressful wise preworkout heavier exacerbated redose olanzapine asshole ahi vomiting pregnancy quickly
147	153	wired	Effect			wired tips spinal noon tamoxifen grogginess christmas overkill dinner sleeping bread lights settled latuda polyphasic alarms brintellix adjustment clonazolam cb bummer unmedicated geodon neurology couch stairs asthma bf paralysis straterra worn longterm restless pep quot shaking metformin tag refreshed stabilize arthritis falling tricks professor combivir drugreview adjusting weaning crashed mirapex
148	153	seizure	Mixed			seizure kratom seizures death threshold opiates documented incident combining shock recorded fatal keppra epilepsy stroke lowering risk beers opioid recommending nerve cases exertion careful opiate unsafe interacting strains preworkout incorrect toxicity injury likelihood suboxone absence died accidentally mobile damage respiratory promethazine overdose withdrawals cardiac depakote zaps combo tramadol od tapering
149	152	armo	Mixed			armo mod peaks adra responder waklert batch wakalert hab bored vendor moda aggrenox responders ran armod mirror burned intermittent anecdotes prob endep handling artvigil ime motivate purchased endurance rant batches importance paleo armoda smallest nmr considerably wtf mods cat sublingual prefrontal retailer uni marathon recieved nrx implications peak laziness concepts

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
150	152	fluency	Effect		Speech	fluency verbal speech sentences words concepts tip conversation confidence memory sentence language creativity expression recall communication skills attribute improving social logical improved word ability accounts clarity talkative sunifiram analytical speaking awkward english lowers inducing negatives lion reasoning poor bpc talking improvement memories dislike interviews rapidly clients negatively twitching inner anecdotal
151	152	smarter	Effect		Cognitive	smarter solving cognitive smart enhancement enhances enhancers limitless learning creatively tasks memory forming enhancer executive improve iq forgive enhance concentrate ability improving prefrontal disorders concluded intelligence cognition increasingly distractions et nzt picamilon abstract communicate enhancing conscious students idra skills improves focussed performance movie concepts bursts tms capacity abilities deficits debate
152	151	karen	Unclear			karen hugs lipoic amantadine acetyl carnitine msers xx christmas lisa ala error concerta luck nubrain costly laurie sarah ms sue everybody supportive killer neuros alex cancer hubby adderrall overdoing gentle acid encouragement wishing fatigue edited transaction gps mary norvir perk ginseng pseudoephedrine tysabri relapse neuro hope awhile clorazepate wishes gp
153	151	blunting	Effect		Mood	blunting emotional irritable emotions cry phd moody yup clears aggressive edgy cried itemprop irritability brilliant depressed div angry hopeless video deficiencies overly entry respect masters lawyer anger emotion fire butalbital haha awkward coupled tears cranky commenttext boy shooting curb playing hmm relieved nauseated creating vent games sadness pdocs profound joe
154	150	enhancers	Enhancement			enhancers cognitive enhancement enhancer enhancing performance improvements smart intelligence blunting smarter ect cognition memory tms students nootropics enhances nootropic versus impairment debate sub newbie enhance alzheimer emotional confusion racetams topics improving neurotoxicity limitless term nzt solving tdcs noot booster improves survey acute ethical artificial fluorenol opinions interest perception abilities chess

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
155	150	asks	Conditions		Fibromylagia HCP	asks fma fa reasoning salts fibromyalgia rheumatologist share idiopathic replying insight rheumy prosteride vit discipline existed tumor apologize dishes hydergine latest lyme tecfidera tysabri ha nubrain ra debilitating discussing gmail neuro biloba neuros modafinal kb norco spare continually narcotic vice vinpocetine hypersomnolence neurontin wheelchair chart ot trusted detailed bitartrate continues
156	150	withdrawals	Withdrawal			withdrawals taper withdrawal cravings turkey tapering withdrawl stopping wd booze quit quitting bupe tapered weaning dependency craving zaps dependence benzo ticket kratom sober addicted opiate addict transition addiction pregabalin sobriety benzos subs opiates crying nine merged smoker cold smoking slowly cigarettes successfully heroin wean exertion pristiq overdo cessation klonopin eliminated
157	149	pots	Conditions			pots autoimmune tachycardia cfs ldn sarah ra viral ebv cardiologist idiopathic dysfunction illness blocker amantadine fibromyalgia sympathetic researchers fatigue disabling exhaustion ppms lifetime beta losartan hypersomnia specialist blockers intolerance immune neuro mentions patients mayo conference bloodwork diseases diagnosed fatigued rheumatologist therapies standing hopeful rheumy raising candidate fibro diagnostic adderrall anemia
158	149	tunnel	Effect			tunnel vision blurry speech sharper fluency clearer screen gear yawning focused lifting distracted sustain ultram amped hyper verbal attitude humor ime visual meditating forehead clients drove focus occurs distractions yawn music confidence outcomes exelon bright tasks eye boring closed language permanently migraine slight bored headed intense tries dull talkative corrected
159	148	blocker	SideEffects		Cardiac	blocker beta blockers tachycardia propranolol atenolol heart rate pots bp adrenaline channel cardiologist pulse waves beat pressure monitor cardiac agonists resting asthma bystolic clonidine heartbeat palpitations blood guanfacine downer blocking racing sarah elevated rapid straterra unfair beats race barrier dilantin addressing ideal raise hypertension kemadrin raises irregular speech cardiovascular bpm

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
160	148	clonazolam	Other Interventions			clonazolam gabapentin etizolam weed benzos clonazepam nicotine meo dxm phenibut quot beers edited vape ot benzo alprazolam kava taper tonight sedative cups mcg trips genuinely tianeptine tripping rebound kratom finals opiates ate mxe apap booze gonna evening stash od benzodiazepines acid pregabalin cup cannabis stoned beer pcp suprax cigarettes lord
161	147	armod	Mixed			armod burnt cos waklert artvigil kicking sharper stays thick rock ringing flow peaks lunchtime bag alive burst creatively bummer gradual hab smoother sibutramine mod walls prescript mate adrenaline hppd unmotivated worn soul shaky rolling tool lazy highs microdose fatigued attractive engineering lifted cleaner thin software bf ear rotating curb married
162	147	deficit	Conditions		ADHD	deficit hyperactivity adults attention adhd systematic hyperactive efficacy sct inattentive disorder children kemadrin adult pmid methylphenidate atomoxetine literature sluggish mph impressive cbt depth psychostimulant bupropion lecture mindfulness nhs academic child tricky norvasc subjects strictly poorly beneficial relaxation dexamphetamine modified sustained strategies add formulary nri remarkable russian vpxl compared disorders thoroughly
163	147	nrx	Acquisition		Online purchasing	nrx md retailer uk geoff modafinilcat vendor site status stock tarka customer chose strips sheffield received united cc iop prescript shed worldwide orders debit email edandmore discreet tpe packaging onemedstore numbers ems visa modafil mods competitive tracking legit contacted cat ship items duty scam ban pr reship code feedbacks envelope
164	147	nighter	Enhancement			nighter pulling nighters pull exam revision exams homework studying deprivation assignments intensive spectrum finals etizolam repeated graduate deadlines rested autism essay attend final tonight levlite finishing skipped classes clonazolam severely semester nmr deprived workload impair broad efficiently season frisium ab intelligent gpa lecture nri cumulative teachers apathetic pulled cilostazol debt
165	146	ur	Unclear			ur ppl bout wouldnt nuvigl narcotic plz lupus pool insights desoxyn yr awhile alter iv isolated units didnt da vicodin idk hill gun wd toxic hrs abused addicting prayers epilepsy couldnt doze biaxin dont hardest karen horribly autoimmune snack bumped acetazolamide lortab rogan drugsgear tapatalk shooting ultra attitude plavix ins

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
166	146	sharing	Information sharing			sharing responder batch expense tracking flucloxacillin reminds appreciation updated update emailed detailed log suspected distinct cigarette thanks thank feedbacks subreddit sorted hello comments responders consists ongoing member identify gmail arrived enjoyed posted mayo received poster bunk cares crossed query placed joined reply nubain modavigil btc yahoo pot touched contacted diethylpropion
167	146	strategies	Mixed			strategies coping anna augmentation fragmented aap abilify xeloda ad cbt print plz australian coaster mindfulness difficulties augment dex pmid cousin bless sclerosis msers inattentive ciprofloxacin progression awhile sdmb clorazepate tca responded remission survey lisa ht heh addressing antipsychotic offset snp nightmares sounded fluorenol hospitalized bupe therapies freely viracept godsend buspirone
168	145	ability	Effect		Cognitive	ability studying welcome monk cognitive distractions curing creativity enhancement downers enhancers enhance memory programming mindfulness creative task boring dull workload tasks wakalert topics dual foundation tdcs improve clarity motivation tunnel iq sobriety suffers downer procrastination psychedelics conversations chess collective retain uppers challenging focus enhances libido horny concentrating hello glycine master
169	145	shaky	SideEffects			shaky shaking hands drawing tremor alex palpitations lexapro heart ephedra sigh nurse teachers toes modafinal spells asthma adrenaline tremors zoloft shortness voices sensations undiagnosed racing amantadine fitness worn adderrall avonex jittery ruining rebif nu ads mentioning draw legs decaf joy wheelchair heartbeat arms laundry caring resting blocker horrific beta amantidine
170	144	psychoactive	Acquisition		Legality	psychoactive bill ban law definition act blanket substances highs regulations medicinal april import legal illegal banned importing policy licensed uk grey laws misuse legality status afaik police imported medicines possession contfzzza vague substance eu txt falls progs litezennopost getstring intent emsam debate nrx established viracept named seized china under prohibited

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
171	144	dreams	Effect		Dreams	dreams vivid dream dreaming lucid paralysis nightmares hh hallucinations fragmented voices woke rem refreshing laughing waking asleep ptsd deep bizarre visual restful refreshed galantamine ibs detail nightly sp scary automatic norm eds remember corner naps endless during practicing fall spells stages falling images simultaneously hungover childhood flip odd sleep fallen
172	142	hcg	Mixed			hcg trt iu testosterone hgh clomid drugsgear injection tren gear dhea ml ai aas hormone ed thyroid test panel steroid labs amps pics vitamin injections estrogen peptides mission gel vit wierd mcg prosteride caps cycle propecia bio blood levels insulin tsh sublingual steroids ct ball restore hormones pharma commented stimulate
173	141	shots	Mixed			shots ebv iphone sleeper multiple institute plavix espresso absorb sublingual avonex vitamin lyme infection copaxone taurine adderrall iron drinks moderately clears crush magnesium tolerant remedies orlistat amantadine wet sublingually sclerosis shot orally bladder tells aspirin chewing ms marathon fatigue muscular sinus quicker drinker therapies constipation aldactazide walmart bull stiffness loses
174	141	onemedstore	Acquisition		Online purchasing - suppliers	onemedstore tracking received feedbacks aus envelope emailed iop edandmore ordered payment confirmation customer package transaction packaging blister discreet recieved arrived cc placed australia packages sample packs worldwide foil site adjustment satisfied ems strip stilnox tpe orders spier pack express btc visa sent rxrex arrives december frisium receive messages transfer email
175	140	nauseous	Effect		Food	nauseous nausea lbs stomach eat vomiting eating pounds bread hungry weight dinner gained appetite dizzy cereal soda decaf carbs food eaten bowl meal gaining shake calories ate nauseated foods yawning scoop empty chocolate meals drank breakfast drinking calorie savella dairy coconut mct protein carb apple snack hungover meat dizziness units
176	139	fluency	Unclear			fluency goodness dogs perk sooner chew aggrenox wheelchair verbal kamagra chewing films quarter oxycocet bpc journalist fu accidentally researcher boy plavix esterom levlite glucose dick viramune modiodal celexa supeudol symmetrel darvocet morphine prescript speech documentary aphthasol instructions cheese uptake alex tramadol viracept gum xl wonderful practically yikes unsafe desperately cleocin

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
177	139	ampyra	Conditions		Multiple Sclerosis	ampyra walking dmd tysabri rrms wheelchair ap relapses progression ccsvi ms neuros neuro msers gilenya foot spasticity aricept ppms relapse insufficiency grapefruit rebif messages amantadine copaxone induction amantidine fatigue spasms avonex fatique reduces drugsgear bladder steroids sclerosis infusion exelon ac pt wet formulary mile flares mri legs improving lesions losartan
178	138	otc	Acquisition			otc countries pharmacists country illegal russia dexone united laws importing thailand states india mexico modiodal import italy fee driver filled western travel depressant aphthasol flu canada unscheduled japan pilagan ponstan itch sinequan crossing sold aids imported indian spain caught inexpensive esterom pharmacy gov nhs supeudol legal fill pbs container manufacturers
179	138	nsi	Mixed		Conditions Emotion OtherInts	nsi tianeptine coluracetam motivating anxiolytic sodium idra handling inner cerebrolysin prl fasoracetam headed cured outrageous amazingly pleased stablon lovely impair nope fuzzy somnolence yikes maximize ly cumulative interestingly memantine synergistic atomoxetine treats noticing anger intuniv accomplished semax imho keen surprisingly growth bid unifiram chapter changer returns casual accupril burnout ime
180	138	cat	Unclear			cat pack sample user noticeable packs clenching distractions blister chaos relapse sociable artvigil wks exertion strips responder lifted reflux incredible modafinilcat finished pieces nitrofurantoin bpm responders enjoyed courses grab modup onemedstore hab bupe litres talkative degrees pc swim quoted climb modapharma envelope modvigil smoker uni breaks foil errors batch grinding
181	137	impairment	Symptom Impact		Cognitive	impairment cognitive cognition memory improving alzheimer memantine paranoia atomoxetine schizophrenia worsen visual enhancer benzos verbal enhancement improvements iq motor benzodiazepines counteract enhancers mindfulness psychosis deficits stablon psychotic memories sedation skills techniques improved aricept idra retention term nerves dementia instance beneficial impacting improve larger cognitively dysfunction sentence enhances reverse relaxation overkill

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
182	137	ama	Unclear			ama ying fu researcher neurology professor science writes thinks substances afaik neuroscience scientists phd qualified field stay exists addicts ss practicing dependence decade prescribes humans ghb advertisement studies she psychiatry ignorant darvocet desired depth western frisium aspects unnecessary ethical depressant ketamine existence factor pmid involved anti knowledge poster endless understood
183	137	benefited	Mixed			benefited adrenaline authoritative coaster blocker blockers tachycardia possession pool amphetamines bummer gingko ranexa estradiol guide critical ciprofloxacin beta pulmonologist frisium cc acetazolamide ginkgo mane noradrenaline godsend finasteride decline carnitine tsh pots sentences moduretic burst lipoic ritilan games attend fighter darvon abilities scott intolerance swimming winter blanket exhausting highs depleted flucloxacillin
184	136	review	Information sharing			review systematic deficit hyperactivity reviewed complaints received update placed modup wrote feedbacks edandmore comprehensive reviews spier arrives recieved spierx essay ocd introduction write reship writing letters waited peer modadropship escitalopram medsforbitcoin cloud modavigil sorted letter elsewhere awesome consists envelope doubts signed spoken modapro hypo realised purchased attention hypomanic ac branded
185	136	gwern	Spam??			gwern net modup table mymodafinil mentions reddit cod survey ships faq delivery sunmodalert www ab qhi oxycocet dual resource darvon sellers overnight authoritative trustworthy creatively discount floxin google shipped guide brugada fedex link buy goodrx itching cheapest nrx worldwide frisium linked vpxl nextday genotype ciprofloxacin utterly purchasing eriacta btc definitive
186	136	upping	Dosage			upping napping fuzzy accommodations responsive dizzy skip nodding awkward hubby alternate weekdays weekends esterom collective crying fingers pulse teachers nap savella atleast forgotten log preparing wore lb march overkill estrace climb highest biaxin limitstart flucloxacillin waited fired aldactazide moderately attractive laughing warm built feb responders remained effexor sleepy adjust cried

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
187	135	uppers	Recreational			uppers downers downer od benzos benzo addicted alters addict marijuana lunesta ferret opiates heroin etizolam psychedelics cb phenergan cleocin weed sets coaster bipap irregular crippling roller mdma addictive fits cd recreationally perk cannabis knee achieved abused heartbeat strains modern pushed taper recreational alcohol dad keen aggrenox sweaty opiate clonazolam dependence
188	135	participants	Education			participants university students ethics enhancing ac enhancers anonymous interviews london smart student importance psychology questions project performance research researcher survey media scientists cognitive interest enhancement field interview society published carried carrying record academic account professor ethical neuroscience services voice nature among article extensively graduate journal science hesitate concerns understood opinions
189	135	rxrex	Spam??			rxrex placed tracking received arrived polar medsforbitcoin modadropship service packaging bi ordered recieved onemedstore customer feedbacks registered discreet package arrive ems july sent confirmation disappointing modafinilcat exelon cc frisium shipment blister envelope communication usps edandmore levlite arrives btc email transaction payment packaged emailed vpxl receiving impressive update seasonal fixing updated
190	134	ringing	SideEffects		Tinnitus	ringing ears tinnitus ear ring loud noise hearing hurts annoying swollen painful inner existing nerves twitching vertigo pains internal vicodin modup worsen hydrocodone lipoic remeron infections sa nitrofurantoin quiet heightened sensory frequency gut comfort hillary requested lips effecting identify itching circle bromazepam nu sensation jaw wakalert soul sores fm sound
191	134	youtube	Media			youtube videos video watch watching tv https music introduction watched darvocet movie spasticity games channel podcast documentary film motivational movies creatively fpm screen commercial com talks playing carvedilol sct experimented www listening repeated limitless upgraded loop radio imgur bbc drawing tk lecture smart aricept html occasion monnieri spasms hillary browser

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
192	132	procrastination	Symptom Impact			procrastination motivation procrastinate fire procrastinating task phenyl distraction laziness distractions pi boring distracted discipline fpm schizophrenia breakthrough motivate aap stability magical oxi evil deprenyl motivational endless motivated megathread happily confidence stimulates energizing movies upcoming antipsychotic cured solved saphris grocery struggle smarter solving lack bored dopamine upgraded stabilized logic happiness levlite
193	131	favourite	Food			favourite fasting addition ladies beans tea green beers intermittent bang drinking absolute cream nu herbal visual dat meal clears combos bread curb suprax arginine butter booze cialis coconut cup fats protein enhances warm raw water redose beautiful target along calorie etizolam coffee pleasure binding oxygen oil hot calcium zone drink
194	131	pbs	Acquisition		Suppliers	pbs aus australia obtain description modavigil au license qualify onemedstore healthcare private latency australian nursing records item attached photos insurances dexamphetamine criteria mins signed importing assessment beans import nhs ah avandia gp hypersomnolence agreement trustworthy officially discounted casual lurking imported pages eu motivating engineering fee apnoea container temperature ab injury
195	131	monk	Unclear			monk master mode chess scare unknown fluency gum hypomanic getelementbyid insufficiency polyphasic green grades compulsive documentary bbc diethylpropion balls hype kidding dual psychologically tactics damaging iq ccsvi ultram tendencies cats model journalist courses obsessive nutrients concepts curve steroids worsen verbal fruits connections brief lipitor creativity considerable bpc acetylcholine incredible impair
196	131	laxative	Unclear			laxative constipation parnate nardil proved subutex marketed beats snris opioids max butt maois jim argument antidepressant bottles ironically offset snri ads augmenting shoot diarrhea maoi suboxone fails originally drowsiness rated augment paying potent pdocs copay tca costco vastly ssris newer rating approve insanely truly ad supply sa substantially sdmb desipramine

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
197	131	retailer	Acquisition		Spam??	retailer iop confirmation nubrain pricing nrx qhi legit payment usd worldwide frisium cc site bitcoins paypal email medsforbitcoin darvon published reship awaiting sheffield motilium ama included eulexin feedbacks sells seller shipping ac sunmodalert timestamp ems send cod flucloxacillin edandmore vice prescript ships onemedstore shipment website fedex trustworthy registered verify sunpharma
198	131	meridia	Other Interventions			meridia noprescription phentermine soma lexotanil sibutramine zolpidem adipex tk ambien imovane lorazepam valium codeine xanax diazepam hydrocodone sale purchase spain ativan alprazolam buy online bromazepam butalbital uk clonazepam order fioricet pharmacy visa cheap tenuate discount stilnox mastercard canada overnight tramadol ultram mexico india cheapest australia cod tadalafil zopiclone delivery tags
199	130	geodon	Other Interventions			geodon seroquel neutral aphthasol aap klonopin sedating hypomanic risperdal wears sedation limitstart saphris latuda abilify pdoc esterom trileptal zyprexa grogginess depakote trazadone stash groggy stabilizer supeudol nighttime mania bedtime offset manic noon awhile flonase lamictal counteract zombie tegretol med lithium asleep revia yawning crashes antipsychotics wide morning pm sedative night
200	130	documentary	Media			documentary films journalist bbc nightmare writes tactics article smart scare takes his bullshit film fucking magazine intellectual replying modifinal ss times piece media watched video impair swear judge ban qualified movie confusion former rats ethical definition cooking banned debate complications section compares circle argue drugs limitless existence frontal advantage blame
201	129	grapefruit	Dosage			grapefruit juice orange enzymes enzyme potentiate cyp inhibits glass hd conversion hydration breakdown meantime synergistic hydrated metabolism drink slow liver cups maximize inhibition white broken induction hypothesis interacts responsible absorbed waves cup cheese blocking thistle water poop absorption apple metabolized interactions sustain bang induces drinking inhibitors tuesday motion chicken sunday

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
202	128	phenyl	Other Interventions			phenyl oxi ani phenylpiracetam retention racetam racetams oxiracetam stacks stacked logic pair megathread pramiracetam prami darvon tweaked topics noopept impressed supps core frisium coaster aniracetam nefiracetam buildup stack endep overdoing motor cdp headed wore adds vice roller ime fire progressive curb ethylphenidate stabilize dislike separately strongest upgraded path procrastination desired
203	128	buspar	Other Interventions		Antidepressants	buspar zoloft abilify pdoc lamictal latuda seroquel geodon lithium depakote welbutrin brintellix effexor remeron pristiq viibryd xl aap wellbutrin topamax prozac celexa sedating cocktail zyprexa cymbalta risperidone bipolar tweaked buproprion titrate risperdal antidepressant augmenting ssri neutral pdocs augment trileptal mirtazapine xr luvox hypomanic benzos antipsychotics stabilizers prn benzo titrating fetzima
204	128	ah	Mixed			ah accutane fri autoimmune spots holiday wed itching mon aggrenox omg ulcers gf darvon slowed diseases tbh blind weekdays double estradiol darvocet theres survey itchy timing sigh unbelievable agreement skin lips pregnancy clorazepate gradual noprescription blurry em itch realise mate arms drugsgear raw keflex joints cycles mum died exciting tarka
205	127	spelling	Unclear			spelling verbal ebv nitrofurantoin floxin meat language sentence oxycocet flares cfs iphone endep aldactazide quoted prednisolone sized fluency hypo saved coupons ranexa tysabri word roof accomplished ciprofloxacin adipex spinal mistakes tested mentioning spell hd pls vb supporting plaquenil darvocet aggrenox english peoples ponstan flare king forgive unmedicated conclusions adderrall apologize
206	126	vigil	Unclear			vigil nu pro modiodal germany skypax cycrin drawing iud lexapro heal birth teen tricks esterom france sclerosis trade avonex paranoia aphthasol cozaar netherlands elite possibilities modafinal vigils art provigal esports needless flucloxacillin orlistat sweden acupuncture brazil sdmb laughed employer dear xl shoes facts contraception interacting poker medsforbitcoin la ephedra boss

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
207	126	authoritative	Unclear			authoritative guide approx darvon gingko citalopram gwern beginners scored skeptical prosteride comprehensive lowers antagonism merged coaster ingest profiles synthetic parkinsons essay frisium cordarone eight answered histaminergic raises adrenaline apologize utterly statements discussion levlite ciprofloxacin elevated anecdotal ratings evista mentions acetazolamide iop estradiol spier meanwhile emailed ginkgo cipro poster thc generation
208	126	depressants	Other Interventions		Antidepressants	depressants depressant anti ar snri snris cymbalta cumulative stabilizers tianeptine incorrect ssris couldnt mile cbt fresh antidepressant psychiatrists mistakes nortriptyline stablon ssri ads grain antidepressants ponstan wellbutrin tool foot lexapro guidelines ibs wort annoyed pristiq taught sodium communicate addressed pilagan evenings tricyclic surprisingly bupropion recipe silicon somebody bless breastfeeding impair
209	125	freely	Enhancement			freely students training medical among enhancing available performance dual student multiple enhancers difficulties professionals abuse instance dexone physicians university aim incorrect ethical survey essay taught field russia authoritative education exams eriacta subjects abilities doctors recreational academic animals applies player mindfulness rates significantly cheating chess suffers abused dependence popular uncommon pilots
210	124	tension	SideEffects		Headaches / Tension	tension forehead neck tense muscle tightness shoulder muscles jaw chest shoulders subsided aches sensations headaches tight spasms muscular head pains relax headache painful exercises arm pressure dull ache tingling clenching hurts ears upper migraines pain hypotension surgeries stiffness dxm sore nerves pounding spine joint knee attributed knees secondly adrenaline imitrex
211	124	mymodafinil	Acquisition		Online purchasing - suppliers	mymodafinil modup net medsforbitcoin font sunmodalert customers dot btc nubrain qhi gwern bitcoins received tracking rxrex arrives ships package affective refund walls cod shipped packages packaging july envelope ship geoff feedbacks customer delivery ordered onemedstore site arrived patch shipping supeudol reddit scam vendor reputation blisters placed legit prednisolone february shipment

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
212	124	buproprion	Other Interventions		Antidepressants	buproprion xl wellbutrin sr megathread consequences washed bupropion welbutrin laundry effexor atm fluency unsafe member drugged endep ulcers celexa crashed hydergine ds sensations cleocin answering videos ime mpa motivate speaking apathy journalist perform medicating nonsense vinpocetine lexapro zyban replaced immensely ssri demand considerably eph tamoxifen imbalance unmotivated aldactazide awakening medium
213	123	va	Acquisition		Prescibed Denied	va emailed osa private acetazolamide shrink referral clinic claimed pulmonologist facing formulary disabled prescribe tpe duty medicaid surgery tag disability addressing ptsd asv army restrictions appointment medicare denied contacted service failing appt war primary apnea specialist rep ampyra union emails nuvigl chair refuses agreed medicated email quoted obstructive prescribing pcp
214	123	snorting	Dosage		Routes Timings	snorting orally absorbed bioavailability sublingually snort swallow soluble absorption crushed bloodstream oral nose administration crush route sublingual crushing quicker faster ingestion tongue ingest fillers dissolve bitter digestive stomach stream empty duration snap suspension enter burn coke redose determined cocaine gi onset hits tastes nasal inactive shorter gross mph approximately peak
215	122	energized	Effect		Energy	energized sobriety charm chess beautiful obsessive refreshed olanzapine ot asleep sleepy wakalert yawning bedtime groggy drives zone quiet burnout foggy afternoon evening cordarone geodon pattern dependence treats limitstart cereal drowsy challenges heal smoked meanwhile saliva redose sociable worn ying alters pilagan parkinson negatively saphris crushing tired accupril energetic forskolin nap
216	122	fever	SideEffects			fever flu shoulder motrin scored zaps mono cold itch infection viral needles hypothyroidism tsh rash coincidence allergic throat advil itchy nose sjs color vomiting sore sick rashes hydrochloride burning ebv swelling arms shortness sickness infections woman broke hands arm immune contfzzza round police oct cough swim win warm swallowing zyban

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
217	122	pea	Spam??			pea yohimbine hcl selegiline deprenyl synergistic stacking phenylalanine mao ephedrine ec preworkout epa stacked tyrosine biloba vinpocetine craze blend ginkgo crisis extract potentiates dha carnitine phenylpiracetam dmaa bro centrophenoxine neurotoxic hypertensive methyl stack supps builds swing boosting artichoke balls nefiracetam phenibut uridine huperzine holy rhodiola conversion adrenals dlpa taxing combining
218	121	shortness	SideEffects		Respiratory	shortness breath breathing deep chest smelling tightness heart breathe oxygen heartbeat palpitations osa cardiologist asv snoring odor cpap pounding sweating machine pulse throat pressure experiencing deeply compliant air pap pulmonologist stairs swelling beating sweat experienced apneas undiagnosed tingling trouble mask bipap difficulty vivid surgery holding irregular bpm yawn clorazepate weather
219	121	nuvigl	Spam??			nuvigl aggrenox orlistat bipap mct italy accutane vs institute researcher neuroscience cousin butter aa automatic stromectol rip neurology hab uppers dave copay fu adderal perscribed armoda drs limitstart interfering san promotion visit ciprofloxacin modapharma cried acupuncture upgraded bpc claimed hack accupril podcast loved heads bf hydergine ebay pdoc vibramycin lasix
220	121	hillary	Conditions		Parkinson's Multiple Sclerosis	hillary parkinson sclerosis alzheimer excessive multiple parkinsons emails neurological treats revealed policy email label treat disease aspirin among scientists hacking relation additionally diagnostic suffering biohacking article seizures students rotating patients sleepiness military elite title uses somnolence records originally conditions methods clinical athletes colleagues foreign decision addicting department pilots missions modanifil
221	120	tactics	Media			tactics bbc scare nightmare documentary journalist films smart writes article media confusion annoyed piece applies monk drugs advantage ban poison endep la news aim banned film engineer character model magazine professionals concept document ject hype de his master ss society chess outcome biology argue getelementbyid debate modifinal population cooking hives

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
222	119	gluten	Food			gluten dairy carbs intolerance diet foods eliminating diets veggies carb paleo sugar meat vegetables items bread meals guanfacine fruits keto allergies killer fats cheese consume cooking ingredient orlistat eliminated elaborate grocery plaquenil cuz lipitor fruit weigh merely msg chicken discussions contain itch gi debilitating chocolate dietary counting binding pwn infections
223	117	nubrain	Acquisition		Online purchasing	nubrain customer vendor qhi rxrex mymodafinil service medsforbitcoin payment sheffield received modup placed iop packaging ship packages customers tpe ordered recieved bitcoins tracking refund site powdercity email ships cc legit package reviews vendors parcel reship foil scammed confirmation websites receive orders arrived edandmore fraud onemedstore sent btc feedbacks shipment discreet
224	117	retention	Effect		Cognitive	retention memory learning learn concentrating dual behaviour book recall motivational sorta phenyl techniques mindfulness forgive memories detrimental swallowing fruits erowid retain sa concepts remembering studying reading trial finishing lesson frisium idra visual improvements chapter prl exams sessions intensely climb cleaned verbal english creativity improved books boosts gingko intellectual upgraded enhanced
225	117	finds	Unclear			finds valley silicon ccsvi curve methylphenidate psychoactive lostfalcos executive marijuana pot workers regret phendimetrazine circadian reset interviews impair uni magazine liking laziness aggressive believes inclined artichoke regulated impulsive definition tryptophan medicinal socially considerable appeals ra aggrenox designed driven stoned brugada doubts outcome darvon improving experiments bills intellectual approximately strains compulsive
226	117	appreciation	Unclear			appreciation humor music psg deprivation philosophy architecture burst dirt latency arousals enjoyed bpd medium neuroscience loosing anxiolytic suprax mslt arousal army favourite sustain ending relaxation anyhow sincerely nod periodic relaxing dual listen concepts ingest rem listening objective rls blessed path youtube deficits taught girl neuroprotective soul pages font lucid pursue

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
227	116	cons	Mixed			cons pros weigh rating stories rheumy hcg breastfeeding physicians sunifiram evista continuing pages galantamine aggressive liking elaborate cipro noprescription vpxl adult chart paragraph advocate comprehensive aderall dec nail tren antidepressants chess children relapses concerta dangers revia hgh sleeper dave enhancers dexone dramatically rheumatologist exploring competitive exelon questionable lasix worthwhile thoroughly
228	116	socially	Effect		Emotions	socially awkward dates conversation conversations clears social masters aggressive augmenting sobriety stoned meetings circle buddy ativan gf chaos sober sociable sentences blunting nicely burn enjoyed messes phd wine cheating fellow lorazepam calming cigarette writer edgy expression solutions relate tendency improving overly died calm microdoses frankly situations chill patterns energizing stressed
229	116	lazy	Unclear			lazy psychoactive physically discipline laziness active methylphenidate fault society behaviour management teach habits bill phendimetrazine mindfulness tool therapist development ban relaxation mentally grey beneficial dysfunction neurotoxic floxin fitness responsibility logical preventing debate modern gain nowadays named hppd variety illnesses humans adults circumstances develop cognitively psychiatry learn substances advantage ject animals
230	115	polar	Conditions		Mental Health	polar bi downs bipolar pap vyvance mania manic psychotic ocd registered gad medicating goodness depressants seroquel rapid stabilizer stabilizers depression depressant mixture trileptal cocktail blown diagnose psychiatrist modanifil react illness recieved realistic disorder resistant cycling packaged deny anti rxrex foil psychiatric latuda straterra newly depressed isnt adhd personality ptsd doesnt
231	115	anybody	Unclear			anybody rls institute pregnant pregnancy accutane wed fluency breast trt joke scares child brazil flucloxacillin shed clears athletes coincidence cerebrolysin anecdotal prednisolone imipramine twitching emotion cialis outlook fri kidney beware trade mirapex sincerely clorazepate nursing meo born community depressants keto failing diets excuse insight preworkout withdraw somebody bummer infection treats

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
232	114	weekdays	Dosage		Temporal	weekdays weekends skipping holidays skip shorter vigils security reset phenyl terrified medicate tolerance alternate breaks alarms friday alternating sunday nightmares accommodations oxi develop progressive booze planned savings xyrem nap greatest coaster nighttime climb revia fire ih meal procrastination extension strict campus manager twenty overdoing holiday monday relapses injection break bupe
233	114	tapatalk	Spam??			tapatalk iphone sent sleeper binge saphris dexone drugsgear darvon pdoc revealed grand emsam trileptal desoxyn america plavix shots drove addressing vyvanse ebv yikes diethylpropion espresso impression brintellix topamax suprax expression aap lithium multiple institute ec depakote ptsd concerta remeron latuda viibryd hillary tren discussed using documentation jim emails tweaked buspar
234	114	viral	Mixed			viral amantadine msers ms hyperactivity ect fatigue deficit label narcotic relapses cfs focalin management medications karen relaxation therapies treatments students goodness neuros phendimetrazine neuro treat plz anti enhancers progression adderrall spasticity persons sclerosis licensed narcotics resistant treatment approved ampyra ccsvi referring mylan aricept symptom forms depressants gps tenuate recieved ldn
235	114	poop	Unclear			poop wk drank explained inpatient pounding merged hangover euphoric mirapex tense coaster elevated bullshit espresso washing abilify tca rude parnate ect screaming kidding mxe fma darvon sigh someday fa tms augment forehead freak proves cups roller sober butt ebv dxm mayo admittedly believes subs pray meo oct ccsvi suck bathroom
236	113	love	Unclear			love symmetrel acting addy loved aldactazide climb adderrall locked asks touched narcotic tag vyvance sad unifiram teacher darvocet park euphoria custom hugs west triggers dishes cried fma ritalin mxe sustainable virginia vape nighters downside addict citalopram smoother tweak comedown understands ha perk acetazolamide tecfidera blah alex motilium wow ethical ritilin

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
237	113	reasonably	SideEffects		Rash	reasonably rule serious sjs johnson possibility coincidence discontinued discontinuing levlite precursor priced rash stevens monitored ih suicide slightest respiratory anastrozole spells flucloxacillin onset mslt subjects impacting periodic latency la adverse flu term pulmonologist fatal architecture scheduled cleared music spine tysabri isolated compliant lesions terrified detailed eeg bound aggrenox itch allergic
238	113	ritilan	Effect			ritilan drugreview pemoline bowl mellow cereal waves miserable accupril wouldnt adderrall pulmicort lessened killed slowed switched yrs patents rheumatologist lbs outcome fm pregnenolone expand aldactazide cried spacey worsened shortness duty randomly wheel treats retired blessed discouraged reacts waited motivate twitching omg assure attached collapse subside egg wada darvocet moody norco
239	112	medschat	Unclear			medschat ingredient dizziness contains getelementbyid title nausea conditions modafanil htm wiki document lists display provided responded ad forming insomnia details nervousness page include listed laurie shift experiencing classified eriacta excessive comments pdf description plavix exceed org commonly sensation discussions explains provigal hello endep gather learn common https below obstructive pressure
240	112	finals	Enhancement			finals gilenya exams studying clonazolam students preparing college upcoming ot student sobriety physics exam successful interview semester season grades centrophenoxine workload revision stash authoritative huperzine nurses litres scrip reality meh error frisium gingko language assessment failing project flucloxacillin nutrients cognitively mxe abilities ranexa concepts practicing programming grogginess tier beast gather
241	111	sclerosis	Conditions		Multiple Sclerosis Neuro	sclerosis multiple ms students neurology amantadine researcher avonex hillary orlistat fatigue aspx fu plavix copaxone professor parkinson therapies debilitating treatments progression rs hopeless org drugsgear training neuros wet ra revealed national science survey geoff msers ampyra ebv relapses sales scores ldn burning rrms introduction wheelchair snp url gluten intolerance flares

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
242	111	holy	Mixed			holy anxiolytic wow depressant dude session sleeps laughing micro scripts sh ncbi bf shout panic casual lion attacks filled shit paranoid nih cleaned monster handful sociable anxiogenic marathon shoulders pmc fucking shaky crap rolling popped tad loosing foggy sooo pubmed lazy hanging trazadone walgreens zolpimist girl hardcore dad experimentation headed
243	110	sobriety	Conditions		Addiction	sobriety sober aa addicts challenges breaking na monk getelementbyid rehab guilty addiction caffeinated cravings behavioral wd struggles challenging meth intention meetings addictive addicted withdrawl nine european methamphetamine stromectol abusing discover style recovery implications sodium rrms candy recreationally document guilt army resource addict oct cycrin fentanyl relevant dependence recovering smoked euphoric
244	110	depressive	Conditions		Mental Health	depressive hair depression episodes bipolar txt episode contfzzza getstring sadness progs litezennopost maoi depressed manic mdd parnate darvon suicidal ect hospitalized loss atypical saphris depakote aldactazide insight apathetic advertisement tms sa participate topics major spoken tca stabilizers adjunct progressive luvox file stabilizer maois illness div age crisis hypomanic obsessive hello
245	109	modiodal	Acquisition		Name variations	modiodal mexico alertec vigil modapro countries dollar imported gmail netherlands modavigil hydergine italy russia france french olmifon modafil japan sweden counterfeit name bought germany generics brand sold movements verbal brazil spier english brands manufactured european indian pro modalert spain mylan pharmaceuticals skypax names chemist united americans standards american modvigil pharma
246	108	jayster	Unclear			jayster roller touched prescribes qhi sdmb prosteride albeit coaster recieved guilt hospitalized describes nine maker elderly sunday sweaty quetiapine feb distance thirty hoops trileptal mile spacey sinequan mylan modanifil mgs antipsychotic adderral effecting crashing modafnil hadn nevertheless ride olmifon arrive prevents flucloxacillin noprescription scammed discontinuing affordable aggrenox wrote geodon toll

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
247	108	journey	Conditions		Sleep	journey hypersomnia idiopathic sucks melatonin diagnosis baclofen hypersomnolence ih diagnoses corner eds mslt dx tweak teenager figuring newly discover diagnosed tech deserve chart nubrain rrms dmd tremor diagnose eulexin rant rem sadly narcolepsy indications truth error spinal walmart struggles root proven cataplexy underlying googling lucky naps specialist gad facing proved
248	108	executive	Effect		Cognitive	executive intelligence iq intelligent memory deprived improving function cognition improve cognitive functioning chronically improves individuals prefrontal abilities val genotype indirectly functions frontal deprivation cognitively performance skills improved impair ocd enhances processing enhance math cortex impaired deficits comt complex attention concentration eric measuring fixing term ability motivate supplementing stacked genes brain
249	108	aricept	Other Interventions			aricept alzheimer bystolic shifts aldactazide students tdcs cognitive existed exelon tms impairment effexor ward narcoleptics memory viibryd bitartrate losartan concentrate enhancers enhancing cranky ect marketed tenuate candidate viracept dementia amino tysabri emotions abilify minded performing esterom diethylpropion propranolol prozac flonase narcotic pots immensely intelligence coaster workers buproprion eriacta amongst requip
250	107	medicating	Effect			medicating hardcore self medicate tms silicon motivate improving executive habits nu lectures cog outlook perceived valley motivational happily exertion participate orlistat recovery prevacid indirectly behavioral mindfulness gather anhedonia circle joined depressive shame referral psych exercise illnesses route polar stablon assessment rogan practicing bi suffers charm regimen meanwhile exelon apologies birth
251	107	sheffield	Acquisition			sheffield judged customers payment modafinils email received gmail scam customer emails placed orders retailer worldwide nrx service geoff recieved paypal code recorded visa confirmation mastercard emailed genuine tracking friendly medsforbitcoin uk supplier kemadrin jerk delivery united site nubrain fraud contact prosteride discounts reputable number spironolactone union btc status services arrive

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
252	107	safely	Unclear			safely manner observations sertraline lortab notes dependence indirectly frisium remarkable pathways imovane def passing adjusting lion darvon wound rogan keppra qualities processes burnout pitolisant eaten tren yo quot neurotoxic train effectively beginners afterwards aas cardio joe lb regardless miss sam modaf zopiclone incredible deficient decaf viable financially cos meal settings
253	106	changer	Outcome			changer game metacam petcam pics connect suspension oral challenges reflux amitriptyline absolute athlete physicians neurotoxicity masters flmodafinil evista lisa pseudoephedrine unhealthy sinequan neuros swear phd sharper playing gmail pump park happier consensus prosteride dig turning impacts retired xeloda lawyer struggled pace treats toes cog illness keys lifesaver marathon considerably stamina
254	106	hassle	Unclear			hassle responsive btc awkward bitcoins scares cc accommodations fingers feb brilliant fellow activated season repeat ly reward essay classified gift transaction charges unheard charged border debit losartan extra sometime responders somnolence software attend cancel frankly transporter meantime york supeudol advised merged glycine campus bucks protection qhi requirements slower mile trustworthy
255	106	casual	Spam??			casual shout lovely nsi progressive libido host ferret amazingly porn moclobemide apathy estradiol latency aldactazide ranexa dmd darvocet radio burnout facebook accomplished sex buddy nine rrms heads norvasc spending anhedonia relationships pleasure pbs se laugh spinal lucid conversation league combivir intolerance unfair yo inner psg bar growth podcast hill laid
256	106	lostfalcos	Spam??			lostfalcos tulip pqq experiments pregnenolone laser extensive coq artichoke pde ciltep ebay zembrin biohacking oxygen potentiates glucose endurance cerebrolysin extract tdcs inhibitors protocol neurofeedback synergistic stacking improvements timestamp clarity racetam prefrontal cortex nootropic wary longecity creatine spot booster favorite stack experiment outweigh forskolin variables curcumin overstimulated stacks epa plate forgive

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
257	106	tendencies	Unclear			tendencies cos viracept adrenaline aggressive analysis bothered cbt california milligrams asshole rip exacerbated cd moclobemide sincerely nutrients boredom stabilizers hooked mono nmr lostfalcos depleted partial aa ocd scoop eulexin sobriety stabilizer aggrenox cb existed hypomania dual bystolic lacks trileptal loving quantity tightness agonism utterly king bpd notably rates volume compulsive
258	105	erection	Effect		Sexual	erection libido sexual sex porn women horny arginine anhedonia men anyhow testosterone arousal pleasure viagra female nri girlfriend eric dysfunction drive emotions cialis attractive collective curing bupropion male dick clomid moclobemide desire reboxetine noradrenaline suffering wreck nardil sertraline struggled hormones confidence emotion spectrum gad ht girl recipe man augmenting culprit
259	105	hoops	Acquisition			hoops jump provider jumping thru medicaid authorization lipitor insurance jumped submit jayster aggrenox odds approved forms progs litezennopost getstring denied filling companies scripts coverage contfzzza approve company motrin spain sdmb agent rip refused txt ssdi handed copay considerable fought americans sheffield tap isomer desoxyn maker aricept zyban dollar paperwork california
260	105	analytical	Effect		Cognitive	analytical tasks performance skills drawing programming improving improved verbal decreases task engineer creativity fluency smarter complex logic cognitive improve concepts ability abilities iphone flow sentences iq deprived awakening solving motor intelligence engineering memory mindset reasoning deprivation inner baseline conclusions pmid blank completing enhancement improvements freely enhanced focus executive forgive clarity
261	104	tremors	SideEffects			tremors mirapex spasms boys shaking bowl ritilan horrid shakes amantadine tremor selank requip savella swings amantidine twitching epilepsy asks blunting impairment retired bladder acetazolamide pilots ap scott battling vertigo ads ampyra shoulder joe moody amazingly spell infusion video doze yrs muscle fatigue restlessness aldactazide fighter doxycycline moduretic rogan kava miserable

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
262	104	gmail	Spam??			gmail email judged contact yahoo spray sheffield nasal pics london dexamphetamine norco drugreview received sunpharma dot youtube hydergine hi pulmicort twitching outlook ac medsforbitcoin dramatically david itching facebook promotes hello movements confirmation introduced nubain reply july eriacta message antihistamine sun flonase presented newbie modiodal percocet bowel packs managing https physiological
263	104	sunifiram	Spam??			sunifiram excitotoxicity idra unifiram glycine longecity nefiracetam coluracetam led pramiracetam inner alleviate refreshed racetams persistent calming remotely oxiracetam experimental attempt combined fluency basis comparable beyond amph intolerable experimented remained nsi discontinuing minerals phenylpiracetam tinnitus furthermore tianeptine verbal ltp pros bpc screwed stacked alcar clarity favourite toxicity doses hppd visited od
264	104	boosting	Effect			boosting hype disturbing modafinils science monk techniques article journalist master iq benefits technology enhancement tactics fighter writer enhancing retailer cognitive radiation jayster nzt mode quoted deficits writes scare norvasc podcast acupuncture abilities craze chasing judged scientific operating cycrin date behaviour documentary unknown alleviate improvements tendency ebv films balanced intellectual intended
265	103	noob	Unclear			noob et newly al beginners beginner gotta jerk buddy ice beers forgive mad shoes ride imipramine organized viracept train ai lord laundry cream paralysis gun treated stabilize dishes feet humor smoker vpxl stacks flucloxacillin adults coaster relationships chess vote shock duty fun polar recipe hallucinations roller rude nerves son homework
266	103	longest	Mixed			longest stayed ying fu slept western hours beers havent geodon sleeper ama awake neurology hallucinations union ssdi researcher night pd fasting withdrawl consecutive fallen hospitalized wasnt rd born effected bus binge estrace hh limitstart pregnancy nights rough relationships adjustment crying dinner meal apneas hypotension hrs cured grogginess distance winter awhile

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
267	103	replacement	Spam??			replacement testosterone original hd her device hormone senior herb sex psychologist french seek records neurologists walgreens younger experts switching plastic herbal likelihood promethazine yr transaction fu ama trt orange bi texas she male monitored laid iud rogan retired erection solved poster hcg age declined researcher narcotics hillary evista cozaar performing
268	102	armoda	Unclear			armoda savella contfzzza progs getstring litezennopost txt amp moda army sickness realise ache noticing secondly hcg cumulative pains noprescription tomorrow resources hitting file FALSE itch workload washed married marathon build disappointment recreationally hang estrace hyped meridia impair progressed positives returns pace pop dick troubles procrastinating megathread task retired fluid shall
269	102	libido	Effect		Sexual	libido collective curing erection decrease anhedonia horny sexual porn sex expression answers da esterom relationship hormones emotions dopa antagonism gaining requip decreased agomelatine fibro acetylcholine ht agonists curb apathy drive women sociable savella restorative nri girlfriend aphthasol oxycontin ra tournaments cereal increasing pregnant ignored downsides inpatient attractive esports estradiol desipramine
270	101	burnout	Symptom Impact			burnout sleepless returns adrenal john adrenals skipped induced multivitamins impair nsi zaps chapter weekends exposed ying polyphasic programs assistance vape ly refreshed join egg stream bpc practicing ptsd excessively fu stacks fm employed cb cures sustain hypertensive hydration loop recover epa beast treats wort boy cycle building ginkgo dinner sadness
271	100	worthless	Symptom Impact			worthless qualities bottle hands animals def wouldnt nubrain crack polyphasic addicted booze pd rated recreationally inc waste dick army riding bother vicodin bus crystal judged terrified ptsd snorting apartment drugged tms fixing meo etizolam addy threads promising sinemet modified dmt dry numbness dozing kid nerves feb employed aderall door bored
272	100	spacey	Symptom Impact			spacey step user heavy litres lethargic wash bread stairs metabolite yawning drained weights underlying meditate groggy lingering shy burned potentiation crave glycine mile inhibitor releasing rls ditch meditating sodium nmr lurking programs excitotoxicity utterly nubrain blah foggy fitness heading fog spaced decided scores profound swallowing medsforbitcoin noise briefly insane gym

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
273	100	blurry	SideEffects		Mixed	blurry sensitivity vision sharper numbness migraine sensation dark smelling visual odd versus intolerance humor neuros migraines lights heartbeat nu nrx headaches color shaky spasms weakness eye debilitating tips technology bright ladies screen karen ringing rebif tunnel painful titrating stiffness clears needless potassium experiencing dizziness extreme brainfog md cell nauseated lurking
274	100	convince	Acquisition		HCPs	convince prescribing prescribe viramune rheumy hesitant label savella prescribes rheumatologist bio vigil fibromyalgia celebrex fibro powdercity pea identical reps harmful cozaar psychiatrist narcotics psychiatrists arthritis pbs autoimmune prodrug criteria addressed revision discussing gram tag viable outrageous supervision aphthasol fatal psychedelics pro declined practitioner comprehensive refills purposes patents shift ebv loving
275	100	thistle	Other Interventions			thistle milk liver enzymes mitigate conversion taxing norco oz raw coconut curcumin shake percocet adrafinil metabolize observation gross protein ice fatty toxic juice eliminating heavier processed enzyme amazed litres rise damage metabolized supplementing nac potentiation kidneys organic stable veggies equal toxicity metabolizes decaf bothering vegetables bars breast raises outcome lifting
276	99	wine	Alcohol			wine glass glasses drunk red beer visual drinking alcohol drank hangover drink bull beers booze vape dinner wearing car bottles plastic aderall clarity blurry drops sober ride vision water awkward birthday gf moved circles endo bottle freaked soda poison fell container oz estrace drinks drove boyfriend suck noon brutal cigarette
277	99	beginners	Information sharing			beginners beginner noob guide buddy darvon lurking frisium ratings humor train definitive subreddit likelihood assessment nubain pics reddit jerk vb pd established ride monnieri newly effectively length china ingest onto nightly french em intend vote powders prescript safely alleviate suprax continually problematic paralysis teaching hoped od fortunately chime il lacks

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
278	99	jet	Mixed			jet lag flight travel missions flying arrive fighter pilots traveling associated split soldiers popping cozaar johnson combat drained clock inc military sonata avandia carry frisium land shift levlite zone presumably participants deadline shortage viracept ems shifts inhibiting air assist tenuate wide apnoea requiring indication men driven broad solely relapses intensive
279	98	japan	Acquisition		Online purchasing - legality	japan japanese importing imported import pemoline caught police illegal customs status possession packages unscheduled legal prohibited germany law countries country eu seized possess legality dexone parcel border letter goods laws australia ireland opened letters destroy analogue china modiodal fee american shipment grey shipments charged recognized bringing powdercity thailand india foreign
280	98	sdmb	Unclear			sdmb muscular ciprofloxacin subsequent motilium chemo voices born breastfeeding baby aphthasol sedated swing esterom pays claims obstructive medically addy supeudol weren jaw childhood dad surgeries breakthrough component adrafanil ibuprofen pulmicort refills ec hoops notably asthma smarter fpm triggered documentation voice federal attribute functional outrageous animal law aggrenox overload tricks qualify
281	98	acids	Supplements			acids amino minerals fatty occurring vitamins phenylalanine carnitine essential acid acetyl taurine tryptophan protein tyrosine supplementation supplementing chain nutrients lipoic supplements shakes arginine conversion glycine precursor converted supplement ensure creatine peptides sufficient supps fats htp acetylcholine dopa bystolic systems vitamin epa zinc cos impacting separately deprenyl nac nature blocks carb
282	98	code	Spam??			code ab reputable coupon judged vpxl discount discounts received nrx prescript sh worldwide prosteride html packaging stumbled stilnox placed exelon net blister lamisil discreet sheffield alabama onemedstore cod edandmore mastercard pharmacy delivery noted de lurking fedex visa modafinilcat order geoff granted adipex btc overnight ht crossed strips palpitations tk eriacta
283	97	quantities	Acquisition			quantities bulk shrink large larger checks ying masters rcs importing fu aus unscheduled australia pbs ject isomers ticket import zaps government btc ama chase brazil unheard rich tarka worldwide items handling charged sunmodalert ap hassle goods turning arent alabama reship quantity sheffield credit skypax swallow amounts lamisil fee produced digestive

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
284	97	ar	Mixed			ar argument impair cumulative tournaments plasma armo everyday buildup blunting redose grain cessation fails suspension sufficient dosing washed apathetic evenings kidding ranexa depressants daily overload ug anecdotes peaks microdoses endep pot asshole nighter deprivation irritable esports established meditating periodic hangover armoda progressively dog se irresponsible presently wound combivir screaming returns
285	96	dirty	Unclear			dirty secret round prescript evil model poison et noots clean bystolic chinese nz wreck connections completing keto mindset steroids ignorant methyl challenging idra modern rat tears ultra exception profiles lives performance laundry norvir society burning dick protective bromantane ridiculous podcast responsibly enhancement design dirt valley producing addicts anastrozole lets differ
286	95	sertraline	Other Interventions			sertraline nortriptyline escitalopram calories zoloft ssri affecting fluoxetine curcumin remeron ocd lion seroquel latuda buspirone agomelatine ssris zaps sensations bupropion xl risperidone venlafaxine gf cereal pregabalin twitching brintellix dave trazadone abilify dhea podcast tren stablon remarkable advertisement shaking unnecessary sexual strattera culprit depression butter cortisol quetiapine pristiq augmenting energizing tapered
287	94	elevated	SideEffects		Cardiac	elevated rate began heart pressure tachycardia pulse bp fancy culprit intolerance cardiovascular resting spells remained breath blood pots sensations sharper arms dizzy cardiologist humor peripheral cardiac assumed hypertension bpm potassium pounding standing led rapid suprax ticket monitor typing incident bupe shortness achieved beat palpitations correlation chest struggled warm crying wheel
288	94	web	Information sharing			web hello site facebook vpxl service biz cordarone messages reviews nhs kemadrin searching parkinsons page pages customer medsforbitcoin association approximately idiopathic pots worldwide portal revision message cipro excellent flucloxacillin injury ject tiny tracking discreet ratings avandia cialis nubrain national design mobile website closing friendly received edandmore feedbacks getelementbyid query sites

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
289	94	meditating	Other Interventions			meditating meditation mindfulness meditate ject practice techniques practicing sustain glycine notably tdcs english neurofeedback development sorta relaxation yoga exercises masters repeatedly habits dual lacks taught behavioral ac cbt hppd memories continuously sentences psychology philosophy sufficient hygiene adjusting newbie calm progressive capable powerful psychologist controlling worthwhile model learning habit stabilize cracked
290	93	reminder	Mixed			reminder friendly ritilin forehead restful horrific wednesday burst eriacta xyrem surgeries surgery hyped needless zaps wed fri cog mirror spine arthritis ineffective nonetheless blurry releases untreated surprise methadone conference spells comfort masks tarka apathy hopeless drugged godsend enormous commit mon grogginess groggy anhedonia birth relieved joy snoring alarm font travel
291	93	dnp	SideEffects		Appetite	dnp eca sibutramine tren appetite suppression sides hgh dmaa lbs fat hunger hcg ephedrine lethargy bf gym weight neurotoxic suppressant lean gains ed ephedra fasting cravings ciprofloxacin summer hungry aas meridia cycle weights sweating fa calories cream lb loop phentermine preworkout minimize session loss lose calorie workout amph lifting ime
292	93	saphris	Other Interventions	Conditions Other Interventions	Mental Health	saphris aap seroquel geodon pdoc depakote lamictal keflex neutral bipolar sedating stabilizer lithium abilify latuda manic zyprexa mgs parnate klonopin mania antipsychotic gad residual episode fetzima depressive prn levlite hypomanic shy trileptal eric groggy compliant pristiq antipsychotics jim awakening drowsiness goodrx tapatalk follows fragmented joined ii stabilized itemprop hangover sedation
293	93	killed	Mixed			killed inflammation inflammatory mdma meet viramune curing yo vigils collective asshole animals xyrem cried blunting htp hypocretin teen gram incorrect millions flush ibuprofen selank auth gmail medically lsd fasoracetam rolling na tears wasnt coke rep sooo snort dad suicide autoimmune wouldnt quoted labs named answers accurately curiosity died criteria miserable
294	92	roller	Mixed			roller coaster ride crashing tweaked riding exhausting emotional lunesta rough grand prosteride norco focalin tsh irritable ladies guilty noon zaps poop meh downs climb nightmares willpower jayster newly moody authoritative anxious offset jan sdmb ritilan elite someday energy vicodin cranky freely lord mxe heavier oct kicking guilt nubain finasteride adrenaline

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
295	92	ly	Unclear			ly keflex meat crave bread eggs approx mobile assistance inhibitor programs fedex worldwide cod carbs classified dexone vpxl impaired raw cognitively cheapest cleocin delivery frisium overnight consultation chemical endep chocolate alertec infections qualify presumably program dairy user breakfast photos discount online floxin cheap feedbacks sunlight hungry premarin structure norvir rs
296	92	dexamphetamine	Mixed			dexamphetamine un de la sa forming prosteride au nu sweaty gun dex mindfulness kidding pbs originally fioricet picked pot rls suspension nortriptyline abusing petcam mum jim youtube psychologist reluctant es reply metacam beginner tournaments tremor scare consultant augmenting propecia amazed exelon supplies cos intention brainfog ai keys organic shock sobriety
297	91	chasing	Recreational			chasing wound heightened institute reward dependence awareness national abuse center addicted potential addictive addiction agreed speed missions unheard abused notably occasion alcohol neuroscience contrary blocked darvon rogan wikipedia addicting enhancing ranexa psychological force decades growing development crack artificial smoked darvocet personal amphetamines attractive occasional addicts purposes concentrations appealing increases value
298	90	quarter	Dosage		Temporal Food	quarter noon quarters overdoing chicken dinner lunch bulletproof tripping espresso half sustain limitstart ug cheese grass finished morning cups changer redose breakfast hour cup beans worn afternoon chores paste meat veggies microdose spasms cut spot fu cleaned followed evening beforehand restful favourite egg brilliant washed wine tablet decaf bike nine
299	90	expired	Spam??			expired consume shelf date bromazepam dairy orlistat lexotanil april alabama levaquin lamisil resulted gluten al lasix getelementbyid intuniv killers display consumption dates germany february development harm women eliminating doxycycline patent netherlands stablon progressive prepare pantoprazole consideration november container cycrin awkward msg exercises fruits grow es complications suit xeloda grain france

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
300	90	getelementbyid	Conditions	Side Effects Other Interventions		getelementbyid display document medschat style discussions community none index nursing grogginess dizziness htm pregnancy hello swings metformin candy medicine insomnia irritability warnings according endep allergy fever complications listed allergies prone stromectol parkinsons refills deficient typical provigal drowsiness lists messages bodies interactions diarrhea reactions unusual frequent artvigil european corrected nutritional forming
301	89	addy	Other Interventions		Stimulants	addy quot adderral worthless meth gotta coke studying adderall abusing grind rcs addicts grinding adderal edgy ir fuck amp clonazolam cocaine hmm tadalafil hyper stim stims sends bout retain dexedrine euphoria focalin wreck amphetamines abuse comedown fucked mistaken shit youre clenching euphoric msg dying trips xr revia buzz clean ritalin
302	89	radio	Unclear			radio commercial shift disorder threatening host ads shifts ad um uses spell disturbing suicide society fladrafinil warned listening mad ru model interests boosting scientists difficulty suicidal woman pictures hospitalized watson phenergan rashes forces death cilostazol concerning sinemet black norvasc esp tv il youtube tarka vomiting staring sincerely treat exelon skin
303	89	ldn	Unclear			ldn naltrexone amantadine fl neuros neuro pots tecfidera ppms downs copaxone ms scrip msers progression crush filler disabling dmd vivid xx fatigue tysabri fatique ampyra rebif nhs gilenya interestingly flares symmetrel refused contraception losartan tarka saved soluble rrms revia ra mess ups relapses raising smallest autoimmune pramipexole vit midday sclerosis
304	87	text	Mixed			text wall merged male warning auto manic classic episodes married old regret added following lb grades yo mph female names wrote pages ah hit grad cough reminds italy freaked final english straterra assignments dog diethylpropion acetyl brought france reads beautiful watched episode march yr edited mri frustration shout stilnox undiagnosed
305	87	sessions	Effect			sessions intensive session neurofeedback training studying graduate od behavioral darvon msg bpm student exams vpxl geoff mindfulness focussed sets season library spaced school suck deadlines poker college tobacco amp degrees die homework unstable rats dependence litres attend projects exam default topics cardio trained app classes bouts cumulative temperature engineering mice

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
306	87	theres	Acquisition		Online purchasing	theres vendor legit counterfeit bro medsforbitcoin rxrex sheffield reviews received mylan pharma blister refund brands faq modvigil packaged recieved modup batches sells scam hab vendors sun modafinilcat retailer cheapest ripped reply ship branded ordered packaging imgur emails cozaar complaints disappointed european delivery brugada sunmodalert ive aas nrx anecdotes spots ppl
307	86	modapharma	Unclear			modapharma def snp rs shed quetiapine ultram skeptical refreshing aa netherlands endo worldwide adenosine steps flush iq marked zopiclone breakfast spots metabolizer theories apologize nubrain shout andme gene drawing balls diethylpropion ibs frisium cleocin prednisolone caring ratings analytical snack celebrex qhi affective ject chaos updates biohacking bonus sertraline sample dedicated
308	86	wort	Other Interventions			wort john st htp tachycardia tyrosine spell hypertensive timestamp biloba ginkgo inositol examine potentiate interact serotonin adult tryptophan emotional describing ct epa closed moods edgy creatine lighter contraindicated avoided snris gingko plz depressant interaction respectively offset shock dlpa depressed winter potentiates divided longecity dha ss anger crisis shoulders oxycocet mood
309	85	yr	Unclear			yr forming old de grades gpa un genius mom ranexa insights teaching daughter yrs female son sex propecia girls interferes chores texas male kids venlafaxine thru cleaned woman laughed cataplexy noob teachers newbie hugs concerta darvocet click quitting habit dependent citalopram assignments undiagnosed grocery girl tissue savella movies afternoons ltp
310	85	sex	Effect		Sexual	sex libido drive erection porn horny sexual relationship arousal girlfriend arginine women anhedonia girl male acne overkill dull behavior attractive movies desire child volume girls relationships connection struggled hormones snap stays casual woman compulsive coincidence boredom conversations creatively expression gaining gad pleasure rem pains occurred female cardiovascular brutal age reverse

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
311	85	downsides	Mixed			downsides semax selank powerful russian significant viracept na follows fasoracetam diminishing flucloxacillin russia benzodiazepines stabilizers cards deprenyl tool lostfalcos nootropics tim experiments model suspension clenching sulfate anxiolytic rep racetams consideration lyme imitrex nootropic prami positively coluracetam credit event chapter feelings spray sparingly fitness handful aspect notable failing likelihood described swings
312	84	binge	Other Interventions		Recreational	binge mdma cocaine fucking man zero fucked self darvon god hungry joy fuck fpm masters coke control ptsd mushrooms sober mad meth impulse beers killed hang booze methamphetamine hangover damn hppd fun hanging psychedelic rolling messed guilt remission ending hardly bars nightmare loving happily waste cb smoked intellectual alcohol mono
313	84	viibryd	Spam??			viibryd prozac sexual dealt lexapro zoloft zaps celexa nightmares ssri closing fetzima effexor depakote itraconazole lithium abilify advertisement buproprion lamictal digestive failure adjunct snris boyfriend mdd ssris luvox marketed brintellix pdoc quicker newer latuda rude restorative seroquel buspirone gad somebody perscribed italy remeron desirable posters buspar welbutrin visiting prosteride activated
314	83	delivery	Spam??		Acquisition	delivery cheap overnight cod fedex perscription consultation membership buy prescription prescript online shipping shipped saturday cash cheapest presciption tarka nextday rx worldwide order script sale usa purchase prescriptions without no ex pharmacy pulmicort buying frisium delivered consult ordering sales watson cilostazol mastercard next accepted alabama line discount generic canada vpxl
315	83	testosterone	Spam??			testosterone trt hormone hgh supplementing thyroid iu gel dhea cortisol levels hcg hormones erection flat deficiency cortex replacement supplementation zolpidem cramps apathetic frontal libido ocd horny restore noprescription arginine sexual emotionally ambien benefited steroids muscle ptsd steroid restless numb depressed prefrontal production temporary culprit magnesium increasing nmda iq dave estrogen

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
316	82	keppra	Unclear			keppra hppd sinemet andme convinced lamotrigine dopaminergic seizure lamictal combos dp tdcs trial particularly nac titration anyhow shot tumor beers epilepsy levlite freaked instant tendencies sam researching stabilizers topamax spaced risperidone rated diethylpropion depressive seizures dilantin stabilizer bothered continuing vpxl depakote biohacking alas poop reporting metabolizer uni lowered chess length
317	82	usd	Acquisition		Spam??	usd brazil fee timestamp hab retailer pharmaceuticals india modafil modafinilcat indian send sun tab tadalafil ems pharma starts nextday waklert base paypal pack oxycocet shipping prepare imported beginners nrx btc name worldwide arent fees supeudol door reship supplier esterom mail scam modup strip aus fedex guarantee payment delivery cialis quote
318	81	mb	Mixed			mb methylene blue fish complex alcar aswell multivitamin ala freely raw pqq bang topics conjunction dual upping hydergine oil moderately jpg observation negatives omega piracetam lurking consciousness dan pissed cycled starts stack skipped mcg cdp material unnecessary supps practicing learning alpha vitamin expecting exams divided noticing cd nevertheless corrected awakeness
319	81	methadone	Recreational			methadone suboxone opioid opioids opiates oxycodone oxycontin tapering opiate nod morphine fibromyalgia withdrawal expand pain randomly heroin nerve subutex wd sincerely reps enzyme subs sum relief taper thc resulted cloud intolerable nonetheless pregabalin knocked remotely presently cereal fibro kicked oxy muscular norco pushed metabolized anytime withdrawl imipramine chime pregnant nubain
320	80	observations	Effect			observations remarkable notes nardil compound hypotension relation blurry gift poop augment shout match arent mr problematic timing pregabalin obtain augmenting campus nmr concentrations lovely ya lyrica detailed replaced peak checks pitolisant cc birthday plasma butalbital experimentation parnate intolerance scott maois refer maoi dunno gradual gps anonymous bump ss rich briefly

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
321	80	twitching	SideEffects			twitching aids arms swollen legs itching spasms muscle eye sensations tingling pains purchased leg shortness breathing needles nerves shaking attributed knee cures knees tremors revealed tren overdose jaw closing seconds wk sociable contains pregabalin movement chest ears fragmented rls apneas hallucinations nauseated migraines aches withdrawl gradual ritilan worsened ringing discontinuing
322	79	yellow	SideEffects			yellow dropping eyes concerning asap white delay circles zone die tight isomer anastrozole comfort feb skin thistle vegetables transporter heartbeat spironolactone coconut piece alternatively land kidneys style potassium poison fruits frequent laid photos beats pics butter display realized liver seconds chest sorted mixture fatal worsening motor jumped vibramycin damage swelling
323	78	dp	Conditions			dp hppd concentrating degree psychology uni comfort lamotrigine cured attempted psych returning exacerbate join mindfulness courses zone thankfully tms techniques makers anxiogenic tdcs smile london deals king sa permanently problematic worsen sentence neurofeedback charm pi schizophrenia tough cognitively modanifil completed google luvox competition dropping ssri discontinuation therapist cognition waste sclerosis
324	77	sadness	Emotions			sadness lingering anger suppression feelings touched emotions intolerable tries swim depressive joy hungry thirsty happiness logic emotional depressed happily pissed struggles growth replied irritability acetazolamide balanced sociable amped talkative dealt trileptal encouragement mdd burnout tarka subs firstly pleasure married cook appetite lethargy sh recognize swings sufficient darvon relationship contfzzza april
325	77	qhi	Spam??			qhi netherlands sunmodalert ships alertec ship btc eu reliable nubrain legit mymodafinil vendors sites medsforbitcoin modup tpe seller payment site bitcoins transaction co vendor reship geoff circle bitcoin rxrex olmifon ordered cc packaging skypax refund modapro prescript shipments supplier keppra card arrived shipment ems source received cheaper iop branded email

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
326	76	pro	Mixed			pro concerns helping tournaments avonex vigil dog banned neurology rheumy dmd player loop esports ms tsh minus flares tad animals rrms ephedra players physicians ldn aphthasol applied drawing copaxone art married hack rated smoker dopa exercising fatigue ruled formulary existed believes neuro apathetic antihistamines ac published marijuana joints enjoyed smoking
327	76	merged	Spam??			merged text auto added following edandmore feedbacks edited transaction spier iop post modapro minutes visa tpe introduced credit authoritative charges last patent modalert account rated brand fraud images bank sun stromectol speedy alertec cc pr bucks comparing reviews card fake isomer product conversion mastercard stilnox classic responsive ticket cephalon prosteride
328	75	depleted	Mixed			depleted nutrients supplement vitamin overkill bystolic lethargic ranexa fentanyl glucose magnesium pa york bars riding narcotic cough ice mcg gaba nutritional aldactazide spray combivir washed intermittent instantly smallest hypotension physiological wind intervals sunlight norco phenibut protein downers fasting vit bus plasma flat tryptophan lostfalcos established ticket sedating endep supplementing dilantin
329	75	impacting	Effect			impacting negatively tryptophan neurotoxicity circadian executive sustainable dxm curve lowering microdoses petcam metacam glycine silicon comprehensive workers adrenals interfere adrenal alternatively viable reducing fetzima consideration suspension deprived alternating chronically memantine specializes replacing milligrams interacts valley diminish rhodiola rhythm sufficient patterns offset flares metabolizer impacts confirmed ama whose tyrosine stacked exists
330	75	tren	Unclear			tren wk hgh aas dnp gear sides hcg testosterone ed lean cycle peptides eca accutane test gains ai cardio running sibutramine drugsgear clomid training iu season trt appetite weights gym finasteride hunger lifting summer workout dmaa acne workouts lethargy dhea tapatalk preworkout sweaty suppression fat calorie propecia raised calories sexual

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
331	75	reflux	Spam??			reflux acid litres discomfort persistent nutrients lipoic carnitine temperature container keto petcam amitriptyline bread metacam suprax anemia cough cholesterol water suspension drinking gut ala ibs soda stomach bullet sickness clue cured oral migraine nauseous migraines decaf plays press connection dust absolute milk flmodafinil limitstart subside pseudoephedrine asthma mb headaches amino
332	74	antihistamine	Effect		Allergy	antihistamine pseudoephedrine hcl antihistamines nasal benadryl histamine allergies allergy itching spray barrier itch hives restlessness nose exertion imitrex losartan tactics drowsiness dust drowsy itchy allergic mirtazapine flu tournaments cross yup coping stiffness mindfulness shots dementia fever baclofen chess ringing crippling eriacta overload amitriptyline un remembered pr winter sinus biology patch
333	73	climb	Effect			climb chaos refuse lie swallow fail topics smoother breaks genius analogy triggers strangely smooth neurofeedback recovering refreshing alternating pdocs estrace ssdi film precise moclobemide directed psychiatrists comedown ordinary locked definite cycrin rational metacam changer neurology leading physicians gains character petcam built traumatic keen flow nzt manic acting ciprofloxacin perscribed sleepless
334	73	industry	Pharma			industry inc kill pharmaceutical shift disease disorder requiring pushing teva selling disorders pharma sell fda played cephalon purpose chemical html workers killing website makers pop sold es marketed patent host companies label company across approved called marketing autoimmune modavigil claim manufacturers treat pharm european maker fake media photos rep revia
335	73	ccsvi	Conditions	Other Intervention HCP	Neuro	ccsvi insufficiency procedure ampyra msers vivid ppms dmd spasms improvements flares foot fatigue tysabri eulexin sdmb shoulder volume spier chronic spinal evista september acetazolamide tap ms neurologists exelon amantadine spasticity california interests progression outcomes relapses tryptophan neck unpredictable reliably flu published injection remembered sue considerable walking million drivers counts tightness

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
336	73	sarah	Pregnancy /Birth			sarah ds moderator itching overload pregnant baby ot pregnancy medschat pilagan flonase eriacta sinus pots asap tachycardia breastfeeding notably xx carvedilol ethylphenidate breast shortage groups infections infection scott getelementbyid drives nose downloads sensory shelf sulfur osa existed propranolol cortex display obstructive ii font warnings interactions pi kb section fluid reported
337	72	neurofeedback	Effect			neurofeedback sessions tdcs training practitioner dual eeg improvements endurance cerebrolysin mindfulness biohacking technology techniques avoided tulip meditation project built intelligence injury traumatic jim exercises behavioral session device improvement psychology stimulate therapist blank remarkable topics lostfalcos depth dp challenging default oxi stamina progress sees prami therapy abilities mct oxygen brain intensity
338	72	swim	SideEffects		Effect Emotion	swim geoff cat hanging coping seconds vomiting talkative energetic swimming report itch sociable euphoric july mdma forces organized bowl lsd pool east recover nauseous fraud darvon leaving herb urgent loses added hrs roll documentary hopes hoped incredible brief twitching contributed relaxing disappointing slight intensely reputation mins tingling consecutive rolling bbc
339	72	european	Acquisition		Licensing	european agency unheard indications licensed apnoea nhs restricted disorders medicines obstructive gp association prescribe label treating indicated conditions cases adults treatment prescribing treat reported gps approved associated patients narcolepsy disease regulations suffer osa countries healthcare referral disorder psychiatric indication excessive msers monitored display fda diagnostic appears promoting ss concluded marketing
340	71	ec	Spam??			ec ephedrine preworkout dick training yohimbine arginine eca stim shortness cracked ticket endurance animal um stacked hated brings ephedra breath units lifting stimulated techniques suppression stacking pea challenge prosteride scare synergistic granted roof tbh workout rise workouts grey sides spell craze drove lurking espresso gym horny scoop bulk insulin officially

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
341	70	fri	Temporal			fri mon wed sat clonazolam till weekends weekdays pm friday wednesday monday infusion sunday tarka build weekend retail weekly lord feed closed mirror skip attend staff tuesday driven week bromantane pharmacology slept excitotoxicity revealed schedule double accutane continuously lifting conference skipping begins sleeper drives everyday band holiday rls reset involves
342	70	eugeroics	Spam??			eugeroics agents promoting eugeroic potency generation iii compound dat fluorenol dri promising beyond agent affinity notably wakefulness comparable profiles aka analog analogues adenosine search mice reuptake ii pitolisant releasing oct methylphenidate actions weak looks literature eph novel compounds paper binding histaminergic hydrafinil action adults potent pharmacology phenyl et text monoamine
343	69	newest	Acquisition		Online purchasing	newest vendor samples credit card accept debit payment claiming btc cards bitcoins worldwide sample bitcoin customer modapharma confirmation account sent site products medsforbitcoin email imgur feedbacks update giving receive emailed sunmodalert discreet ship exelon orlistat contacted received seller drugsgear reporting counterfeit nrx service edandmore reship refund eriacta paypal reddit customers
344	69	asshole	Unclear			asshole mdma intensely anti death cranky sad absolutely idiot binge booze redose tendencies social situations depressants guilt incorrect tears girl benzodiazepines moody psychedelics schizophrenia regret gather emotional forgive microdose depressant remission sobriety fucking shit personality ha hmmm angry illnesses psychedelic animals desirable ptsd soul hurts induced deserve lsd employed gf
345	67	pitolisant	Effect		MoA	pitolisant subjective compound idra observations promising lion prl histaminergic extensively trials wakefulness isomers eugeroics histamine relation headed furthermore dislike poorly compounds structure comparable remarkable ih worthless placebo systems interested modapharma disclaimer tapatalk cns iii novel desirable shown fluorenol feedback pmc promotes participants improvements clinically antagonist looks fladrafinil claiming ject reverse

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
346	67	pseudoephedrine	Other Interventions			pseudoephedrine amitriptyline ephedrine pseudo flu definately neurotoxicity petcam metacam methamphetamine eulexin neurotoxic amp intensive antihistamine alleviate ringing bright enhances neuros unhealthy advise mpa classical chemically reflux resulting peripheral awkward prosteride lacking curb considerably suspension passes antihistamines parkinsons meth msers evista ring drawn simultaneously cardio ears sinequan extent elaborate sets adderral
347	67	chess	Effect		Cognitive	chess rating vastly players olanzapine antipsychotic obsessive played game dumb games clarity play determine aim dig improve cognition draw playing park input performance score ratings heading consistently master improved obtained regard tournaments player ten escitalopram creativity changer appreciated hypomanic enhance poker zyprexa superior beginner cognitive banned central lastly competition memory
348	67	round	Unclear			round scored white pictures confirm plavix fever tylenol batch tablet nz blurry pulmicort blister foil dexone exelon screen included tablets hydrochloride jpg modafanil swallow heres november yellow popped images clenching prevacid cycrin tongue packaging imgur genetic discreet progs taste mirror swallowing sheet color letters registered film overseas pill split received
349	66	milligrams	Dosage			milligrams craze scoop vice viracept chapter preworkout podcast lostfalcos phenyl recipe spells evista experiments hundred shy advanced plate bodybuilding tools analog weigh sheet eulexin experimentation dedicated bothered revealed appeared workout doxycycline cycrin reliably skipping spread intend cos kg blend profit germany measured progressive nutrients shaking grogginess intelligent tim metacam tulip
350	65	italy	Acquisition		Online purchasing	italy eu countries germany european skypax ireland modiodal netherlands country shipments russia france orlistat strict goods institute english ships named europe ship parcel brazil india san china reship sweden spain bipap packages cleocin customs french officially nursing interests experimental banned japan nowadays pharmacists cd prosteride send containing south checks police

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
351	64	msg	Unclear			msg amp topic index php vpxl ab radiation battling jim chemo joe forums reply responded thread disappeared oxycocet cancer threads describing orlistat showthread offered scammed nonetheless sarah majority aggrenox progression motrin dot suprax positively questionable div forum neuros persistent goodness bone realise url dad comparing tumor persons vb discussions valuable
352	64	pregabalin	Other Interventions			pregabalin lyrica nardil booze paradoxical dizzy withdrawals irritated lovely rls pls cos drunk gabapentin tiredness upped reads corrected temporary ruining benzos gad dydrogesterone cycrin permanently craving fatigued fuzzy exhausted vigil addicted factor fibro relieve sedated wore nurse realise hypotension utterly fortunate breathe bumped overdoing combos darvocet horribly fatigue mary ponstan
353	64	polyphasic	Effect			polyphasic longterm deprivation intervals army deprived attempting adapt ideal biology maintaining sleep sleeping forces schedules adjustment nightmares minimize mask schedule impacts advocate neurotoxic molecules military scare settings device alarms bipap math nights debt ptsd hygiene refreshed rested cb tren monk adjusted tactics specialists doze missions nightly pdf slept paralysis waking
354	62	waves	Conditions		Neuro	waves eeg measured wave rem epilepsy beta latency kemadrin nighttime blocker showed randomized neurofeedback ih deep stage arousals mslt executive monitoring participants arousal vigilance objective sedative spironolactone wakeful blind study revealed hh motilium correlation mice beats measure sleep slow histamine brain genotype restful fragmented activity scientists efficiency unclear guidelines phenergan
355	62	quot	Recreational			quot tea rcs rough stim crappy coke trips green clonazolam phenibut addy etizolam combos favourite compulsive hangover stims stimulation cocaine decaf mushrooms birthday grinding kratom releasing cup ginseng resort buzz tadalafil gingko leaves fpm visual edited beers bromantane drinks pillow booze meth euphoric analogues kava edgy spending theanine reached aid

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
356	59	delivery	Spam??			delivery worldwide cod overnight fedex nextday cheap shipping buy online saturday consultation membership shipped cheapest presciption prescription mastercard perscription cash prescript pulmicort visa rx sale order script purchase dexone imitrex discount pharmacy consult tarka express usa no without fast accepted generic cordarone ordering discreet internet ject alabama frisium vpxl required
357	59	potentiation	Effect			potentiation via camp spacey programs heavy assistance keflex step correlation scan calming preworkout notable glucose definite sunlight scoop contrast annual program consistently stream bid potentiate exertion interacting unknown application win investigate training ct hoped presumably keen scores beast shed digestive intake selective hitting hype dreams guidance ingest protein wild creatine
358	57	alters	Effect			alters decision creatively drawing making stevens prednisolone uppers drunk adenosine crisis concluded ciprofloxacin increasingly solving shy supeudol biological conclusions downers overstimulated steven dependence lexotanil kemadrin decisions decreases fl differ butalbital johnson font gilenya hunger lord naive heavier um suppress ponstan imovane floxin forgive quiet erection sleeps caffeinated fatique heightened animals
359	57	kb	Unclear			kb downloads attached jpg bpd official photos molecule imgur pdf maximize mb eulexin longterm cats errors impulsive avandia nutritional verify psg evaluated fladrafinil image dislike shall device mslt color skipped european apologies nmr lurking prosteride ring adjustment unifiram officially ps alter cycrin evista hello magical description panel update tegretol sarah
360	54	gap	Unclear			gap competition opened resources symptom edge discuss increasing smart intelligence txt world neurotoxicity iq secret naive contfzzza image idra getstring attributed shout chess broad negligible ignorant crystal reduces monnieri decade enhancement filler conclusions phenergan statement aricept anastrozole litezennopost drugsgear performance wound tweaked usage feb FALSE processes heroin darvon statements serotonergic

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
361	53	viracept	Spam??			viracept ciprofloxacin viramune cod delivery levlite overnigh frisium fedex overnight perscription cash shipping darvocet aldactazide presciption membership supeudol saturday cheap consult clorazepate shipped ponstan ranexa euthyrox oxycocet script required prescription consultation rx ex keflex zolpimist online delivered buy aphthasol no fees next darvon esterom aggrenox vb without prescriptions discount nextday
362	53	finasteride	Spam??			finasteride pathway hair create cyp enzyme propecia loss authoritative query hospitalized estradiol aggrenox saliva gilenya darvon acetazolamide female male enzymes educated viagra somnolence ginkgo thick beginners carvedilol stabilized lowers metabolic oxycocet toxicity aap metabolism ranexa plaquenil prosteride metabolites mistaken eriacta medicating induces breakdown erowid metabolite armod decreasing grow stores lead
363	51	glucose	Dosage			glucose potentiation insulin sugar intake metabolic workouts carbs workout calorie meal fasting bread consuming protein intermittent followed blood meals pqq preworkout fruits cardio frontal dates juice apple imbalance carb hormones responsible afterwards converts metabolize fruit acetylcholine bull oxygen fuel nutritional eat foods limiting energy hydration drinks adrenaline fat consume chocolate
364	48	session	Outcome			session sessions retain hurt learn distractions senior learning relaxation finishing bpm studying math passes exams physics marathon driver awkward incredible projects cbt lifting app troubles weights exercises holy yoga ibs material gym exam recall programming cardio trazadone loosing appropriate downers boring completing meditation min grind january socially memories mindfulness model
365	45	showthread	Acquisition		Online purchasing	showthread vpxl url tid endep php ru cycrin premarin plavix floxin fedex delivery losartan membership cheap cash norvasc orlistat canada avandia shipping prescript cheapest buy saturday cod prescription canadian usa online overnight paxil consultation nextday counter discounted celebrex topic alesse shipped omeprazole price coupons rx script frisium sale pharmacy perscription

Topic No	No of Docs	TopWord	Topic	Other Topics	SubTopic	Topic Words
366	45	valley	Biohacking		Enhancement	valley silicon tech marijuana popular workers pot microdoses limitless executive industry reality manager news possession fwiw movie toll types stoned asprey challenge drawn clients square consuming charm itraconazole recreationally smart nzt bouts intelligence cannabis granted heh ebay media johnson aids employee systems bupe remotely relationships aging upgraded hack addicts america

Appendix N Conditions extracted from site names

Unique Site Conditions	No Posts
Narcolepsy	4910
Multiple Sclerosis	2953
ADD/ADHD	2206
Social Anxiety	1158
Sleep Apnea	711
Depression	655
Bipolar	618
ME/CFS	612
Fibromyalgia	329
Chronic Disease	294
Mental Health	282
Schizophrenia	201
Sjogrens	185
Hypersomnia	181
Dysautonomia	177
Lupus	132
Cancer (Breast)	121
Sleep Disorder	113
Depersonalisation	89
Cancer	79
Parkinsons Disease	77
Neuro	74

Unique Site Conditions	No Posts
Crohns	72
Anxiety	68
Restless Leg Syndrome (RLS)	67
Aspergers	66
Addiction	58
Brugarda	56
Addiction (Alcohol)	55
HPPD	53
Arthritis	50
Cancer (Lung)	50
Adrenomyeloneneuropathy (AMN)	50
Alzheimers	49
Ankylosing Spondylitis	49
Epilepsy	46
PBC	46
Rare Disease	41
Idiopathic Hypersomnia	41
ALS/MND	35
PTSD	35
Schizophrenia / Psychosis	33
Lewy Body Dementia	33
DSPD	32

Unique Site Conditions	No Posts
Hepatitis C	32
Rheumatoid Arthritis (RA)	30
Obesity	28
Chronic Pain	28
Ehlers Danlos	27
Tinnitus	24
Arthritis - Psoriatic	23
Thyroid Disorders	21
Epstein Barr Virus (EBV)	21
Adrenal	21
Diabetes	20
Insomnia	20
Addiction (Substance Abuse)	19
IBS	17
LymeDisease	17
Pain	16
Hypothyroidism	16
Migraine	16
Borderline Personality Disorder	15
Sexual Health	15
Conduct Disorders	13
Disassociative Identity Disorder (DID)	10
Chronic Lymphocytic Leukaemia (CLL)	10
Withdrawal	10

Unique Site Conditions	No Posts
Psoriasis	10
Huntingdons Disease	10
Acne	10
Hepatitis	9
TBI	9
Panic	9
Propecia	9
Trigeminal Neuralgia	8
Kleine Levin Syndrome (KLS)	8
Avoidant Personality	8
Suicide	8
Cardiac	7
Behcets	7
Progressive Supranuclear Palsy (PSP)	7
Addiction (PrescriptionDrugs)	7
HairLoss	7
Hughes Syndrome	6
Polyneuropathy	6
Joint Replacement	6
Celiac	6
Contraception	6
Stroke	6
Cancer (Colon)	6
Spine	6

Unique Site Conditions	No Posts
Chiari	5
Facial Pain	5
Melanoma	5
Ejaculatory Anhedonia	5
OCD	5
Pregnancy	5
Vasculitis	4
Cystic Fibrosis (CF)	4
Muscular Dystrophy	4
PSSD	4
Autism	4
Tendonitis	4
Anxiety Panic Phobias	4
AVM	4
Organ Transplant	4
Myeloma	4
Premature Ejaculation	4
Morgellons Disease	4
Wegeners Granulomatosis	4
Mast Cell Disorders	4
Myasthenia Gravis	3
Errectile Dysfunction	3
Vertebral Artery Dissection (VAD)	3
Tourettes	3

Unique Site Conditions	No Posts
Dercums Disease	3
COPD	3
Cancer (Prostate)	3
Liver Disease	3
DVT	3
Reflex Sympathetic Dystrophy (RSD)	3
Epilepsy - VNS device	3
GERD	2
IBD	2
Ataxia	2
Myelitis	2
Hypertension	2
Interstitial Cystitis	2
Raynauds	2
HRT	2
Acoustic Neuroma	2
Bone Marrow	2
Tension Myosoytis Syndrome (TMS)	2
Chlamydia Pneumoniae (CPN)	2
Primary Biliary Cirrhosis (PBC)	2
LVAD	2
Chronic Inflammatory Demylinating Polyneuropathy (CIDP)	2
Rosacea	1

Unique Site Conditions	No Posts
Lipoma	1
Myositis	1
TMJ	1
ICD device	1
Addisons Disease	1
Hidradentis Suppurativa (HS)	1
HHV-6	1
Schizoid Personality Disorder	1
PCOS	1
Idiopathic Intracranial Hypertension (IIH)	1
Atrial Fibrillation	1
Acute Disseminated Encephalomyelitis (ADEM)	1
Brain Fog	1

Unique Site Conditions	No Posts
Hypopituitarism	1
AutoImmune	1
Lymphoma	1
CKD	1
Glossopharyngeal Neuralgia (GPN)	1
Brain Aneurysm	1
Hyperacusis	1
Cancer (Testicular)	1
Valve Replacement	1
Scoliosis	1
Chronic Myelogenous Leukemia (CML)	1
Cancer (Bile Duct)	1
Anaemia (Pernicious)	1
Functional Neurological Disorder	1

Appendix O Top 100 KeyWords and Themes

Item	Freq	No Posts	Score	**Theme**
modafinil	47585	31252	2730.9	Modafinil
nuvigil	22333	13760	2604.1	Modafinil
provigil	30483	19833	1713.6	Modafinil
narcolepsy	5337	4015	464.6	Reason
armodafinil	3727	2978	440.7	Modafinil
adrafinil	3209	2086	378.2	OtherInt
modalert	2907	1900	337	Modafinil
adderall	8120	5365	333.5	OtherInt
nootropic	3079	2258	332.6	OtherInt
stimulant	7076	5308	260	OtherInt
xyrem	2097	1202	244	OtherInt
ritalin	5223	3938	232.1	OtherInt
piracetam	1981	1336	229.3	OtherInt
noopept	1586	1114	191.2	OtherInt
moda	1901	1375	178.6	Modafinil
amphetamine	3487	2665	162.5	OtherInt
sleepiness	2120	1775	162	Reason
vyvanse	1364	977	153.5	OtherInt
caffeine	5854	4226	139.6	OtherInt
wakefulness	1505	1314	134.2	Mixed
pdoc	1265	919	129.5	?
cataplexy	1098	731	126	Reason
racetam	1031	809	124.4	OtherInt
choline	1379	936	119	Nutrient
aniracetam	889	666	107.5	OtherInt
wellbutrin	1785	1241	105.7	OtherInt
amantadine	957	804	105.6	OtherInt
phenibut	865	606	103.7	OtherInt
fatigue	10294	6935	102.8	Reason

Item	Freq	No Posts	Score	**Theme**
phenylpiracetam	834	677	101.5	OtherInt
methylphenidate	1035	838	99.3	OtherInt
l-theanine	856	726	97.4	AminoAcid
apnea	2132	1602	96	Reason
neuro	1347	1088	95.8	Reason
hypersomnia	842	691	95.7	Reason
cod	4541	273	94.6	Acquisition
med	9284	6415	91.3	?
dexedrine	806	622	87.9	OtherInt
nardil	671	317	78.7	OtherInt
nootropics	658	584	78.7	OtherInt
cephalon	697	496	78	Company
dopamine	2065	1277	78	Mixed
concerta	863	667	76.4	OtherInt
cpap	1095	748	75.3	OtherInt
stim	800	639	75.1	OtherInt
adhd	3161	2190	73.7	Reason
rx	2485	1022	73.4	Dosage
mslt	606	414	73	Investigation
adderal	616	521	72.8	OtherInt
lamictal	795	598	72.8	OtherInt
fedex	1953	241	71.9	Acquisition
lyrica	686	386	70.6	OtherInt
awake	6082	4982	69.1	Effect
afinil	546	458	66.9	Modafinil
waklert	546	405	66.4	Modafinil
neurologist	1339	1108	65.2	HCP

Item	Freq	No Posts	Score	**Theme**
tianeptine	535	376	64.7	OtherInt
nap	3026	2168	64.4	Reason
ssri	781	653	62.3	OtherInt
strattera	627	416	61.4	OtherInt
parnate	513	278	60.1	OtherInt
prescription	12456	3593	59.6	Acquisition
benzo	771	595	59.1	OtherInt
theanine	477	416	56.4	AminoAcid
modvigil	453	323	55.3	Modafinil
oxiracetam	450	372	55.1	OtherInt
selegiline	480	307	55	OtherInt
tiredness	864	767	53.3	Reason
jittery	626	602	53.3	SideEffect
sleepy	1735	1485	52.9	Reason
cymbalta	632	487	52.7	OtherInt
alertness	785	719	51.7	Effect
effexor	628	412	51.6	OtherInt
sjs	465	323	51.1	SideEffect
bacopa	436	340	51	OtherInt
perscription	476	213	50.4	Acquisition
fibro	533	411	49.8	Reason
narcoleptic	422	384	49	Reason
dosage	3129	2549	48.3	Dosage
pill	6731	4783	47	Dosage
bupropion	514	370	46.7	OtherInt
modifinil	374	320	46.2	Modafinil
xr	901	699	45.2	?
melatonin	789	623	45.1	OtherInt
alcar	369	294	44.8	OtherInt
histamine	571	349	44.4	?
pramiracetam	355	304	43.7	OtherInt
dose	12635	8566	43.2	Dosage

Item	Freq	No Posts	Score	**Theme**
antidepressant	1400	997	42.9	OtherInt
eds	534	412	42.5	Reason
sulbutiamine	338	276	41.6	OtherInt
prescribe	6011	4958	40.5	Acquisition
commenttext	325	325	40.3	NotRel
entry-content	325	325	40.2	NotRel
itemprop	325	325	39.6	NotRel
seroquel	572	407	39	OtherInt
focalin	327	260	39	OtherInt
sleep	25387	13079	38.4	Mixed
ih	493	353	38.1	Reason
dextroamphetamine	331	284	37.9	OtherInt

Appendix P Top 1000 Keywords

Keyword	Freq	No Posts	Score
modafinil	47585	31252	2730.9
nuvigil	22333	13760	2604.1
provigil	30483	19833	1713.6
narcolepsy	5337	4015	464.6
armodafinil	3727	2978	440.7
adrafinil	3209	2086	378.2
modalert	2907	1900	337
adderall	8120	5365	333.5
nootropic	3079	2258	332.6
stimulant	7076	5308	260
xyrem	2097	1202	244
ritalin	5223	3938	232.1
piracetam	1981	1336	229.3
noopept	1586	1114	191.2
moda	1901	1375	178.6
amphetamine	3487	2665	162.5
sleepiness	2120	1775	162
vyvanse	1364	977	153.5
caffeine	5854	4226	139.6
wakefulness	1505	1314	134.2
pdoc	1265	919	129.5
cataplexy	1098	731	126
racetam	1031	809	124.4
choline	1379	936	119
aniracetam	889	666	107.5
wellbutrin	1785	1241	105.7

Keyword	Freq	No Posts	Score
amantadine	957	804	105.6
phenibut	865	606	103.7
fatigue	10294	6935	102.8
phenylpiracetam	834	677	101.5
methylphenidate	1035	838	99.3
l-theanine	856	726	97.4
apnea	2132	1602	96
neuro	1347	1088	95.8
hypersomnia	842	691	95.7
cod	4541	273	94.6
med	9284	6415	91.3
dexedrine	806	622	87.9
nardil	671	317	78.7
nootropics	658	584	78.7
cephalon	697	496	78
dopamine	2065	1277	78
concerta	863	667	76.4
cpap	1095	748	75.3
stim	800	639	75.1
adhd	3161	2190	73.7
rx	2485	1022	73.4
mslt	606	414	73
adderal	616	521	72.8
lamictal	795	598	72.8
fedex	1953	241	71.9
lyrica	686	386	70.6

Keyword	Freq	No Posts	Score
awake	6082	4982	69.1
afinil	546	458	66.9
waklert	546	405	66.4
neurologist	1339	1108	65.2
tianeptine	535	376	64.7
nap	3026	2168	64.4
ssri	781	653	62.3
strattera	627	416	61.4
parnate	513	278	60.1
prescription	12456	3593	59.6
benzo	771	595	59.1
theanine	477	416	56.4
modvigil	453	323	55.3
oxiracetam	450	372	55.1
selegiline	480	307	55
tiredness	864	767	53.3
jittery	626	602	53.3
sleepy	1735	1485	52.9
cymbalta	632	487	52.7
alertness	785	719	51.7
effexor	628	412	51.6
sjs	465	323	51.1
bacopa	436	340	51
perscription	476	213	50.4
fibro	533	411	49.8
narcoleptic	422	384	49
dosage	3129	2549	48.3
pill	6731	4783	47
bupropion	514	370	46.7

Keyword	Freq	No Posts	Score
modifinil	374	320	46.2
xr	901	699	45.2
melatonin	789	623	45.1
alcar	369	294	44.8
histamine	571	349	44.4
pramiracetam	355	304	43.7
dose	12635	8566	43.2
antidepressant	1400	997	42.9
eds	534	412	42.5
sulbutiamine	338	276	41.6
prescribe	6011	4958	40.5
commenttext	325	325	40.3
entry-content	325	325	40.2
itemprop	325	325	39.6
seroquel	572	407	39
focalin	327	260	39
sleep	25387	13079	38.4
ih	493	353	38.1
dextroamphetamine	331	284	37.9
norvir	318	4	37.6
rhodiola	334	281	37.4
semax	301	221	37.3
asleep	3270	2520	36.7
copaxone	314	253	36.5
kratom	662	390	36.4
baclofen	445	270	36.2
sunifiram	288	157	35.8
latuda	293	200	35.6
gabapentin	437	324	35.4

Keyword	Freq	No Posts	Score
alesse	313	6	34.9
noot	287	234	34.6
modafinilcat	277	249	34.5
prozac	647	486	34.2
maoi	303	244	34
lexapro	472	331	33.7
ciltep	271	175	33.7
insomnia	1106	914	33.6
memantine	297	211	33.5
klonopin	633	446	33.4
stims	270	248	33
abilify	699	522	33
gpc	378	301	32.5
norvasc	379	7	32.2
comedown	279	229	32.1
paxil	417	166	31.9
overnight	4970	471	31.8
generic	5413	2843	31.4
rem	709	465	30.6
pantoprazole	269	9	30.2
alertec	244	190	30.2
mdma	421	302	30
symmetrel	241	17	29.8
geodon	255	159	29.7
stromectol	249	3	29.3
anhedonia	257	175	29.3
zoloft	427	329	29.2
flmodafinil	232	154	29
gaba	389	281	28.9

Keyword	Freq	No Posts	Score
flucloxacillin	229	2	28.2
daytime	1406	1234	28.1
orexin	238	153	28.1
edandmore	224	172	28.1
euphoria	524	439	28
copay	299	254	28
anti-depressant	360	300	27.7
citalopram	298	117	27.6
artvigil	220	167	27.5
zyprexa	305	205	27.3
tolerance	2968	2330	27.1
combivir	241	2	27.1
off-label	339	302	27
estrace	226	5	26.9
meth	613	416	26.6
dopaminergic	288	259	26.6
tired	3659	3070	26.5
reuptake	302	251	26.5
doxycycline	466	16	26.1
bipolar	1012	724	26.1
l-tyrosine	216	180	25.8
headache	2645	2043	25.7
accupril	206	1	25.5
spier	221	151	25.2
doc	3441	2650	25.2
lamisil	326	6	25.2
pristiq	210	153	25.1
ahi	266	166	24.9
modafanil	197	165	24.8

Keyword	Freq	No Posts	Score
medication	7595	5319	24.7
cfs	526	360	24.7
drowsiness	361	304	24.7
metabolize	416	364	24.5
dexamphetamine	197	167	24.4
legit	599	543	24.4
idiopathic	342	314	24.3
pramipexole	202	57	24.2
euphoric	312	273	24.1
bromazepam	194	31	24
side-effect	562	471	24
psychiatrist	1172	969	24
modafinal	190	152	23.9
pulmicort	206	1	23.8
desoxyn	195	136	23.8
nitrofurantoin	231	7	23.8
ssris	282	244	23.7
adderrall	188	139	23.6
manic	511	403	23.5
moduretic	184	1	23
snri	191	175	23
psych	487	419	22.9
clonazepam	401	292	22.6
litezennopost	179	1	22.6
contfzzza	179	1	22.6
emsam	181	87	22.4
hydrochlorothiazide	223	4	22.4
xanax	755	535	22.3
fioricet	408	50	22

Keyword	Freq	No Posts	Score
buspirone	206	52	22
norepinephrine	260	201	21.9
nicotine	761	554	21.9
getstring	179	1	21.8
ephedrine	413	332	21.8
pharmacy	3580	1812	21.7
sedate	369	312	21.7
caffiene	179	145	21.6
teva	260	167	21.6
serotonin	516	368	21.4
ashwagandha	196	164	21.3
dex	348	239	21.3
armo	171	123	21.2
b12	461	379	21.2
anxiety	4507	3248	21.1
cleocin	210	3	21
enhancer	396	323	21
atomoxetine	175	131	20.9
ambien	522	361	20.8
agonist	407	291	20.7
zopiclone	210	67	20.6
mirtazapine	183	132	20.6
prn	234	174	20.6
lethargy	278	254	20.5
appt	249	227	20.5
acetazolamide	170	3	20.4
ject	166	3	20.3
levlite	160	1	20.3
frisium	160	1	20.3

Keyword	Freq	No Posts	Score
addictive	791	675	20.3
prednisolone	225	8	20.1
vit	255	203	20
pilagan	157	1	20
xeloda	161	4	19.8
sublingual	200	167	19.7
ldn	192	114	19.7
lunesta	217	142	19.6
ponstan	154	1	19.5
hypomania	170	133	19.5
drowsy	260	237	19.5
promethazine	176	14	19.5
irritable	373	351	19.5
clorazepate	156	2	19.5
anxiolytic	183	162	19.5
groggy	203	193	19.4
aldactazide	152	1	19.3
eugeroic	152	131	19.3
lexotanil	151	36	19.2
uridine	161	123	19.2
ranexa	152	2	19.2
fog	1431	1167	19.2
lsd	398	293	19.1
longecity	150	122	19.1
titrate	193	177	19.1
presciption	157	120	19.1
jitter	329	313	19
dx	469	381	18.9
trazodone	177	144	18.9

Keyword	Freq	No Posts	Score
supeudol	147	1	18.8
esterom	147	1	18.8
aphthasol	147	1	18.7
viracept	151	3	18.7
naltrexone	205	88	18.7
prodrug	176	159	18.6
creatine	275	237	18.6
fibromyalgia	418	337	18.6
deprenyl	149	110	18.4
pharma	771	627	18.3
bulletproof	252	196	18.3
hydrafinil	143	101	18.3
potentiate	188	173	18.1
carvedilol	152	5	18
neurontin	196	154	18
tysabri	147	96	17.9
osa	279	189	17.8
euthyrox	139	2	17.7
r-modafinil	138	101	17.7
nsi-189	138	98	17.6
zolpimist	138	1	17.6
depression	4432	3147	17.5
hycodan	138	1	17.5
tamoxifen	239	35	17.5
reddit	437	385	17.4
armoda	135	93	17.3
celexa	186	147	17.2
remeron	152	110	17.2
psychostimulant	145	127	17.1

Keyword	Freq	No Posts	Score
huperzine	136	108	17
alpha-gpc	133	111	17
mg	3345	2012	17
lamotrigine	155	130	16.9
viramune	147	1	16.9
nauseous	197	181	16.5
sublingually	133	115	16.5
antipsychotic	270	214	16.5
opiate	381	294	16.3
prevacid	199	8	16.3
bystolic	138	6	16.3
palpitation	209	189	16.3
darvon	165	3	16.3
ampyra	128	103	16.2
subreddit	170	154	16.2
co-pay	188	147	16.1
modiodal	125	95	16.1
savella	127	87	16
lethargic	203	190	15.8
welbutrin	126	99	15.8
rebif	127	100	15.7
multivitamin	200	187	15.7
neurotransmitter	325	261	15.7
diagnose	2791	2400	15.7
avonex	128	104	15.7
mania	379	301	15.6
modapro	121	98	15.6
tramadol	647	279	15.5
saphris	121	82	15.4

Keyword	Freq	No Posts	Score
wakalert	119	97	15.4
non-responder	131	114	15.4
codeine	440	165	15.4
isomer	190	145	15.2
combo	1134	966	15.2
plaquenil	136	59	15.1
pcp	255	217	15.1
nmda	154	94	15.1
dilantin	142	12	15.1
vibramycin	127	4	15.1
sertraline	162	122	15
spierx	115	72	14.9
l-carnitine	137	113	14.8
meds	135	124	14.8
pwn	128	87	14.8
idk	220	198	14.7
bipap	123	75	14.7
ive	808	663	14.6
rosea	130	122	14.6
intuniv	114	85	14.6
hypomanic	120	107	14.5
modavigil	112	80	14.5
medicate	225	209	14.4
phenyl	145	104	14.4
gilenya	114	78	14.3
placebo	638	526	14.2
irritability	218	201	14.2
etizolam	111	87	14.2
diethylpropion	197	7	14.1

Keyword	Freq	No Posts	Score
ymmv	137	136	14.1
valium	403	279	14.1
armod	108	80	14
nausea	568	486	14
ocd	302	221	14
revia	111	8	14
asprey	113	93	13.9
keflex	158	4	13.9
clonidine	134	100	13.9
coluracetam	107	81	13.9
stilnox	110	8	13.9
obstructive	265	245	13.8
amph	109	69	13.8
motrin	131	23	13.8
risperdal	129	99	13.7
ativan	332	168	13.7
eulexin	106	3	13.7
dydrogesterone	106	1	13.7
venlafaxine	134	107	13.6
ginseng	208	166	13.6
appetite	1010	837	13.5
fluoxetine	174	125	13.5
drug	16016	10599	13.5
diazepam	332	136	13.4
psychosis	302	213	13.3
cognition	517	422	13.3
tarka	116	2	13.3
sinequan	105	1	13.2
sleeping	913	834	13.2

Keyword	Freq	No Posts	Score
vinpocetine	105	87	13.2
benadryl	130	113	13.2
topamax	168	129	13.1
dmae	106	81	13.1
aderall	101	82	13.1
redose	100	94	13
rash	660	435	13
doctor	11343	7646	13
hallucination	319	253	12.9
mood	2359	1934	12.9
nzt	104	83	12.9
limitless	318	292	12.9
suboxone	144	96	12.9
buspar	116	90	12.9
div	325	325	12.9
microdosing	102	76	12.9
racetams	98	89	12.8
wake	4337	3313	12.8
skypax	97	67	12.7
kemadrin	98	1	12.7
otc	323	290	12.7
somnolence	110	93	12.7
tyrosine	199	147	12.6
evista	106	3	12.6
enantiomer	118	89	12.5
forskolin	104	73	12.5
motilium	105	3	12.5
migraine	502	375	12.5
half-life	279	248	12.4

Keyword	Freq	No Posts	Score
depressant	152	144	12.4
maois	102	88	12.3
foggy	246	232	12.3
depressed	474	442	12.3
adra	102	77	12.2
ms	3198	2105	12.2
depersonalization	105	71	12.2
bitartrate	99	80	12.1
noprescription	92	39	12.1
lorazepam	214	91	12.1
wakefullness	92	86	12.1
cycrin	92	1	12.1
dmaa	95	78	12.1
ciprofloxacin	154	3	12.1
nighter	101	94	12.1
depressive	284	249	12
sedation	227	176	12
bedtime	340	309	12
depakote	109	91	11.9
depress	622	567	11.9
guanfacine	93	67	11.9
cdp	192	155	11.9
noradrenaline	105	91	11.9
cilostazol	93	2	11.9
phendimetrazine	142	4	11.9
galantamine	94	64	11.8
floxin	95	2	11.8
vpxl	90	3	11.8
modifinal	89	65	11.7

Keyword	Freq	No Posts	Score
methamphetamine	244	201	11.7
mirapex	93	71	11.7
racemic	102	86	11.7
hypocretin	91	54	11.7
stimulating	325	308	11.6
prescribed	534	516	11.5
trazadone	91	78	11.5
exhaustion	395	346	11.5
libido	205	156	11.5
viibryd	88	57	11.5
spironolactone	105	7	11.5
cialis	343	66	11.5
glutamate	164	106	11.5
psg	181	120	11.4
rheumy	100	85	11.4
acetyl	109	101	11.4
nortriptyline	93	64	11.4
dxm	93	65	11.3
ceretropic	85	77	11.3
anastrozole	102	3	11.2
adderral	85	73	11.2
antihistamine	160	140	11.2
ritilin	85	74	11.2
neuroprotective	107	91	11.2
cognitive	1771	1379	11.2
escitalopram	96	77	11.2
addy	127	97	11.2
histaminergic	85	51	11.2
rxrex	84	69	11.1

Keyword	Freq	No Posts	Score
biaxin	103	11	11.1
inattentive	113	101	11.1
im	2474	1661	11.1
adrenal	249	159	11.1
endep	89	1	11.1
haha	482	456	11
hyper	258	235	11
motivation	1672	1385	11
eriacta	84	5	11
yohimbine	89	68	10.9
anecdotal	260	249	10.9
modaf	83	59	10.9
wakefulness-promoting	82	78	10.8
pregabalin	94	82	10.8
ketamine	138	112	10.8
fluorenol	81	57	10.8
bromantane	81	61	10.8
stack	2301	1604	10.8
vyvance	81	68	10.8
comt	86	55	10.7
prednisone	169	124	10.6
spasticity	106	85	10.6
dexone	80	1	10.6
ghb	105	82	10.6
propranolol	119	98	10.5
anxious	848	779	10.5
petcam	79	1	10.5
counteract	284	274	10.5
cheap	5672	1891	10.5

Keyword	Freq	No Posts	Score
deprivation	389	338	10.5
cyp3a4	98	68	10.5
biloba	101	94	10.5
mser	80	72	10.5
symptom	4446	3440	10.5
keppra	84	52	10.5
estradiol	173	18	10.4
hangover	226	204	10.4
all-nighter	92	85	10.4
cozaar	109	2	10.4
wakeful	89	82	10.4
nubain	79	11	10.4
cdp-choline	78	66	10.3
microdose	79	67	10.3
modup	77	62	10.3
omeprazole	104	14	10.2
brintellix	77	56	10.2
havent	234	217	10.2
circadian	171	140	10.2
amantidine	76	63	10.2
tegretol	83	56	10.1
orlistat	99	8	10.1
viagra	511	110	10.1
onemedstore	75	60	10.1
itraconazole	98	1	10.1
luck	2553	2471	10
mild	1350	1208	10
bp	760	563	10
atenolol	123	24	10

Keyword	Freq	No Posts	Score
hydergine	75	59	9.9
lol	1680	1512	9.9
lipoic	88	86	9.9
relapse	372	289	9.9
olmifon	74	60	9.9
cipro	121	12	9.9
ethylphenidate	74	56	9.9
noticeable	610	581	9.9
metacam	79	1	9.9
alot	649	588	9.9
thyroid	540	428	9.9
pharm	133	121	9.9
darvocet	151	2	9.8
shit	1445	1204	9.8
prog	180	2	9.8
anxiogenic	73	61	9.7
levaquin	119	4	9.7
rheumatologist	103	96	9.7
avandia	92	3	9.7
desipramine	76	51	9.6
modafil	71	56	9.6
pqq	75	42	9.6
agonism	74	57	9.6
zembrin	71	32	9.6
oxycocet	71	2	9.6
cerebrolysin	71	57	9.5
unmotivated	91	89	9.5
suppressant	105	98	9.5
non-stimulant	72	71	9.5

Keyword	Freq	No Posts	Score
shitty	198	180	9.5
anyways	446	424	9.5
lithium	393	305	9.5
nalt	70	57	9.4
snris	73	65	9.4
effect	19206	13300	9.4
caffine	72	61	9.4
trt	94	65	9.4
rls	107	67	9.4
dri	106	82	9.4
asv	88	52	9.3
sunpharma	69	62	9.3
straterra	70	63	9.3
doesnt	480	424	9.3
refill	360	296	9.3
b-12	88	72	9.3
costco	176	129	9.2
magnesium	460	399	9.2
lipitor	126	19	9.2
nac	135	102	9.2
acetylcholine	104	78	9.2
oxi	70	58	9.2
centrophenoxine	68	58	9.2
supplement	2302	1859	9.2
suprax	98	3	9.2
d3	201	177	9.1
psychotic	204	169	9.1
monnieri	69	60	9.1
relaxer	83	76	9.1

Keyword	Freq	No Posts	Score	Keyword	Freq	No Posts	Score
ginkgo	106	98	9.1	reboxetine	66	54	8.7
crl-40	67	48	9.1	losartan	94	6	8.7
benzos	69	61	9.1	phentermine	255	170	8.7
agomelatine	68	59	9.1	tho	346	307	8.7
nextday	68	61	9.1	l-dopa	72	55	8.7
pre-workout	77	69	9.1	psychoactive	119	102	8.7
apap	90	68	9.1	metabolizer	68	42	8.6
pitolisant	67	46	9.1	powdercity	63	57	8.6
picamilon	67	54	9.1	finil	63	59	8.6
tren	77	46	9	debilitating	200	196	8.6
potent	539	483	9	nubrain	63	43	8.6
stabilizer	200	172	9	sorem	63	48	8.6
withdrawl	86	72	9	anti-anxiety	79	75	8.6
procrastinate	122	111	9	medsforbitcoin	62	47	8.5
taurine	89	75	9	hypo	91	78	8.5
cocaine	460	375	9	cns	218	188	8.5
coffee	3698	2869	9	receptor	920	594	8.4
lasix	111	13	8.9	phenergan	87	8	8.4
aggrenox	72	3	8.9	oxycodone	188	110	8.4
moclobemide	68	48	8.9	awhile	453	429	8.4
ect	202	158	8.9	diagnosis	1783	1410	8.4
coq10	92	75	8.9	celebrex	111	15	8.4
tablet	1965	1440	8.9	hi	3012	2967	8.4
atypical	193	171	8.9	hab	94	89	8.3
lupus	199	145	8.8	brainfog	61	53	8.3
selank	65	55	8.8	prescript	64	42	8.3
nri	100	75	8.8	delivery	4831	554	8.3
preworkout	65	52	8.8	overdo	160	154	8.2
rs4680	64	40	8.7	trileptal	66	49	8.2

Keyword	Freq	No Posts	Score
artichoke	127	96	8.2
didnt	552	490	8.1
dmd	90	78	8.1
bpc	73	53	8.1
tca	104	90	8.1
nzt-48	59	56	8.1
serotonergic	70	62	8.1
gingko	68	61	8.1
sinemet	66	43	8.1
citicoline	60	50	8.1
taper	476	374	8.1
lifesaver	96	96	8
soma	240	127	8
hypopnea	63	46	8
imovane	83	29	8
brain	5037	3738	7.9
fatique	59	56	7.9
ugh	184	176	7.9
horrible	858	815	7.9
dizziness	181	168	7.9
tecfidera	59	51	7.9
pulmonologist	68	59	7.9
dont	1719	1352	7.9
dizzy	196	179	7.9
liver	1153	863	7.9
iop	108	94	7.8
olanzapine	72	52	7.8
auto-merged	58	55	7.8
kinda	702	661	7.8

Keyword	Freq	No Posts	Score
tbh	108	107	7.8
titration	87	79	7.8
metabolise	68	59	7.8
pharmacist	441	376	7.8
shipping	1381	572	7.8
prami	56	47	7.8
pregnenolone	60	45	7.7
hydrocodone	220	114	7.7
rrms	60	56	7.7
inositol	70	62	7.7
stablon	56	46	7.7
d-amphetamine	57	50	7.7
zaleplon	58	13	7.7
amitriptyline	75	62	7.7
n-acetyl	59	41	7.7
cordarone	60	3	7.7
downer	120	107	7.7
mxe	56	33	7.7
prl-8-53	55	43	7.6
requip	65	51	7.6
wean	196	183	7.6
ir	433	332	7.6
godsend	87	85	7.6
exelon	83	5	7.6
mdd	73	61	7.6
doze	112	104	7.6
tapatalk	102	100	7.5
benzodiazepine	133	114	7.5
withdrawal	964	744	7.5

Keyword	Freq	No Posts	Score
adrafanil	54	49	7.5
imo	410	398	7.5
addicted	112	110	7.5
stimulation	469	419	7.5
anti-psychotic	63	59	7.5
perscribed	56	53	7.5
premarin	107	8	7.5
ppms	57	37	7.5
tachycardia	93	81	7.5
hypersomnolence	54	47	7.5
allergic	363	312	7.5
wake-promoting	54	38	7.5
disorder	2736	2187	7.4
scrip	82	77	7.4
antagonist	243	180	7.4
chemo	143	103	7.4
tpe	66	33	7.4
jayster	53	51	7.4
d2	152	64	7.4
pee	228	192	7.3
dlpa	53	42	7.3
anecdotally	72	72	7.3
buproprion	53	48	7.3
regimen	333	308	7.3
monoamine	68	59	7.3
overstimulated	56	54	7.3
spasm	165	133	7.3
clonazolam	52	36	7.3
gp	806	616	7.3

Keyword	Freq	No Posts	Score
bloodwork	61	54	7.2
ibuprofen	112	97	7.2
hubby	179	157	7.2
formulary	85	70	7.2
undiagnosed	93	88	7.2
awakeness	52	48	7.2
procrastination	100	87	7.2
prosteride	51	1	7.2
luvox	59	39	7.2
overnigh	51	41	7.1
addict	570	477	7.1
cardio	173	153	7.1
mct	70	64	7.1
adhd-pi	51	47	7.1
tire	1540	1382	7.1
neurotoxic	64	59	7.1
dehydrate	150	131	7.1
methadone	189	88	7.1
zolpidem	162	76	7.1
modadropship	50	35	7
redbull	54	50	7
modanifil	50	44	7
bupe	55	31	7
feel	23117	14555	7
plavix	101	8	7
btw	489	477	7
curcumin	83	62	7
vitamin	1425	1162	7
cognitively	79	76	6.9

Keyword	Freq	No Posts	Score
yawn	193	171	6.9
timestamp	138	33	6.9
stomach	1076	923	6.9
provigal	49	44	6.9
fyi	148	148	6.9
ritilan	49	42	6.9
zyban	65	34	6.9
excessive	921	836	6.9
anymore	1393	1305	6.9
reship	50	45	6.9
paralysis	192	152	6.9
amphetamine-like	50	49	6.8
sedative	115	102	6.8
recreationally	60	58	6.8
shift-work	49	47	6.8
oxy	77	64	6.8
noticable	67	62	6.8
biohacking	50	39	6.8
dmt	67	54	6.8
me	50185	27300	6.8
ime	67	65	6.7
cant	678	589	6.7
gwern	48	46	6.7
mylan	63	41	6.7
grogginess	50	45	6.7
oversleep	56	53	6.7
try	24823	17820	6.7
neurotoxicity	61	53	6.7
restful	117	109	6.7

Keyword	Freq	No Posts	Score
energize	266	242	6.7
capsule	549	343	6.7
fetzima	47	38	6.7
latency	260	202	6.6
synergistic	103	98	6.6
talkative	78	74	6.6
adaptogen	53	49	6.6
shroom	58	43	6.6
tricyclic	63	59	6.6
definitely	3177	2884	6.6
buy	11858	2872	6.6
metabolite	166	127	6.6
subutex	51	31	6.6
insurance	5338	3944	6.6
bi-polar	59	47	6.6
non-addictive	50	47	6.6
pseudoephedrine	58	48	6.6
sct	70	41	6.6
neurofeedback	56	38	6.6
pro-drug	48	46	6.6
fasoracetam	46	39	6.5
nsi	59	36	6.5
pharmacological	121	107	6.5
p-doc	46	27	6.5
contraindicate	74	67	6.5
hypothalamus	77	62	6.5
i	315187	56752	6.5
apnoea	56	36	6.5
hallucinate	71	64	6.5

Keyword	Freq	No Posts	Score
proair	46	1	6.5
pemoline	46	29	6.5
anti-fatigue	48	44	6.5
honestly	828	801	6.4
dehydration	154	144	6.4
prolintane	45	28	6.4
unifiram	45	37	6.4
tylenol	120	111	6.4
imipramine	51	40	6.4
kava	69	50	6.4
psychedelic	192	165	6.4
weird	1088	993	6.4
googled	92	89	6.4
cortisol	122	91	6.4
tried	79	77	6.4
sparingly	105	100	6.4
iirc	108	108	6.3
coke	280	231	6.3
mastercard	197	112	6.3
self-medicating	48	45	6.3
carb	316	248	6.3
tolerate	553	515	6.3
adrenergic	54	47	6.3
risperidone	53	44	6.3
prescripton	44	32	6.3
fda	808	679	6.3
stevens-johnson	47	47	6.3
dexadrine	44	40	6.3
potency	161	149	6.3

Keyword	Freq	No Posts	Score
subside	181	174	6.2
numbness	104	96	6.2
dha	103	97	6.2
ginko	47	41	6.2
alert	1585	1463	6.2
jitteriness	45	45	6.2
productive	888	813	6.2
worse	496	485	6.2
nuvagil	43	33	6.2
supp	130	124	6.2
decaf	55	46	6.2
itchy	125	111	6.2
ndri	44	40	6.2
caffeinated	62	61	6.2
substance	1930	1485	6.2
excitotoxicity	45	34	6.1
hppd	44	31	6.1
c-pap	43	37	6.1
advil	56	52	6.1
focused	180	173	6.1
tl	126	125	6.1
fogginess	44	43	6.1
morning	5895	4772	6
eph	74	44	6
yea	232	217	6
paradoxical	101	87	6
wont	379	359	6
synthroid	85	69	6
amped	56	55	6

Keyword	Freq	No Posts	Score
hypotension	69	58	6
polyphasic	42	24	6
exhaust	951	870	6
provogil	41	33	6
op	453	416	6
nefiracetam	41	36	5.9
wouldnt	145	134	5.9
prob	99	96	5.9
binge	145	124	5.9
ampakine	41	35	5.9
gad	78	75	5.9
low-dose	68	65	5.9
numb	183	159	5.9
dr	2256	1683	5.9
alprazolam	115	89	5.9
quetiapine	48	38	5.9
sjogren	46	39	5.9
betaseron	42	36	5.9
citrate	83	69	5.9
ssdi	59	44	5.9
my	77353	31950	5.9
chronic	1227	1089	5.9
hydrated	79	78	5.9
yr	387	327	5.9
unmedicated	43	42	5.9
keto	89	69	5.8
modafinl	40	37	5.8
cipla	46	27	5.8
idra-21	40	26	5.8

Keyword	Freq	No Posts	Score
arousal	125	96	5.8
vicodin	154	74	5.8
iud	66	47	5.8
prefrontal	71	56	5.8
tremor	144	117	5.8
afaik	67	66	5.8
kamagra	76	14	5.8
redosing	40	39	5.8
addiction	828	657	5.8
speedy	240	223	5.8
spansule	40	32	5.8
snort	168	120	5.8
restless	193	182	5.8
poop	160	137	5.7
nervousness	95	85	5.7
agmatine	40	28	5.7
isnt	191	185	5.7
worsen	323	301	5.7
googling	87	86	5.7
script	2339	1310	5.7
paranoid	181	171	5.7
mdpv	40	30	5.7
quit	885	748	5.7
dysthymia	43	35	5.7
newmind	39	35	5.7
inducer	55	51	5.7
milligram	106	87	5.7
stimulant-like	39	39	5.7
auth	68	54	5.7

Keyword	Freq	No Posts	Score
glycinate	40	39	5.7
r-enantiomer	39	34	5.7
goodrx	40	33	5.7
diphenhydramine	45	39	5.7
ltp	48	29	5.7
ashwaganda	39	36	5.7
vigil	163	142	5.7
flonase	47	40	5.6
dysautonomia	41	31	5.6
anti-histamine	41	37	5.6
lioresal	41	2	5.6
tremendously	186	181	5.6
kaufen	42	25	5.6
hyperactive	67	63	5.6
drink	3716	2772	5.6
amidate	38	35	5.6
steroid	402	304	5.6
stimulatory	44	41	5.6
imitrex	64	8	5.6
wk	84	64	5.6
catecholamine	48	42	5.5
cuz	156	140	5.5
testosterone	287	230	5.5
microsleep	38	31	5.5
tinnitus	90	60	5.5
cholinergic	49	44	5.5
expensive	2081	1920	5.5
intas	38	32	5.5
mentally	453	427	5.5

Keyword	Freq	No Posts	Score
longterm	62	51	5.5
overstimulation	41	38	5.5
t3	114	88	5.5
re-uptake	39	35	5.5
fladrafinil	37	30	5.5
suicidal	177	165	5.5
hesitant	153	149	5.5
carnitine	46	39	5.5
orexinergic	37	16	5.5
sudafed	41	31	5.5
crap	516	484	5.4
redditor	42	39	5.4
downregulation	46	43	5.4
anyone	6405	5638	5.4
dexmethylphenidate	37	31	5.4
tolerable	95	91	5.4
myself	3909	3317	5.4
narcotic	212	186	5.4
thats	206	192	5.4
maybe	5647	4811	5.4
dnp	56	36	5.3
uni	172	148	5.3
atleast	88	83	5.3
inhibitor	410	346	5.3
cipralex	37	17	5.3
blister	212	178	5.3
ppl	121	99	5.3
take	56523	30681	5.3
dextro	38	30	5.3

Keyword	Freq	No Posts	Score
motivated	169	167	5.3
cyproheptadine	40	9	5.3
bioperine	37	21	5.3
vasoconstriction	42	40	5.3
hungover	45	44	5.3
schizophrenia	187	129	5.3
brugada	37	37	5.3
couldnt	124	110	5.3
orally	132	113	5.3
sorry	1882	1790	5.3

Keyword	Freq	No Posts	Score
seizure	403	301	5.2
pharmaceuticals	242	205	5.2
oxybate	36	36	5.2
na-semax	35	28	5.2
zyrem	35	25	5.2
resmed	42	28	5.2
d-amp	35	31	5.2
clozapine	43	29	5.2
til	189	178	5.2
ketosis	46	34	5.2

Appendix Q Top 100 KeyTerms and Themes

KeyTerm	Freq	No of Posts	Score	Theme
sleep study	1273	1023	130.9	Investigation
overnight delivery	969	197	98	Acquisition
saturday delivery	766	213	90.3	Acquisition
brain fog	854	714	89.5	Reason
sleep apnea	1633	1315	87.2	Reason
daytime sleepiness	779	713	83.5	Reason
low dose	835	767	63.6	Dosage
sleep disorder	597	547	59.3	Reason
shift work	563	524	56.4	Mixed
next day delivery	554	193	52.2	Acquisition
prescription buy	404	141	48.3	Acquisition
day delivery	585	202	47	Acquisition
chronic fatigue	654	598	45.7	Reason
buy provigil	371	46	45.6	Acquisition
ms fatigue	369	323	45.3	Reason
side effect	1214	1120	42.9	SideEffect
half life	408	365	41.5	Dosage
buy nuvigil	330	19	40.8	Acquisition
excessive daytime sleepiness	345	334	40.5	Reason
div itemprop	325	325	40.3	NotRel
sleep doctor	325	281	39.5	HCP
sleep specialist	320	293	38.1	HCP
post entry-content	304	304	37.7	NotRel
online pharmacy	658	296	36.3	Acquisition
cheap provigil	286	33	35.4	Acquisition
prescription provigil	276	37	34.3	Acquisition
empty stomach	414	383	34.3	Dosage

KeyTerm	Freq	No of Posts	Score	Theme
overnight shipping	299	165	34.2	Acquisition
generic provigil	254	183	31.4	Mixed
sleep doc	252	201	31.2	HCP
prescription nuvigil	245	22	30.6	Acquisition
online buy	294	143	30.1	Acquisition
non prescription	260	149	30	Acquisition
delivery provigil	232	28	29	Acquisition
provigil provigil	229	26	28.7	Modafinil
work disorder	226	207	28.1	Reason
excessive sleepiness	231	202	28	Reason
cheap nuvigil	223	17	27.9	Acquisition
cod saturday	225	123	27.9	Acquisition
cod delivery	222	111	27.6	Acquisition
sleep schedule	246	227	27.5	Mixed
overnight cod	217	121	27	Acquisition
taking provigil	214	208	26.8	Dosage
overnight fedex	213	128	26.6	Acquisition
delivery cod	212	106	26.5	Acquisition
prescription cod	212	124	26.5	Acquisition
small dose	241	225	26.2	Dosage
sleep deprivation	340	296	25.5	Mixed
shift work disorder	202	189	25.3	Reason
extreme fatigue	222	216	25.2	Reason
delivery buy	201	98	25.1	Acquisition
nuvigil cod	196	20	24.7	Acquisition
high dose	300	287	24.4	Dosage
prescription overnight delivery	191	124	23.9	Acquisition

KeyTerm	Freq	No of Posts	Score	Theme
order provigil	188	31	23.6	Acquisition
provigil cod	186	24	23.5	Acquisition
fedex delivery	187	113	23.4	Acquisition
work sleep	187	180	23.2	Mixed
cod nuvigil	182	17	23	Acquisition
shift work sleep	175	168	21.9	Reason
delivery nuvigil	172	19	21.8	Acquisition
obstructive sleep apnea	227	212	21.8	Reason
work sleep disorder	173	166	21.6	Reason
obstructive sleep	234	219	21.4	Reason
first dose	214	205	20.9	Dosage
day supply	176	162	20.5	Acquisition
shift work sleep disorder	162	155	20.3	Reason
free fedex	160	84	20.2	Acquisition
nuvigil nuvigil	158	19	20.1	Modafinil
dry mouth	239	196	19.5	SideEffect
miracle drug	159	158	19.1	Outcome
cod provigil	150	29	19.1	Acquisition
fish oil	300	278	19	OtherInt
cod saturday delivery	150	103	19	Acquisition
sleep latency	154	129	18.9	Investigation
extended release	161	146	18.9	Dosage
order nuvigil	148	22	18.9	Acquisition
good sleep	189	182	18.8	Effect

KeyTerm	Freq	No of Posts	Score	Theme
prescription order	152	104	18.8	Acquisition
term memory	185	162	18.8	Effect
severe fatigue	151	144	18.5	Reason
cognitive enhancement	150	130	18.4	?
online cod	152	95	18.1	Acquisition
second dose	164	140	18	Dosage
good luck	1585	1564	18	?
enough sleep	183	174	17.7	?
taking nuvigil	137	136	17.6	Dosage
insurance company	705	604	17.4	Acquisition
fedex cod	136	88	17.4	Acquisition
prescription next day	136	103	17.3	Acquisition
quote name	139	30	17.2	NotRel
post count	148	116	16.7	NotRel
short term memory	151	143	16.7	Reason
fatigue syndrome	193	179	16.6	Reason
smart drug	130	116	16.4	?
cod next day	125	77	16	Acquisition
prescription fedex	125	96	16	Acquisition
chronic fatigue syndrome	183	172	16	Reason
overnight buy	125	90	16	Acquisition
prescription next day delivery	124	95	15.9	Acquisition

Appendix R Top 1000 KeyTerms

KeyTerm	Freq	No Posts	Score	KeyTerm	Freq	No Posts	Score
sleep study	1273	1023	130.9	generic provigil	254	183	31.4
overnight delivery	969	197	98	sleep doc	252	201	31.2
saturday delivery	766	213	90.3	prescription nuvigil	245	22	30.6
brain fog	854	714	89.5	online buy	294	143	30.1
sleep apnea	1633	1315	87.2	non prescription	260	149	30
daytime sleepiness	779	713	83.5	delivery provigil	232	28	29
low dose	835	767	63.6	provigil provigil	229	26	28.7
sleep disorder	597	547	59.3	work disorder	226	207	28.1
shift work	563	524	56.4	excessive sleepiness	231	202	28
next day delivery	554	193	52.2	cheap nuvigil	223	17	27.9
prescription buy	404	141	48.3	cod saturday	225	123	27.9
day delivery	585	202	47	cod delivery	222	111	27.6
chronic fatigue	654	598	45.7	sleep schedule	246	227	27.5
buy provigil	371	46	45.6	overnight cod	217	121	27
ms fatigue	369	323	45.3	taking provigil	214	208	26.8
side effect	1214	1120	42.9	overnight fedex	213	128	26.6
half life	408	365	41.5	delivery cod	212	106	26.5
buy nuvigil	330	19	40.8	prescription cod	212	124	26.5
excessive daytime sleepiness	345	334	40.5	small dose	241	225	26.2
div itemprop	325	325	40.3	sleep deprivation	340	296	25.5
sleep doctor	325	281	39.5	shift work disorder	202	189	25.3
sleep specialist	320	293	38.1	extreme fatigue	222	216	25.2
post entry-content	304	304	37.7	delivery buy	201	98	25.1
online pharmacy	658	296	36.3	nuvigil cod	196	20	24.7
cheap provigil	286	33	35.4	high dose	300	287	24.4
prescription provigil	276	37	34.3	prescription overnight delivery	191	124	23.9
empty stomach	414	383	34.3	order provigil	188	31	23.6
overnight shipping	299	165	34.2	provigil cod	186	24	23.5

KeyTerm	Freq	No Posts	Score	KeyTerm	Freq	No Posts	Score
fedex delivery	187	113	23.4	enough sleep	183	174	17.7
work sleep	187	180	23.2	taking nuvigil	137	136	17.6
cod nuvigil	182	17	23	insurance company	705	604	17.4
shift work sleep	175	168	21.9	fedex cod	136	88	17.4
delivery nuvigil	172	19	21.8	prescription next day	136	103	17.3
obstructive sleep apnea	227	212	21.8	quote name	139	30	17.2
work sleep disorder	173	166	21.6	post count	148	116	16.7
obstructive sleep	234	219	21.4	short term memory	151	143	16.7
first dose	214	205	20.9	fatigue syndrome	193	179	16.6
day supply	176	162	20.5	smart drug	130	116	16.4
shift work sleep disorder	162	155	20.3	cod next day	125	77	16
free fedex	160	84	20.2	prescription fedex	125	96	16
nuvigil nuvigil	158	19	20.1	chronic fatigue syndrome	183	172	16
dry mouth	239	196	19.5	overnight buy	125	90	16
miracle drug	159	158	19.1	prescription next day delivery	124	95	15.9
cod provigil	150	29	19.1	social anxiety	189	169	15.9
fish oil	300	278	19	online provigil	122	33	15.7
cod saturday delivery	150	103	19	fast delivery	184	117	15.6
sleep latency	154	129	18.9	generic version	148	133	15.6
extended release	161	146	18.9	sleep paralysis	141	113	15.5
order nuvigil	148	22	18.9	bipolar depression	128	110	15.4
good sleep	189	182	18.8	bad side	155	152	15.2
prescription order	152	104	18.8	delivery order	124	86	15.1
term memory	185	162	18.8	cash delivery	118	99	15.1
severe fatigue	151	144	18.5	cold turkey	156	142	15
cognitive enhancement	150	130	18.4	mood stabilizer	122	113	14.9
online cod	152	95	18.1	wonder drug	125	124	14.9
second dose	164	140	18	mental clarity	132	127	14.9
good luck	1585	1564	18	full dose	127	114	14.9

KeyTerm	Freq	No Posts	Score
online next day	115	85	14.8
name brand	154	125	14.6
off label	116	112	14.5
placebo effect	147	136	14.4
online nuvigil	111	15	14.4
free credit	142	3	14.3
next day	1937	895	14.2
sample pack	118	113	14.2
ms nurse	111	102	14.2
reuptake inhibitor	120	107	14.2
heart rate	542	440	14.1
overnight nuvigil	108	17	14
sleep quality	142	132	14
sleep cycle	123	117	14
sleep hygiene	118	113	14
prior authorization	137	121	14
cod buy	106	76	13.8
month supply	114	107	13.7
new doctor	126	119	13.6
term use	131	124	13.5
cod fedex	104	70	13.5
negative side	170	169	13.5
prescription overnight shipping	102	77	13.3
dopamine reuptake	103	93	13.2
overnight provigil	101	23	13.2
tunnel vision	123	115	13.1
controlled substance	141	130	13
shipping buy	99	72	12.9
generic form	113	105	12.9

KeyTerm	Freq	No Posts	Score
provigil next day	98	24	12.8
free shipping	309	179	12.8
brand name	411	343	12.8
max dose	99	98	12.7
long term use	117	114	12.7
online overnight delivery	98	79	12.6
online order	123	89	12.5
sleep test	96	88	12.3
cognitive enhancer	94	86	12.3
pharmacy cod	93	59	12.2
circadian rhythm	118	98	12.1
blood work	131	121	12.1
life saver	110	108	12
working memory	156	127	12
fedex shipping	90	71	11.8
energy level	153	148	11.8
mental fatigue	97	90	11.8
energy boost	101	96	11.8
day time	143	129	11.7
online doctor	92	74	11.7
online consultation	120	95	11.6
deep sleep	147	114	11.6
online doctor consultation	88	71	11.6
abuse potential	96	94	11.5
doctor consultation	89	71	11.5
tracking number	119	94	11.5
sleep debt	90	77	11.4
fedex provigil	86	24	11.4
other medication	95	92	11.2

KeyTerm	Freq	No Posts	Score	KeyTerm	Freq	No Posts	Score
energy drink	113	103	11.2	sleep pattern	80	76	9.7
sleep dr	84	67	11.1	stimulant effect	74	71	9.7
same effect	211	208	11	common side effect	82	81	9.7
long half	85	81	11	new medication	82	78	9.7
taking anything	95	92	10.9	pharmacy provigil	72	32	9.7
promoting agent	82	81	10.8	online drug	76	76	9.7
trouble sleeping	99	98	10.8	allergic reaction	143	129	9.6
night shift	135	126	10.7	free credit report	77	2	9.6
resistant depression	82	77	10.7	bulletproof coffee	72	68	9.6
online prescription	93	56	10.7	bad batch	75	70	9.6
same drug	91	88	10.7	buy cod	71	48	9.5
long half life	81	77	10.7	day shipping	76	60	9.4
daytime fatigue	80	79	10.5	drinking coffee	93	91	9.4
choline bitartrate	79	66	10.5	huge difference	194	191	9.4
much caffeine	85	82	10.5	appetite suppression	72	69	9.4
brain chemistry	95	90	10.4	hour energy	70	65	9.3
cognitive function	128	118	10.4	adhd medication	72	68	9.3
choline source	77	72	10.3	delivery fedex	69	59	9.3
cod order	77	53	10.3	bad reaction	75	74	9.2
common side	123	121	10.2	daily use	116	110	9.2
only side	85	85	10.2	new drug	139	135	9.2
cod next day delivery	76	62	10.2	similar experience	100	100	9.2
hour nap	78	70	10.2	wakefulness promoting agent	68	67	9.2
work day	130	116	10.1	caffeine pill	68	60	9.2
nuvigil next day	75	18	10.1	label use	69	68	9.1
taking adderall	74	73	9.9	prescription drug	193	189	9.1
sleep clinic	75	69	9.8	same dose	74	73	9
cod pharmacy	73	60	9.8	panic attack	105	98	9
prescription fast delivery	73	71	9.8	prior prescription	71	66	9

KeyTerm	Freq	No Posts	Score	KeyTerm	Freq	No Posts	Score
online cash	72	64	8.9	recommended dose	76	72	8.3
many side	75	74	8.9	other stimulant	61	60	8.3
last dose	74	71	8.8	cheap buy	62	55	8.3
study aid	67	66	8.8	noticeable effect	66	64	8.3
normal person	94	91	8.8	saturday delivery provigil	60	21	8.2
mild stimulant	65	65	8.7	shipping provigil	60	22	8.2
green tea	186	168	8.7	canadian pharmacy	120	89	8.2
mental energy	76	74	8.7	fedex buy	60	51	8.2
discount card	74	64	8.6	overnight delivery cod	60	47	8.2
mild sleep	64	62	8.6	daytime drowsiness	61	56	8.2
alpha gpc	63	55	8.6	only side effect	62	62	8.2
birth control	253	200	8.6	usa pharmacy	60	50	8.1
personal experience	274	271	8.6	sexual side	65	56	8.1
dr approval	63	54	8.6	taking ritalin	60	60	8.1
caffeine intake	69	67	8.6	normal dose	62	61	8.1
much coffee	69	68	8.6	poor sleep	71	67	8.1
restful sleep	74	68	8.5	next appointment	70	69	8.1
blood pressure	767	607	8.5	mild sleep apnea	59	57	8.1
much sleep	73	70	8.5	morning dose	60	59	8.1
vitamin d	184	167	8.5	artichoke extract	59	51	8.1
sex drive	103	84	8.5	entire day	116	113	8
only drug	67	65	8.5	long half-life	62	60	8
overnight delivery buy	62	45	8.5	ms specialist	59	55	8
money order	115	87	8.4	other drug	87	86	8
getting enough sleep	71	69	8.4	fedex nuvigil	58	16	8
sleep aid	71	70	8.4	urine smell	59	58	8
red bull	65	61	8.4	new med	59	57	8
prescription buy provigil	61	25	8.4	major depression	83	71	8
severe sleep	63	61	8.4	dopamine reuptake inhibitor	58	55	8

KeyTerm	Freq	No Posts	Score	KeyTerm	Freq	No Posts	Score
normal sleep	62	60	7.9	verbal fluency	57	49	7.6
sleeping pill	68	62	7.9	quality sleep	62	62	7.6
memory loss	113	91	7.9	instant release	56	53	7.6
prescription buy nuvigil	57	18	7.9	low energy	112	104	7.6
saturday delivery cod	57	49	7.9	shift disorder	54	52	7.5
next day cod	57	46	7.9	provigil prescription	54	25	7.5
day cod	57	46	7.9	free overnight fedex	54	48	7.5
treatment resistant depression	57	56	7.8	nootropic stack	54	50	7.5
free consultation	94	72	7.8	severe depression	72	70	7.5
provigil fedex	56	23	7.8	other stuff	180	172	7.5
nuvigil fedex	56	18	7.8	express delivery	60	47	7.5
provigil overnight delivery	56	19	7.8	anti depressant	55	54	7.4
generic nuvigil	56	24	7.8	adrenal fatigue	58	38	7.4
provigil buy	56	30	7.8	bipolar disorder	145	118	7.4
sun pharma	56	50	7.8	nuvigil overnight delivery	53	17	7.4
prior script	56	42	7.7	half dose	54	51	7.4
high tolerance	63	61	7.7	overnight fedex delivery	53	47	7.4
sleeping disorder	58	51	7.7	stimulant medication	55	54	7.4
time release	61	57	7.7	fibro fog	53	49	7.3
hard time	370	358	7.7	generic brand	56	54	7.3
same boat	96	96	7.7	only thing	753	734	7.3
new doc	58	56	7.7	great drug	53	52	7.3
taking something	67	66	7.6	free overnight fedex delivery	52	47	7.3
local pharmacy	65	63	7.6	online purchase	65	54	7.3
rx cod	55	53	7.6	narcolepsy drug	52	51	7.3
same experience	84	81	7.6	sleep attack	52	48	7.3
order prescription	56	54	7.6	bad anxiety	53	52	7.3
new insurance	63	62	7.6	ex delivery	52	52	7.3
weight gain	201	171	7.6	drug company	73	65	7.2

KeyTerm	Freq	No Posts	Score
physical fatigue	54	50	7.2
nerve pain	65	59	7.2
chest pain	106	94	7.2
bed time	64	63	7.2
study drug	57	55	7.2
prescribed provigil	51	50	7.2
right dose	56	55	7.2
extra boost	60	60	7.2
provigil order	51	24	7.2
pharmacy fedex	51	49	7.2
early afternoon	90	88	7.2
free prescription	53	50	7.1
mood lift	51	50	7.1
mental performance	55	51	7.1
online cheap	54	47	7.1
low dosage	53	52	7.1
major side	54	52	7
whole day	139	134	7
effect profile	53	50	7
taking moda	50	50	7
prescription overnight cod	50	49	7
purchase provigil	50	24	7
nasty side	54	54	7
patient assistance	53	50	7
cheap cod	50	48	7
serotonin syndrome	55	49	7
other sleep	53	52	7
next day shipping	50	43	7
prescription uk	50	29	7

KeyTerm	Freq	No Posts	Score
lipoic acid	56	55	6.9
recreational value	51	48	6.9
rx provigil	49	21	6.9
possible side	77	77	6.9
second opinion	84	77	6.9
saturday delivery fedex	49	47	6.9
script fedex	49	48	6.9
online pharmacy cod	49	37	6.9
script buy	49	41	6.9
first couple	120	119	6.9
buy discount	50	42	6.9
provigil cash	48	23	6.8
drug interaction	55	50	6.8
non prescription fedex	48	47	6.8
cod saturday delivery fedex	48	47	6.8
few side	56	56	6.8
joint pain	95	80	6.8
extra energy	61	58	6.8
online saturday	48	46	6.8
slight headache	49	49	6.7
serious side	81	80	6.7
beta blocker	52	48	6.7
atypical depression	48	38	6.7
rx nuvigil	47	18	6.7
purchase nuvigil	47	17	6.7
nuvigil order	47	18	6.7
online pharmacy fedex	47	46	6.7
quick delivery	58	53	6.7
daily dose	79	72	6.7

KeyTerm	Freq	No Posts	Score
dopamine transporter	49	45	6.6
current stack	48	45	6.6
side note	129	126	6.6
restless leg	52	51	6.6
opposite effect	74	74	6.6
buy fioricet	47	5	6.6
pill form	53	52	6.6
similar effect	68	67	6.6
grapefruit juice	57	45	6.6
saturday delivery nuvigil	46	14	6.6
script provigil	46	21	6.6
provigil online provigil	46	17	6.6
sweet spot	102	98	6.6
online pharmacy fedex cod	46	45	6.6
pharmacy fedex cod	46	45	6.6
discount fedex	46	43	6.6
magic pill	50	50	6.6
enough energy	75	75	6.5
side effect profile	48	45	6.5
first post	126	121	6.5
new sleep	47	44	6.5
little sleep	55	53	6.5
proper sleep	48	47	6.5
histamine release	47	27	6.5
nicotine gum	48	45	6.5
sleep architecture	46	36	6.5
shipping nuvigil	45	17	6.4
provigil saturday	45	22	6.4
pharmacy nuvigil	45	17	6.4

KeyTerm	Freq	No Posts	Score
prescription modafinil	45	18	6.4
provigil buy provigil	45	25	6.4
buy t	45	3	6.4
buying provigil	45	26	6.4
long story	126	126	6.4
adderall xr	46	42	6.4
prescription purchase	45	39	6.4
possible quality	52	52	6.4
sleep time	53	44	6.4
attention span	74	70	6.4
shipping order	45	40	6.4
night sleep	51	49	6.4
cognitive performance	56	51	6.4
long term	1000	908	6.3
doing anything	172	171	6.3
prescription saturday delivery	44	43	6.3
mail order	96	76	6.3
wakefulness-promoting agent	44	41	6.3
high dosage	47	47	6.3
online saturday delivery	44	44	6.3
prescription saturday	44	43	6.3
single dose	69	65	6.3
good mood	69	69	6.3
new prescription	50	49	6.3
occasional use	53	52	6.3
related fatigue	44	43	6.3
night time	106	100	6.3
short term	414	384	6.3
research chemical	44	42	6.3

KeyTerm	Freq	No Posts	Score
drug test	76	65	6.2
expensive drug	45	44	6.2
dopamine release	46	45	6.2
large dose	50	46	6.2
off-label use	49	48	6.2
other day	348	332	6.2
mental fog	44	44	6.2
cod overnight delivery	43	31	6.2
generic name	73	71	6.2
family doctor	64	58	6.2
stimulant use	44	41	6.2
effective dose	52	48	6.1
multiple sleep	43	39	6.1
latency test	43	39	6.1
multiple sclerosis	136	130	6.1
blood test	111	94	6.1
free online doctor consultation	42	38	6.1
free online doctor	42	38	6.1
alpha lipoic acid	44	43	6.1
overnight order	42	37	6.1
buy codeine	42	10	6.1
grad school	76	67	6
whole pill	42	40	6
following text	58	55	6
sleep issue	42	42	6
old male	55	55	6
doctor today	44	42	6
med school	50	46	6
sleep lab	43	36	6

KeyTerm	Freq	No Posts	Score
day delivery provigil	41	23	6
nuvigil prescription	41	19	6
pharmacy cod saturday	41	40	6
narcolepsy diagnosis	41	39	5.9
day time sleepiness	41	37	5.9
time sleepiness	41	37	5.9
rx buy	41	32	5.9
cross tolerance	41	38	5.9
depressive disorder	59	57	5.9
slow release	49	46	5.9
mid day	46	42	5.9
standard dose	44	42	5.9
dopamine agonist	42	39	5.9
cpap machine	41	38	5.9
same problem	164	162	5.8
delivery buy nuvigil	40	17	5.8
nuvigil cash	40	17	5.8
muscle relaxer	42	39	5.8
membership buy	40	34	5.8
pharmacy cod saturday delivery	40	40	5.8
buy lamisil	40	1	5.8
bad sleep	41	34	5.8
good diet	49	47	5.8
sleep latency test	40	37	5.8
hair loss	140	104	5.8
debilitating fatigue	40	40	5.8
light therapy	52	46	5.8
time sleep	40	36	5.8
strong coffee	44	42	5.7

KeyTerm	Freq	No Posts	Score
starting dose	45	45	5.7
potential side	52	48	5.7
work shift	43	42	5.7
nuvigil cod nuvigil	39	14	5.7
generic modafinil	39	38	5.7
adhd treatment	40	38	5.7
prescription doxycycline	39	4	5.7
non responder	39	37	5.7
delivery saturday	39	34	5.7
new neuro	39	38	5.7
low price	114	69	5.7
new neurologist	39	38	5.7
off label use	39	39	5.7
maximum dose	44	40	5.6
prefrontal cortex	60	46	5.6
official diagnosis	40	38	5.6
sleep medicine	49	43	5.6
good drug	40	38	5.6
second day	153	148	5.6
pro vigil	38	37	5.6
next day fedex	38	38	5.6
day fedex	38	38	5.6
racing heart	41	38	5.6
daytime tiredness	38	36	5.6
pharmacy buy	38	31	5.6
physical energy	44	44	5.5
only medication	39	38	5.5
bad idea	149	149	5.5
online overnight shipping	38	33	5.5

KeyTerm	Freq	No Posts	Score
mental alertness	41	39	5.5
discount coupon	43	43	5.5
taking half	38	38	5.5
manic episode	40	38	5.5
unpleasant side	41	40	5.5
nuvigil online nuvigil	37	12	5.5
script nuvigil	37	14	5.5
prescription cheap provigil	37	19	5.5
informative x	37	37	5.5
nuvigil buy	37	15	5.5
overnight free delivery	37	37	5.5
choline supplement	37	36	5.5
rx order	37	32	5.5
cheap doxycycline	37	4	5.5
prior auth	37	32	5.5
heart race	39	39	5.5
credit report	98	2	5.4
taking caffeine	37	37	5.4
stimulating effect	39	37	5.4
first day	461	440	5.4
severe anxiety	42	41	5.4
sleep center	38	36	5.4
mood boost	37	36	5.4
bulk powder	37	32	5.4
severe sleep apnea	37	37	5.4
right dosage	38	38	5.4
priority mail	40	23	5.4
dopamine system	38	34	5.4
anxiety attack	41	40	5.4

KeyTerm	Freq	No Posts	Score
cognitive impairment	70	59	5.4
different drug	39	39	5.4
bad experience	57	56	5.4
good combo	38	38	5.4
next day delivery provigil	36	21	5.3
bromazepam online pharmacy	36	3	5.3
cheap modafinil	36	13	5.3
online overnight delivery cod	36	33	5.3
cod free fedex	36	32	5.3
cod cod	36	26	5.3
shipping cod	36	34	5.3
prescription cash	36	33	5.3
cod cash	36	33	5.3
adverse reaction	48	48	5.3
much difference	72	72	5.3
good stack	36	35	5.3
big pharma	44	42	5.3
cheap order	36	32	5.3
addiction potential	36	35	5.3
prescription medication	53	50	5.3
liver damage	52	48	5.3
low carb	60	54	5.3
minute nap	36	36	5.3
mental focus	40	36	5.3
good day	136	131	5.3
provigil saturday delivery	35	22	5.2
cash delivery cod	35	35	5.2
perscription buy	35	29	5.2
next day delivery cod	35	35	5.2

KeyTerm	Freq	No Posts	Score
day delivery cod	35	35	5.2
tolerance buildup	35	34	5.2
bad fatigue	35	35	5.2
anecdotal evidence	65	61	5.2
online overnight buy	35	35	5.2
nighttime sleep	36	33	5.2
normal life	87	82	5.2
half pill	35	34	5.2
real deal	82	81	5.2
caffeine tolerance	35	34	5.2
online consultant	35	35	5.2
afternoon nap	39	39	5.2
sleep efficiency	36	29	5.2
term side	37	37	5.2
multiple sleep latency	35	32	5.2
productive day	38	38	5.2
doctor prescription	36	30	5.1
sleep phase	36	33	5.1
reaction time	60	41	5.1
full stomach	38	36	5.1
anxiety disorder	63	60	5.1
right medication	36	35	5.1
partial agonist	36	19	5.1
shift worker	35	34	5.1
assistance program	54	52	5.1
wakefulness agent	34	34	5.1
mild depression	37	34	5.1
cog fog	34	28	5.1
pay cod	34	30	5.1

KeyTerm	Freq	No Posts	Score
online discount	37	35	5.1
central sleep apnea	35	29	5.1
only medicine	35	35	5.1
extreme tiredness	35	35	5.1
good option	78	74	5.1
same thing	654	639	5.1
restorative sleep	35	32	5.1
old self	47	47	5
central sleep	35	29	5
sugar pill	35	35	5
cognitive ability	45	44	5
year old male	41	41	5
vitamin b	42	41	5
overnight sleep	34	33	5
mild headache	34	33	5
perscription provigil	33	17	5
t ject	33	3	5
mild cataplexy	33	31	5
using provigil	33	33	5
mg prn	33	33	5
overnight shipping buy	33	29	5
cdp choline	33	29	5
medication cod	33	28	5
rx saturday	33	33	5
recreational drug	39	39	5
didnt work	37	36	5
resting heart	37	36	5
protein shake	37	31	4.9
automatic behavior	33	28	4.9

KeyTerm	Freq	No Posts	Score
body detox	34	2	4.9
resting heart rate	36	35	4.9
blister pack	35	30	4.9
big difference	147	145	4.9
jet lag	45	43	4.9
horrible side	33	33	4.9
ciltep stack	32	29	4.9
provigil cod saturday	32	17	4.9
next day fedex shipping	32	32	4.9
day fedex shipping	32	32	4.9
real provigil	32	30	4.9
online buy provigil	32	18	4.9
free fedex shipping	32	24	4.9
asleep driving	32	31	4.9
cheap fedex	32	30	4.9
regular sleep	34	32	4.9
rx saturday delivery	32	32	4.9
overnight saturday	32	31	4.9
next day cash	32	32	4.9
reputable pharmacy	32	32	4.9
buy doxycycline	32	4	4.8
panic disorder	47	40	4.8
racemic mixture	33	27	4.8
overwhelming fatigue	32	32	4.8
generic adderall	32	30	4.8
generic drug	46	40	4.8
only problem	120	120	4.8
personal use	96	83	4.8
school work	50	47	4.8

KeyTerm	Freq	No Posts	Score
addictive personality	33	33	4.8
body chemistry	36	36	4.8
time job	56	55	4.8
leg syndrome	35	35	4.8
positive effect	78	78	4.8
lunch time	56	55	4.8
old thread	39	37	4.7
buy flucloxacillin	31	2	4.7
online pharmacy provigil	31	20	4.7
narcolepsy w	31	30	4.7
daily stack	31	29	4.7
first month	85	83	4.7
overnight delivery saturday	31	30	4.7
next day cod fedex	31	30	4.7
day cod fedex	31	30	4.7
cheap norvasc	31	4	4.7
pharma grade	31	31	4.7
good insurance	34	34	4.7
mct oil	32	29	4.7
dopamine receptor	33	31	4.7
physician approval	31	31	4.7
memory recall	33	31	4.7
smoking weed	35	34	4.7
brain function	59	56	4.7
anxiolytic effect	31	30	4.7
short nap	33	30	4.7
prescription canada	31	26	4.7
long-term use	46	43	4.7
working full time	43	43	4.7

KeyTerm	Freq	No Posts	Score
day cash	32	32	4.7
tiny bit	69	63	4.7
excessive fatigue	31	30	4.7
sleep apnoea	36	25	4.6
anti anxiety	31	30	4.6
free fedex cod	30	26	4.6
online buy nuvigil	30	14	4.6
rare side	32	32	4.6
prescription overnight cod delivery	30	30	4.6
provigil next day delivery	30	17	4.6
overnight cod delivery	30	30	4.6
sale cod	30	27	4.6
sleep onset	32	24	4.6
horrible fatigue	30	30	4.6
fedex order	30	30	4.6
buy diazepam	33	8	4.6
order cod	30	28	4.6
terrible fatigue	30	30	4.6
full pill	30	29	4.6
line cash	30	25	4.6
building tolerance	30	30	4.6
hour half	30	30	4.6
rhodiola rosea	30	30	4.6
little energy	37	37	4.6
new pdoc	30	30	4.6
much energy	70	67	4.6
first week	182	171	4.6
brain training	33	19	4.6
sore throat	61	55	4.6

KeyTerm	Freq	No Posts	Score
drug holiday	30	28	4.6
acid reflux	50	46	4.6
tired feeling	30	30	4.6
daytime sleep	30	27	4.6
new dr	30	26	4.5
discount coupon code	30	30	4.5
clinical depression	39	36	4.5
nasal spray	41	39	4.5
safety profile	38	37	4.5
drink coffee	31	31	4.5
term solution	37	36	4.5
good effect	46	43	4.5
perscription nuvigil	29	14	4.5
nuvigil saturday	29	16	4.5
day nuvigil	29	15	4.5
overnight shipping provigil	29	16	4.5
life changer	30	30	4.5
b complex	31	30	4.5
l theanine	29	25	4.5
severe narcolepsy	29	29	4.5
cheap non prescription	29	27	4.5
script mastercard	29	26	4.5
first thing	339	328	4.5
short-term memory	42	38	4.5
drug provigil	29	29	4.5
insurance wont	29	29	4.5
same company	58	55	4.5
different medication	30	29	4.5
online next day delivery	29	27	4.5

KeyTerm	Freq	No Posts	Score
rapid cycling	30	27	4.5
multiple sleep latency test	29	26	4.5
restless leg syndrome	32	32	4.5
powder form	35	34	4.5
low tolerance	32	29	4.5
milk thistle	34	32	4.5
prescription xanax	29	17	4.5
noticeable side	29	29	4.5
cognitive improvement	29	27	4.5
skin rash	37	33	4.4
treating depression	32	29	4.4
constant fatigue	29	29	4.4
great focus	30	30	4.4
upset stomach	38	38	4.4
last post	84	81	4.4
good info	38	38	4.4
memory retention	30	25	4.4
long term side	30	30	4.4
active ingredient	68	63	4.4
executive function	37	35	4.4
liver toxicity	31	30	4.4
enough water	56	54	4.4
hour shift	32	31	4.4
doing nothing	89	88	4.4
much info	37	37	4.4
double dose	34	30	4.4
prescription cheap nuvigil	28	13	4.4
overnight delivery nuvigil	28	15	4.4
delivery buy provigil	28	14	4.4

KeyTerm	Freq	No Posts	Score	KeyTerm	Freq	No Posts	Score
order overnight shipping	28	28	4.4	reliable source	50	50	4.3
usa fedex	28	28	4.4	taking medication	34	34	4.3
buy norvasc	28	4	4.4	nuvigil buy nuvigil	27	15	4.3
cod purchase	28	26	4.4	saturday provigil	27	14	4.3
cod cash delivery	28	26	4.4	alarm clock	54	51	4.3
hour half life	28	28	4.4	taking piracetam	27	26	4.3
fioricet fioricet	28	3	4.4	brand provigil	27	25	4.3
high blood pressure	124	117	4.4	saturday delivery order	27	22	4.3
google search	45	42	4.4	dr consultation	27	25	4.3
day break	31	29	4.4	prescription usa fedex	27	27	4.3
hypopnea syndrome	28	23	4.4	prescription norvasc	27	4	4.3
uk delivery	28	24	4.4	watching tv	64	61	4.3
body load	28	28	4.4	little effect	52	51	4.3
prescription online pharmacy	28	27	4.3	mg adderall	27	25	4.3
prescription usa	28	28	4.3	depressive episode	30	29	4.2
doctor shopping	29	28	4.3	increasing dopamine	27	26	4.2
much effect	34	34	4.3	own experience	99	97	4.2
good experience	57	57	4.3	polyphasic sleep	27	19	4.2
therapeutic dose	30	29	4.3	regular dose	28	28	4.2
extreme pain	34	25	4.3	using caffeine	27	26	4.2
attention deficit	54	53	4.3	real stuff	32	29	4.2
hormonal birth	29	26	4.3	chronic use	29	27	4.2
first time today	31	31	4.3	good doctor	36	32	4.2
brain power	36	35	4.3	hormonal birth control	28	26	4.2
euphoric feeling	28	24	4.3	good sleep hygiene	27	27	4.2
late afternoon	86	84	4.3	doesnt work	33	33	4.2
long term memory	30	25	4.3	weird side	27	27	4.2
free month	29	26	4.3	delayed sleep	27	26	4.2
order pharmacy	28	27	4.3	new person	42	42	4.2

KeyTerm	Freq	No Posts	Score
low blood	53	50	4.2
first sleep	27	23	4.2
mid afternoon	32	31	4.2
pharmaceutical company	54	52	4.2
losing weight	65	62	4.2
free credit score	27	2	4.2
adverse side	32	32	4.2
only downside	43	40	4.1
overnight delivery provigil	26	14	4.1
provigil overnight cod	26	11	4.1
buy norvir	26	1	4.1
delivery doxycycline	26	4	4.1
taking xyrem	26	23	4.1
o cataplexy	26	25	4.1
much anxiety	29	28	4.1
generic fedex	26	24	4.1
free dr consultation	26	24	4.1
cheap overnight fedex	26	25	4.1
only stimulant	26	26	4.1
free dr	26	24	4.1
cheap lamisil	26	1	4.1
sister drug	26	25	4.1
stimulant type	26	26	4.1
sleep problem	27	27	4.1
day buy	26	24	4.1
delivery purchase	26	26	4.1
prescription coverage	27	24	4.1
delivery next day	26	23	4.1
other med	26	26	4.1

KeyTerm	Freq	No Posts	Score
third shift	28	23	4.1
group buy	28	23	4.1
online uk	29	23	4.1
good stuff	104	100	4.1
short half	27	26	4.1
low dopamine	26	22	4.1
order fioricet	26	3	4.1
same issue	68	68	4.1
sleep last night	27	27	4.1
drug seeker	26	24	4.1
yerba mate	28	27	4.1
chemical structure	37	34	4.1
buy diethylpropion	27	2	4.1
first order	64	59	4.1
executive functioning	29	26	4.1
full time job	38	38	4.1
non stop	36	35	4.1
main problem	69	69	4.1
time sleeping	27	27	4.1
nicotine patch	27	25	4.1
bad headache	27	27	4.1
third day	75	74	4.1
coupon code	52	48	4.1
rapid heart	28	28	4.1
same side	51	50	4.1
chronic pain	110	95	4
antidepressant effect	26	25	4
first pill	26	26	4
other thing	106	106	4

KeyTerm	Freq	No Posts	Score	KeyTerm	Freq	No Posts	Score
clear head	28	27	4	prescribed dose	26	26	4
back pain	113	106	4	power nap	26	25	4
hypertensive crisis	26	25	4	reuptake inhibition	25	25	4
delivery cheap provigil	25	14	4	primary doctor	26	25	4
nuvigil saturday delivery	25	16	4	mild form	28	28	4
buy frisium	25	1	4	sleep inertia	25	21	4
trying provigil	25	25	4	patient assistance program	25	24	4
day provigil	25	18	4	taking melatonin	25	25	4
traditional stimulant	25	24	4	work week	42	41	4
medical advice	52	52	4	purchase discount	25	25	4
narcolepsy medication	25	24	4	long term solution	29	29	4
coffee drinker	27	27	4	healthy diet	60	59	4
full time	267	253	4	trying something	35	35	4
next day buy	25	23	4	western union	30	23	4
prescription discount	26	25	4	different brand	30	29	4
name provigil	25	25	4	caffeine withdrawal	25	22	3.9
water intake	35	33	4	negative effect	57	53	3.9
skin reaction	27	27	4	new script	29	29	3.9
same dosage	26	25	4	next visit	39	38	3.9
high blood	128	121	4	good amount	60	59	3.9
online cheap buy	25	24	4	weird feeling	27	24	3.9
online canadian	25	25	4	low motivation	25	23	3.9
good med	25	25	4	stomach pain	33	28	3.9
different effect	29	28	4	pre workout	25	25	3.9
jittery feeling	25	24	4	brain damage	57	47	3.9
different doctor	26	26	4	black coffee	30	29	3.9
bad day	54	53	4	next day nuvigil	24	13	3.9
major depressive disorder	34	34	4	nuvigil overnight nuvigil	24	14	3.9
muscle tension	33	30	4	day delivery nuvigil	24	11	3.9

KeyTerm	Freq	No Posts	Score
provigil cheap provigil	24	14	3.9
prescription order provigil	24	15	3.9
limitless drug	24	24	3.9
mild narcolepsy	24	24	3.9
work shift disorder	24	23	3.9
cheap saturday	24	23	3.9
bulletproof diet	24	19	3.9
weak dopamine	24	22	3.9

KeyTerm	Freq	No Posts	Score
speedy feeling	24	24	3.9
report free credit	24	2	3.9
delivery cash	24	23	3.9
weight loss	297	255	3.9
cognitive boost	24	24	3.9
fedex overnight delivery	24	24	3.9
adhd diagnosis	25	24	3.9
recommended dosage	29	29	3.9

Appendix S P3 Top 100 KeyNgrams and Associated Theme

Key Ngram	Freq	NoDocs	Score	Theme
in the morning	3016	2627	252.7	Dosage
i have been	2983	2646	185.2	
during the day	1649	1476	169.3	
i was on	1354	1266	148.6	OtherInt
i have a	2217	2104	133.1	
without a prescription	1108	313	132.4	Acquisition
i take it	1105	1050	129.9	Dosage
i have to	1869	1712	127.1	
i feel like	1180	1089	119.8	
the side effects	976	903	114.4	SideEffects
i need to	1406	1316	113.6	
to take it	1007	958	109.7	
i used to	1133	1055	106.4	SymptomImpact
i had to	1394	1290	104.6	
to stay awake	833	781	100.9	
i am on	857	828	99.3	
feel like i	902	844	98.3	Effect
i had a	1220	1162	97.2	
and i have	1247	1206	96.8	
but i do	1174	1145	94.8	
i think i	1183	1135	94.2	
when i was	1257	1191	93.5	

Key Ngram	Freq	NoDocs	Score	Theme
i am not	1339	1241	93.5	
but i have	1020	998	93.1	
put me on	779	728	93.1	Acquisition
i have had	943	886	92.8	
i have tried	822	785	91.3	OtherInt
cash on delivery	739	187	89.9	Acquisition
i took it	760	717	89.9	Dosage
i want to	1585	1443	89.4	
have been on	783	745	88.3	
for a few	1052	1008	87.3	
i was diagnosed	718	677	85.9	Reason
to help with	806	779	85.9	Reason
a few days	995	949	85.1	
all the time	1022	955	83.6	
to see if	1026	983	83	
a lot of	3356	3019	80.1	
and i am	979	937	79.8	
i think it	1141	1109	77.8	
thanks for the	841	829	77.7	
to help me	691	658	76.7	
i have no	891	860	76.6	
not sure if	746	724	76.5	
i was taking	633	611	75.8	
for a while	911	880	75.7	
i take provigil	617	612	75.6	

Key Ngram	Freq	NoDocs	Score	Theme
the morning and	656	633	75.4	
next day delivery	619	198	74.7	Acquisition
in the morning and	643	621	74.6	
and i do	829	809	73.8	
that i have	931	890	73.4	
a sleep study	600	552	73.4	Investigations
i have been on	608	579	72.9	OtherInt
i can get	684	675	72.7	
seems to be	1045	1008	72.3	Belief
but i am	753	727	72.2	
i do not	992	887	71.3	
have been taking	599	584	71.1	
i wake up	589	555	70.5	
it seems to	735	722	69.5	Belief
i also have	629	608	69.4	
if i take	569	532	68.9	
and i was	812	761	68.6	
as far as	962	927	67.5	
twice a day	580	508	67.5	Dosage
when i take	547	513	66.4	
the only thing	702	685	66.2	
a prescription for	545	526	65.7	Acquisition
i have taken	549	534	64.5	
work for me	546	531	64.5	Belief
have you tried	537	524	64.1	

Key Ngram	Freq	NoDocs	Score	Theme
for me to	718	693	63.7	
a couple of	1259	1179	63.6	
i started taking	521	504	63.3	
have to take	579	560	62.7	
i also take	511	503	62.4	OtherInt
in the afternoon	567	539	61.1	Reason
was diagnosed with	529	514	60.9	Reason
i can not	735	675	60.7	
the next day	640	593	60.5	Effect
it for a	552	545	60.4	
to get a	893	845	60	
i take modafinil	479	473	58.9	
i feel like i	514	494	58.8	Belief
i have been taking	483	472	58.6	
i have not	602	589	58.4	
it makes me	507	493	58.2	Belief
it made me	494	483	57.7	Belief
modafinil is a	469	462	57.6	
to try it	495	487	57.3	
if i do	524	506	57.1	
i am going	564	540	57.1	
i am a	727	705	56.9	
the effects of	669	634	56.7	
in the past	967	926	56.4	
so i can	554	539	56.4	

Key Ngram	Freq	NoDocs	Score	Theme
i take nuvigil	458	456	56.3	
you can get	690	659	56.3	
provigil and nuvigil	456	446	56.1	

Appendix T P3 Top 1000 ngrams

Key Ngram	Freq	NoDocs	Score
in the morning	3016	2627	252.7
i have been	2983	2646	185.2
during the day	1649	1476	169.3
i was on	1354	1266	148.6
i have a	2217	2104	133.1
without a prescription	1108	313	132.4
i take it	1105	1050	129.9
i have to	1869	1712	127.1
i feel like	1180	1089	119.8
the side effects	976	903	114.4
i need to	1406	1316	113.6
to take it	1007	958	109.7
i used to	1133	1055	106.4
i had to	1394	1290	104.6
to stay awake	833	781	100.9
i am on	857	828	99.3
feel like i	902	844	98.3
i had a	1220	1162	97.2
and i have	1247	1206	96.8
but i do	1174	1145	94.8
i think i	1183	1135	94.2
when i was	1257	1191	93.5
i am not	1339	1241	93.5
but i have	1020	998	93.1
put me on	779	728	93.1
i have had	943	886	92.8
i have tried	822	785	91.3
cash on delivery	739	187	89.9
i took it	760	717	89.9
i want to	1585	1443	89.4
have been on	783	745	88.3
for a few	1052	1008	87.3
i was diagnosed	718	677	85.9
to help with	806	779	85.9
a few days	995	949	85.1
all the time	1022	955	83.6
to see if	1026	983	83
a lot of	3356	3019	80.1
and i am	979	937	79.8
i think it	1141	1109	77.8
thanks for the	841	829	77.7
to help me	691	658	76.7
i have no	891	860	76.6
not sure if	746	724	76.5

Key Ngram	Freq	NoDocs	Score
i was taking	633	611	75.8
for a while	911	880	75.7
i take provigil	617	612	75.6
the morning and	656	633	75.4
next day delivery	619	198	74.7
in the morning and	643	621	74.6
and i do	829	809	73.8
that i have	931	890	73.4
a sleep study	600	552	73.4
i have been on	608	579	72.9
i can get	684	675	72.7
seems to be	1045	1008	72.3
but i am	753	727	72.2
i do not	992	887	71.3
have been taking	599	584	71.1
i wake up	589	555	70.5
it seems to	735	722	69.5
i also have	629	608	69.4
if i take	569	532	68.9
and i was	812	761	68.6
as far as	962	927	67.5
twice a day	580	508	67.5

Key Ngram	Freq	NoDocs	Score
when i take	547	513	66.4
the only thing	702	685	66.2
a prescription for	545	526	65.7
i have taken	549	534	64.5
work for me	546	531	64.5
have you tried	537	524	64.1
for me to	718	693	63.7
a couple of	1259	1179	63.6
i started taking	521	504	63.3
have to take	579	560	62.7
i also take	511	503	62.4
in the afternoon	567	539	61.1
was diagnosed with	529	514	60.9
i can not	735	675	60.7
the next day	640	593	60.5
it for a	552	545	60.4
to get a	893	845	60
i take modafinil	479	473	58.9
i feel like i	514	494	58.8
i have been taking	483	472	58.6
i have not	602	589	58.4

Key Ngram	Freq	NoDocs	Score
it makes me	507	493	58.2
it made me	494	483	57.7
modafinil is a	469	462	57.6
to try it	495	487	57.3
if i do	524	506	57.1
i am going	564	540	57.1
i am a	727	705	56.9
the effects of	669	634	56.7
in the past	967	926	56.4
so i can	554	539	56.4
i take nuvigil	458	456	56.3
you can get	690	659	56.3
provigil and nuvigil	456	446	56.1
seem to be	679	660	56
i know i	545	531	55.8
if you have	1385	1306	55.6
i know that	611	581	55.5
from what i	513	499	55.4
throughout the day	497	469	55.4
in the day	527	506	55
able to get	562	541	54.5
online without prescription	443	182	54.4
and it was	772	757	54.2

Key Ngram	Freq	NoDocs	Score
out of bed	464	436	54.2
made me feel	462	444	54.1
i still have	497	491	54
of the day	711	684	53.8
hours of sleep	443	405	53.7
i went to	577	556	53.6
a long time	722	708	53.6
be able to	1863	1733	53.4
i did not	606	557	53.4
makes me feel	449	433	52.8
but i was	523	513	52.7
that i am	599	565	52.6
i know it	499	486	52.6
well for me	434	425	52.5
works for you	439	425	52.3
provigil or nuvigil	424	414	52.2
am going to	506	484	52.1
a few weeks	543	525	52.1
to take a	652	632	52
if you can	638	605	51.9
if you do	910	863	51.9
go back to	555	531	51.6
to fall asleep	423	395	51.3
to go to	694	677	50.9

Key Ngram	Freq	NoDocs	Score
no prescription required	414	168	50.8
no prescription overnight	412	170	50.8
i have never	506	489	50.8
i wanted to	656	632	50.6
i was diagnosed with	416	403	50.5
and i can	495	489	50.5
and i feel	439	430	50.3
i would be	521	506	50.2
through the day	416	405	50.1
so i do	460	454	50.1
a few hours	480	463	50
keep me awake	406	396	50
talk to your	436	423	49.9
you take it	412	392	49.8
i was just	467	452	49.5
that i was	557	519	49.2
works for me	408	398	49.2
in the us	623	575	49.2
i take a	410	402	49.1
for a couple	466	451	49
i have found	458	440	48.8

Key Ngram	Freq	NoDocs	Score
i am going to	468	448	48.6
but it does	488	485	48.3
with no prescription	390	184	47.9
it would be	939	908	47.8
you have to	858	801	47.8
i do have	428	415	47.7
but i think	536	531	47.5
help with the	414	402	47.5
i was wondering	419	414	47.4
last edited by	433	433	47.4
if i can	449	440	47.3
to get it	476	461	47.2
a few months	483	476	47.1
so i have	427	420	47.1
i am taking	392	380	47
the first time	911	858	46.9
i will be	532	506	46.9
times a week	403	390	46.8
go to sleep	391	376	46.7
but it is	722	711	46.6
i was prescribed	378	373	46.6
to keep me	388	380	46.4
sleep at night	380	367	46.2

Key Ngram	Freq	NoDocs	Score
i find that	403	388	46.1
was able to	610	579	46
a low dose	374	359	46
i hope you	470	458	45.9
no side effects	373	366	45.6
that i can	474	460	45.4
online no prescription	372	182	45.4
i felt like	396	377	45.2
a bit of	684	652	45.2
if i have	412	399	45.2
even though i	427	417	45.1
i am still	407	393	45
a day and	384	377	44.8
and see if	414	413	44.8
the same thing	493	483	44.6
it might be	487	473	44.5
i find it	407	400	44.4
if you are	1271	1174	44.3
anyone have any	369	361	44.3
gave me a	395	389	44.3
i guess i	416	404	44
hope this helps	371	371	43.9
if you take	387	371	43.9
you need to	941	876	43.8

Key Ngram	Freq	NoDocs	Score
to your doctor	369	359	43.8
in my experience	377	365	43.8
of the time	541	501	43.6
give it a	413	404	43.6
i was able	431	418	43.5
i was able to	430	417	43.5
that i do	417	412	43.2
days a week	393	368	43
i took a	379	368	43
as long as	676	662	42.7
excessive daytime sleepiness	345	334	42.6
because i have	381	380	42.6
cup of coffee	361	347	42.5
for me and	376	376	42.5
see if it	373	364	42.4
was wondering if	364	355	42.4
so i am	380	373	42.4
is that i	400	389	42.2
thank you for	604	588	42.2
want to try	367	359	42.1
wondering if anyone	347	339	42.1

Key Ngram	Freq	NoDocs	Score
for about a	361	354	42.1
for the first	750	718	42
no prescription buy	340	158	42
when i have	367	360	42
for a week	375	360	42
if it is	558	528	41.9
best of luck	354	353	41.8
your doctor about	347	338	41.8
i got a	396	379	41.8
at a time	537	499	41.8
side effects of	359	347	41.7
i try to	408	389	41.7
trying to get	427	407	41.4
i decided to	445	410	41.4
i can take	346	341	41.4
pay for it	351	337	41.3
when i do	353	343	41.3
it could be	473	454	41.3
because i was	377	363	41.2
do you think	494	463	41.2
let me know	431	421	41
it is a	1055	1010	40.8
modafinil in the	329	322	40.7

Key Ngram	Freq	NoDocs	Score
when i first	377	370	40.7
but i did	389	386	40.7
i have read	364	350	40.5
been on it	327	323	40.4
diagnosed with narcolepsy	325	321	40.3
the day and	358	354	40.2
but if you	473	465	40.2
you have a	663	650	40.2
i would like	525	493	40.2
in my life	396	391	40.1
but i would	377	376	40.1
taking it for	324	319	40
that i had	418	394	39.9
used to take	329	324	39.9
if anyone has	339	330	39.9
times a day	350	315	39.7
a few years	470	463	39.6
and it has	389	385	39.6
do you have	438	425	39.5
to wake up	337	313	39.5
and have been	380	372	39.3
i am also	356	351	39.3
i am so	375	358	39.3
felt like i	334	317	39.3

Key Ngram	Freq	NoDocs	Score	Key Ngram	Freq	NoDocs	Score
to the point	427	413	39.2	on it for	318	315	38.5
i thought i	395	384	39.2	going to try	332	317	38.5
it seems like	368	356	39.2	but i still	335	334	38.2
but when i	358	350	39.1	being able to	451	429	38.2
thanks for your	363	357	39.1	i was in	407	393	38.1
need to take	361	356	39.1	go to bed	321	295	38.1
for a few days	340	323	39.1	like i was	327	317	38.1
i woke up	332	312	38.9	a few times	357	351	38.1
i have the	384	382	38.9	i just started	312	310	38
at the moment	463	448	38.9	a script for	308	302	38
i would have	440	424	38.9	i had the	375	372	37.9
for a long	440	435	38.9	does anyone have	317	313	37.8
it sounds like	350	343	38.9	when i am	333	313	37.8
i think that	506	490	38.8	get out of	382	365	37.7
want to be	497	476	38.8	it in the	439	423	37.6
and i had	380	366	38.7	in my case	325	309	37.6
after a few	361	353	38.6	i would say	356	345	37.5
been able to	457	440	38.6	an empty stomach	306	287	37.5
to get my	343	337	38.6	it works for	310	305	37.5
my insurance company	312	294	38.6	i stopped taking	303	296	37.5
only thing that	341	336	38.6	going to be	680	639	37.5
i think you	398	386	38.5	and it did	318	314	37.3
for the past	417	401	38.5	most of the	807	781	37.3
to have a	721	703	38.5				

Key Ngram	Freq	NoDocs	Score
hours a day	341	322	37.3
the only thing that	328	324	37.2
now i am	323	315	37.2
i think the	457	454	37.1
i have an	336	333	37
a little bit	452	431	37
worked for me	305	298	36.9
on an empty	302	283	36.9
just wanted to	350	340	36.8
and it works	310	306	36.8
due to the	648	611	36.8
a side effect	301	285	36.7
work for you	317	305	36.6
and then i	346	335	36.6
i am now	325	314	36.5
lot of people	373	369	36.5
take a nap	296	280	36.5
after taking it	294	286	36.4
the same time	695	654	36.3
on an empty stomach	295	277	36.3
the time i	338	328	36.2
but i can	334	330	36.1
a lot of people	367	363	36.1
i am in	336	328	36.1

Key Ngram	Freq	NoDocs	Score
i am currently	314	306	36.1
do you take	294	279	36.1
i only take	291	284	36.1
i have heard	320	315	36
to deal with	483	469	35.9
i would like to	444	423	35.9
it is not	759	719	35.8
if i am	322	302	35.8
a good idea	384	376	35.8
used to be	384	368	35.8
it has been	538	518	35.7
in a row	342	320	35.7
but if i	311	307	35.7
how to get	365	246	35.7
of the drug	304	287	35.6
just want to	346	340	35.6
make me feel	295	284	35.6
and it does	321	319	35.6
i took provigil	286	284	35.6
i get a	318	311	35.5
side effects and	291	284	35.5
been taking it	286	281	35.5
to get the	546	534	35.4
in the uk	514	454	35.4

Key Ngram	Freq	NoDocs	Score
buy modafinil online	284	84	35.3
if i had	344	341	35.3
at the same	698	660	35.3
where to buy	304	199	35.2
to get up	308	285	35.2
on provigil for	283	282	35.2
used to treat	307	283	35.1
to be able to	510	494	35.1
to be able	512	496	35.1
supposed to be	380	359	35.1
so i was	316	310	35
when i get	307	300	35
i take the	289	275	35
be a good	370	362	34.9
in the am	282	257	34.9
at the same time	629	594	34.8
i was so	317	304	34.8
i just want	315	307	34.8
but it was	445	441	34.8
is supposed to	334	324	34.8
if you want	622	582	34.8
does anyone know	298	290	34.7
because of the	576	561	34.7

Key Ngram	Freq	NoDocs	Score
awake during the	279	269	34.7
i wonder if	325	318	34.6
back to sleep	284	259	34.6
i was a	385	376	34.5
if i was	315	307	34.5
to do with	535	514	34.5
at this point	419	395	34.4
not going to	413	396	34.4
and it is	567	547	34.4
not sure what	315	307	34.4
when i started	298	288	34.4
as much as	494	478	34.4
for a long time	362	359	34.3
as i have	336	330	34.3
i can tell	318	308	34.3
what do you	403	401	34.2
i tend to	301	293	34.2
in terms of	629	590	34.1
a lot more	384	375	34
would like to	687	637	33.9
a bit more	359	342	33.9
for the last	371	360	33.9
off of it	276	256	33.9
have the same	349	346	33.9

Key Ngram	Freq	NoDocs	Score
i seem to	291	289	33.8
effects of modafinil	271	259	33.7
which is a	422	415	33.7
and now i	299	297	33.6
out of my	316	308	33.6
for me it	280	275	33.4
of my life	323	316	33.4
in the evening	316	299	33.4
a week and	284	282	33.4
for some reason	330	324	33.3
with the fatigue	267	263	33.2
up in the	411	403	33.2
i was wondering if	280	277	33.2
me and i	299	298	33.2
i used to take	267	265	33.1
that i could	324	319	32.9
you want to	787	712	32.9
i have used	286	280	32.9
and if i	292	287	32.8
now that i	294	288	32.7
i have some	287	286	32.7

Key Ngram	Freq	NoDocs	Score
awake during the day	262	252	32.6
i have also	290	285	32.6
and when i	305	300	32.6
would be a	447	437	32.6
for the fatigue	261	259	32.5
i had been	308	295	32.4
and see how	284	282	32.4
when i took	265	254	32.4
to take the	389	380	32.3
i really do	292	285	32.3
and i think	368	363	32.3
to be a	924	895	32.2
for a couple of	290	280	32.1
if i could	294	287	32.1
over the counter	267	191	32
i go to	280	273	32
of modafinil and	256	253	31.9
seem to have	318	315	31.9
to be on	316	305	31.9
to give me	273	266	31.9
on top of	404	389	31.9
have no idea	311	306	31.9
need to be	584	565	31.9

Key Ngram	Freq	NoDocs	Score
i ended up	282	270	31.9
a hard time	286	279	31.9
you could try	262	254	31.9
i wish i	295	288	31.8
is a good	404	398	31.8
it helps me	258	253	31.8
i found that	279	272	31.8
sleep apnea and	255	244	31.7
want to take	288	285	31.7
am not sure	285	278	31.7
stop taking it	255	253	31.7
and i would	316	311	31.6
that it is	578	560	31.6
it may be	413	397	31.6
is that it	369	369	31.6
to find a	390	384	31.6
about a week	266	265	31.5
since i was	279	271	31.5
see how it	272	270	31.5
when i need	258	251	31.5
has anyone else	256	253	31.5
if it was	329	326	31.4
i am very	304	301	31.4
first time i	285	280	31.4

Key Ngram	Freq	NoDocs	Score
side effect of	258	252	31.4
about a year	270	266	31.3
have to be	486	476	31.3
to try and	317	307	31.2
there is a	989	955	31.2
but i ca	281	275	31.1
to talk to	318	309	31
was going to	387	377	31
get out of bed	252	243	31
i can do	291	288	31
is going to	476	457	30.9
a cup of	271	260	30.9
it every day	252	240	30.9
it was a	644	628	30.9
provigil and it	247	246	30.8
to try to	347	332	30.8
long as i	265	262	30.8
you have any	351	343	30.8
not sure how	274	271	30.8
have a lot	311	306	30.8
is not a	514	507	30.8
it feels like	268	257	30.8
i just do	273	267	30.7
far as i	297	291	30.7
i was told	274	261	30.7

Key Ngram	Freq	NoDocs	Score
what to do	345	326	30.7
i would not	287	282	30.6
it can be	492	475	30.6
in the brain	269	250	30.6
seems to help	247	238	30.6
have sleep apnea	245	235	30.6
as far as i	290	284	30.5
i was going	291	286	30.4
nothing for me	244	240	30.3
to go back	297	289	30.3
had to stop	248	245	30.2
i was thinking	270	263	30.1
might want to	289	276	30.1
for a month	254	250	30
was on provigil	240	240	30
and i did	278	271	30
i had no	280	277	30
to sleep at	242	237	29.9
most of the time	287	276	29.9
at night and	248	241	29.9
with your doctor	250	238	29.9
more of a	301	296	29.8
days in a	244	232	29.8

Key Ngram	Freq	NoDocs	Score
in my opinion	301	293	29.7
this is my	281	279	29.7
want to do	313	304	29.7
i use it	248	235	29.7
that if i	259	256	29.7
and it seems	258	257	29.7
is the only	370	366	29.7
that would be	349	342	29.6
as i am	274	263	29.6
the same as	359	353	29.6
if this is	301	296	29.5
sounds like you	246	242	29.5
nuvigil and provigil	236	231	29.5
me on nuvigil	236	235	29.5
the first day	280	271	29.4
stopped taking it	235	231	29.4
a week or	252	251	29.3
ask your doctor	239	235	29.3
but now i	251	248	29.2
to see if it	247	241	29.2
talk to your doctor	239	232	29.2
feel like a	250	247	29.1

Key Ngram	Freq	NoDocs	Score
know if it	249	244	29.1
is what i	273	266	29.1
days in a row	235	223	29.1
i am trying	256	250	29.1
i have noticed	242	235	29.1
to see how	320	318	29
to pay for	308	294	29
online without a	232	119	29
couple of weeks	264	255	29
no prescription needed	232	154	29
even if i	251	247	28.9
right now i	243	236	28.9
as it is	396	391	28.9
keep you awake	231	229	28.8
to get through	248	242	28.8
to get me	241	233	28.8
need to get	283	275	28.8
over a year	253	248	28.8
but i feel	241	237	28.7
in the mornings	234	223	28.7
it does not	354	338	28.7

Key Ngram	Freq	NoDocs	Score
for the first time	428	413	28.7
it to be	306	305	28.6
i need it	233	229	28.6
i just got	243	240	28.6
and i ca	253	248	28.6
i am just	247	237	28.6
i took the	246	231	28.5
when i wake	229	222	28.5
provigil no prescription	227	41	28.4
out of it	270	262	28.4
give me a	252	247	28.3
as i can	266	261	28.3
as long as i	241	238	28.3
any side effects	228	226	28.3
is a stimulant	226	225	28.2
so i would	242	240	28.2
obstructive sleep apnea	227	212	28.2
to figure out	295	290	28.2
i am not sure	251	244	28.2
about a month	236	232	28.1
once a day	233	216	28.1
couple of days	252	244	28

Key Ngram	Freq	NoDocs	Score	Key Ngram	Freq	NoDocs	Score
it gives me	230	226	27.9	it when i	229	227	27.5
to kick in	225	217	27.9	told me that	260	246	27.5
to my doctor	224	221	27.9	it seems to be	237	237	27.4
it and i	237	236	27.9	go to the	353	343	27.4
want to go	269	264	27.9	i could not	263	250	27.4
never heard of	242	236	27.8	how to buy	223	120	27.4
but i just	240	237	27.8	sure if it	227	225	27.4
if i were	251	248	27.7	try to get	247	245	27.4
i tried it	226	220	27.7	thanks in advance	229	229	27.4
i will try	239	231	27.7	have to go	268	265	27.3
i have to take	224	219	27.7	buy provigil online	218	55	27.3
a little more	302	294	27.7	keeps me awake	218	217	27.3
to get some	252	250	27.6	what i was	251	245	27.3
tired all the	221	211	27.6	just have to	253	243	27.3
the first few	246	241	27.6	a bunch of	309	297	27.3
i was having	229	221	27.6	has helped me	225	219	27.2
but it did	242	242	27.6	shift work disorder	217	199	27.2
seems to work	229	226	27.6	what you are	271	256	27.2
i have narcolepsy	220	212	27.6	do not have	352	342	27.2
but i know	242	240	27.6	tired all the time	217	207	27.1
and a half	298	282	27.6	that works for	224	221	27.1
because i am	240	235	27.6				
the insurance company	226	200	27.6				
i agree with	262	261	27.5				

Key Ngram	Freq	NoDocs	Score
low dose of	217	212	27.1
because i do	236	233	27.1
you should be	272	267	27.1
online without a prescription	216	114	27.1
have a prescription	216	209	27
see if i	234	232	27
i just want to	236	231	27
any of the	388	381	27
are there any	234	226	27
would love to	267	265	26.9
effect on me	216	212	26.9
to stop taking	217	214	26.9
me feel like	220	216	26.9
it and it	228	227	26.9
quite a bit	255	252	26.9
had the same	231	229	26.8
like i have	220	218	26.8
i could get	229	228	26.8
where can i	226	163	26.7
i do take	214	213	26.7
know what to	250	243	26.7
as soon as	403	388	26.7
you are taking	226	214	26.7
about an hour	228	222	26.7

Key Ngram	Freq	NoDocs	Score
it gave me	218	213	26.7
me on provigil	212	210	26.6
able to sleep	214	211	26.6
so that i	243	239	26.6
changed my life	216	209	26.5
get a prescription	212	208	26.5
and if you	318	315	26.5
i live in	239	236	26.5
and i could	239	238	26.5
my question is	222	220	26.4
is why i	238	236	26.4
nuvigil no prescription	210	18	26.4
and have a	273	269	26.4
told me to	223	213	26.3
soon as i	231	227	26.3
but it seems	231	231	26.3
i suffer from	211	206	26.3
gives me a	216	210	26.3
the rest of	557	533	26.3
to be honest	243	241	26.3
think it is	255	247	26.3
there is no	715	683	26.2
is the best	309	302	26.2

Key Ngram	Freq	NoDocs	Score
and i just	231	227	26.2
want to get	261	257	26.2
has anyone tried	209	206	26.1
cod no prescription	208	110	26.1
that modafinil is	208	206	26.1
i really need	215	212	26.1
one of my	323	317	26.1
to do anything	233	224	26.1
up all night	210	203	26
that i would	256	246	26
are you taking	208	201	25.9
as soon as i	227	223	25.9
modafinil is not	206	203	25.9
to see what	286	282	25.9
when i wake up	207	201	25.9
had to take	218	210	25.8
i was also	220	217	25.8
for a year	229	225	25.8
i tried provigil	205	205	25.8
how it goes	210	206	25.7
hard to get	223	220	25.7

Key Ngram	Freq	NoDocs	Score
one of those	317	313	25.7
to get to	295	283	25.7
you might want	235	227	25.7
on and off	231	225	25.7
if there is	322	313	25.7
some sort of	267	262	25.7
that you are	339	329	25.6
a tolerance to	204	192	25.6
i needed to	231	223	25.6
if it does	232	227	25.6
that being said	221	213	25.6
i can say	229	218	25.6
it does help	205	204	25.6
but i also	220	217	25.6
for me is	212	208	25.5
some kind of	266	257	25.5
any of you	215	210	25.5
my ability to	208	199	25.5
but i had	221	217	25.5
not want to	265	249	25.5
while i was	231	229	25.5
when i had	217	213	25.5
get to sleep	204	192	25.4
i tried to	249	243	25.4

Key Ngram	Freq	NoDocs	Score
by the way	282	276	25.4
take it for	206	205	25.4
a cup of coffee	208	203	25.4
nuvigil or provigil	202	202	25.4
modafinil no prescription	202	35	25.4
it to me	215	210	25.4
i still feel	206	200	25.4
bit of a	259	256	25.3
to use it	250	242	25.3
for some people	211	208	25.3
can i buy	206	138	25.3
get up and	210	204	25.3
was put on	207	192	25.2
you might want to	229	221	25.2
in my head	219	216	25.2
because it is	287	282	25.2
does anyone have any	204	201	25.2
a drug that	202	200	25.1
much for me	202	202	25.1
in your system	202	191	25
i was on provigil	199	199	25

Key Ngram	Freq	NoDocs	Score
on provigil and	199	198	25
modafinil for a	199	197	25
what i have	231	226	25
i would take	204	192	25
my experience with	204	201	25
i was still	211	203	24.9
the first time i	222	217	24.9
if anyone else	202	200	24.9
take it on	200	196	24.9
thing in the	210	207	24.9
up in the morning	203	195	24.9
i got the	222	217	24.8
quality of life	250	242	24.8
able to do	240	232	24.8
used to it	206	197	24.8
for it to	225	222	24.8
and see what	223	218	24.8
and i know	228	225	24.8
to make sure	354	339	24.8
thought i was	211	204	24.7
only thing i	209	209	24.7
me to take	204	200	24.7
to know if	217	214	24.7
it is very	271	268	24.7

Key Ngram	Freq	NoDocs	Score
been on provigil	196	195	24.7
i was going to	227	224	24.7
cups of coffee	197	183	24.6
thought it was	237	233	24.6
the problem is	256	252	24.6
out of pocket	199	188	24.6
has been a	351	348	24.6
been taking modafinil	195	195	24.6
i hope this	210	209	24.6
modafinil is the	195	193	24.6
i have no idea	219	216	24.6
it has a	271	269	24.5
what i am	220	209	24.5
or at least	261	260	24.5
i should be	213	207	24.5
since i have	206	203	24.5
me stay awake	194	190	24.4
side effects from	196	193	24.4
so far i	203	200	24.4
no rx needed	194	102	24.4
i got my	207	203	24.4
much of a	230	228	24.4

Key Ngram	Freq	NoDocs	Score
my blood pressure	196	171	24.4
went back to	218	213	24.4
and it helps	196	194	24.4
i first started	199	197	24.3
know if i	208	204	24.3
just wondering if	197	193	24.3
effects of the	217	212	24.3
i am looking	219	200	24.3
i am having	202	194	24.3
the fact that	576	538	24.3
to make me	202	199	24.3
to be more	261	256	24.3
i hope it	205	202	24.3
make sure you	275	269	24.2
am trying to	211	205	24.2
provigil for a	192	191	24.2
as i was	228	224	24.2
any experience with	194	189	24.2
i felt like i	198	188	24.1
prescription overnight delivery	191	124	24.1
now i have	208	207	24

Key Ngram	Freq	NoDocs	Score
see if that	195	187	24
i found it	209	202	24
a higher dose	191	181	24
if you need	262	253	24
know how it	199	192	24
first thing in	193	189	23.9
everyone is different	191	189	23.9
if you want to	352	337	23.9
i could take	193	191	23.9
which is why	230	229	23.9
for over a	208	204	23.9
i thought it	229	227	23.9
in a few	244	241	23.8
sorry to hear	198	197	23.8
it was the	332	327	23.8
know what you	223	218	23.8
so if you	248	245	23.8
prescription overnight shipping	188	115	23.7
without prescription buy	188	123	23.7
i get the	207	206	23.6
i need a	202	201	23.6

Key Ngram	Freq	NoDocs	Score
to be the	440	434	23.6
make you feel	200	193	23.6
how do you	264	253	23.6
days when i	189	178	23.6
have a lot of	224	222	23.6
the point where	216	209	23.5
that it was	343	332	23.5
in the same	395	387	23.5
been diagnosed with	193	189	23.5
worth a try	189	187	23.5
seems to have	242	234	23.5
like i am	191	187	23.5
stay awake and	186	184	23.4
me to get	197	193	23.4
luck to you	188	187	23.4
i may have	201	198	23.4
first thing in the	188	185	23.4
you can take	219	212	23.3
if i take it	185	180	23.3
started taking it	185	184	23.3
i will have	205	203	23.3
a year ago	219	216	23.3

Key Ngram	Freq	NoDocs	Score
once a week	199	191	23.2
nuvigil and it	184	184	23.2
it helps with	185	183	23.2
dose of modafinil	184	178	23.2
at the time	456	426	23.2
it has helped	187	185	23.2
it was like	208	204	23.2
it for me	192	190	23.2
it seems that	232	227	23.2
is there any	209	206	23.2
some of the	838	804	23.2
have to say	226	223	23.2
i know this	200	197	23.2
i take my	186	179	23.1
so i ca	192	191	23.1
provigil in the	183	177	23.1
i would love	211	210	23.1
and i take	185	185	23.1
i guess it	200	199	23.1
at least i	197	197	23
did nothing for	183	181	23
this is the	559	541	23
modafinil and armodafinil	182	181	23

Key Ngram	Freq	NoDocs	Score
shift work sleep	182	175	23
feel like it	190	186	22.9
because of my	191	188	22.9
to help with the	188	185	22.9
the only one	232	231	22.9
am able to	189	178	22.9
to take modafinil	181	179	22.9
take it every	181	176	22.8
only take it	181	180	22.8
and i still	192	187	22.8
great for me	182	182	22.8
keep in mind	233	228	22.8
chronic fatigue syndrome	183	172	22.8
i can sleep	181	177	22.8
i started to	197	186	22.8
not sure if it	186	185	22.8
it took me	195	195	22.8
and it worked	186	186	22.7
when it comes	339	326	22.7
have found that	198	195	22.7
can get it	188	186	22.7

Key Ngram	Freq	NoDocs	Score
early in the	215	214	22.7
a bit of a	226	224	22.7
give you a	228	226	22.7
to your doctor about	181	176	22.6
going to have	231	230	22.6
to try modafinil	179	177	22.6
wish i could	194	192	22.6
and i felt	188	184	22.6
might be a	218	215	22.6
if i did	195	194	22.5
it a try	188	185	22.5
of side effects	181	178	22.5
could be a	230	225	22.5
in my system	179	166	22.5
to take provigil	178	175	22.5
really want to	206	205	22.5
when i got	199	199	22.5
the reason i	191	191	22.5
when i first started	182	180	22.5
be able to get	198	195	22.4
even if you	241	238	22.4
find something that	180	179	22.4

Key Ngram	Freq	NoDocs	Score
most of my	200	197	22.4
when it comes to	324	311	22.4
buy nuvigil online	177	22	22.4
to prescribe it	177	170	22.4
get through the	183	179	22.3
the only thing i	187	187	22.3
a couple of weeks	194	186	22.3
to give it	200	198	22.3
the effects are	178	178	22.3
i took my	182	170	22.3
an hour or	187	183	22.3
been on nuvigil	176	173	22.3
modafinil and it	176	175	22.3
it wears off	176	172	22.2
i would suggest	188	183	22.2
wake up and	181	179	22.2
to get out	216	212	22.2
a huge difference	183	180	22.2
it comes to	326	313	22.2
end of the day	201	196	22.2

Key Ngram	Freq	NoDocs	Score
thing in the morning	178	176	22.2
i can only	198	196	22.2
lack of sleep	178	167	22.2
better for me	177	176	22.2
this is a	642	627	22.2
i was doing	190	185	22.2
hope you find	179	179	22.1
to keep me awake	175	172	22.1
but it has	201	200	22.1
as i know	191	189	22.1
of the side	180	178	22.1
a day or	188	184	22.1
not being able	189	187	22.1
it is the	401	390	22.1
is a very	283	281	22.1
there are a	304	296	22.1
the same effect	181	178	22.1
there are other	206	203	22
i used to be	180	175	22
i took modafinil	174	172	22
get used to	189	183	22
the sleep study	174	150	22

Key Ngram	Freq	NoDocs	Score
not being able to	188	186	22
may want to	214	211	22
that you have	276	266	22
first thing in the morning	176	174	22
to go through	205	200	22
that is a	234	230	21.9
period of time	228	217	21.9
to worry about	213	210	21.9
to go back to	194	187	21.9
i am trying to	189	185	21.9
went to the	225	222	21.9
provigil for fatigue	173	170	21.9
of the brain	194	163	21.9
to get off	182	168	21.9
had a sleep	173	170	21.9
work sleep disorder	173	166	21.9
in this forum	178	145	21.9
even when i	179	178	21.8
i dont know	182	177	21.8
not work for	177	176	21.8
is there a	212	204	21.8

Key Ngram	Freq	NoDocs	Score
a sleep disorder	172	166	21.7
let you know	198	192	21.7
a couple of days	187	183	21.7
get rid of	223	216	21.7

Key Ngram	Freq	NoDocs	Score
i found out	187	181	21.7
to help you	285	276	21.7
to have to	212	211	21.7
i just wanted	190	186	21.7
i know what	189	186	21.7
twice a week	178	175	21.7

Appendix U Word sketch of 'feel'

Gramrel	Collocate	Freq
modifiers of feel		6842
	still	589
	awake	227
	never	242
	absolutely	46
	kinda	36
	alone	25
	constantly	22
	ill	20
	wide	14
	emotionally	13
	sorry	12
	fully	12
	strongly	9
	tight	7
	afterwards	7
	overly	7
	guilty	6
	close	6
objects of feel		4354
	bit	252
	effect	477
	way	220
	need	134
	nothing	167
	fine	101
	difference	91
	pain	64
	alert	54
	today	61
	kind	54
	urge	24
	energy	31
	emotion	22
	day	46
	alot	21
	crash	19
	sense	21
	spacey	17
	driving	17

Gramrel	Collocate	Freq
	euphoria	15
	crappy	14
	sensation	12
	brain	13
	buzz	11
	rush	11
	hyper	11
	opposite	10
	pressure	10
	pleasure	9
	hunger	8
	exhaustion	7
	crap	7
	etc.	7
	affect	7
	while	7
	discomfort	6
	decrease	6
	light	6
	level	7
	doctor	6
subjects of feel		2032
	i	634
	dont	37
	body	45
	day	38
	head	22
	eye	22
	people	72
	heart	17
	Modalert	16
	morning	13
	stomach	11
	everything	13
	life	13
	eyelid	8
	chest	8
	user	8
	question	8
	im	8
	throat	6

Gramrel	Collocate	Freq
	havent	6
feel and/or ...		150
	yawn	8
	think	9
	look	6
	sleep	10
prepositional phrases		2163
	"%w" like ...	1590
	"%w" in ...	98
	"%w" on ...	69
	"%w" of ...	47
	"%w" for ...	43
	"%w" if ...	34
	"%w" about ...	33
	"%w" with ...	31
	"%w" at ...	25
	"%w" after ...	23
	"%w" though ...	22
	"%w" to ...	21
	"%w" from ...	21
	"%w" during ...	20
	"%w" as ...	18
	"%w" before ...	8
	"%w" over ...	6
particles after feel		74
	down	13
	off	15
	out	21
	up	23
particles after feel with object		19
	out	6
	up	7
pronominal objects of feel		689
	myself	35
	it	540
	them	24

Gramrel	Collocate	Freq
	you	74
pronominal subjects of feel		12026
	me	1467
	I	7620
	you	1641
	it	839
	he	129
	they	129
	she	61
	we	39
	him	25
	them	24
	myself	14
	one	14
	us	11
	her	9
wh-words following feel		179
	when	125
	that	14
	which	6
	what	20
	how	8
infinitive objects of feel		37
	be	13
-ing objects of feel		81
	amaze	7
	fuck	6
	take	12
complement s of feel		289
	different	19
	weird	7
	full	6
	alert	7
adjectives after feel		5342
	tired	492
	good	782
	free	223
	normal	170
	awake	100

Gramrel	Collocate	Freq
	comfortable	81
	different	93
	weird	70
	terrible	53
	guilty	51
	amazing	49
	confident	30
	happy	32
	shitty	27
	alive	22
	heavy	21
	fine	23
	hopeless	17
	sluggish	17
	euphoric	16
	full	16
	cold	15
	helpless	12
	suicidal	12
	human	11
	dirty	11
	stable	10
	flat	9
	itchy	8
	ready	8
	fresh	7
	natural	7
	super	7
	slow	7
	most	7
	fake	6
	everyday	6
	useless	6
feel like ...		1590
	shit	101
	i	99
	zombie	68
	person	68
	self	23
	stimulant	18
	heart	15
	im	14
	superman	11
	head	11
	human	11

Gramrel	Collocate	Freq
	time	11
	death	10
	something	10
	effect	9
	idiot	8
	kid	7
	version	7
	attack	6
feel in ...		98
	control	7
	college	7
	year	12
	life	7
	morning	7
feel on ...		69
	top	7
feel of ...		47
	anything	6
feel for ...		43
	day	6
feel during ...		20
	day	16
usage patterns		0
	as reflexive	35
	in passive	17

Appendix V P1 Exploratory study – journal paper copy

frontiers in Digital Health

ORIGINAL RESEARCH
published: 17 February 2021
doi: 10.3389/fdgth.2021.598431

Spontaneously Generated Online Patient Experience of Modafinil: A Qualitative and NLP Analysis

Julia Walsh[1*], *Jonathan Cave*[2] *and Frances Griffiths*[1]

[1] *Warwick Medical School, University of Warwick, Coventry, United Kingdom,* [2] *Department of Economics, University of Warwick, Coventry, United Kingdom*

Objective: To compare the findings from a qualitative and a natural language processing (NLP) based analysis of online patient experience posts on patient experience of the effectiveness and impact of the drug Modafinil.

Methods: Posts ($n = 260$) from 5 online social media platforms where posts were publicly available formed the dataset/corpus. Three platforms asked posters to give a numerical rating of Modafinil. Thematic analysis: data was coded and themes generated. Data were categorized into PreModafinil, Acquisition, Dosage, and PostModafinil and compared to identify each poster's own view of whether taking Modafinil was linked to an identifiable outcome. We classified this as positive, mixed, negative, or neutral and compared this with numerical ratings. NLP: Corpus text was speech tagged and keywords and key terms extracted. We identified the following entities: drug names, condition names, symptoms, actions, and side-effects. We searched for simple relationships, collocations, and co-occurrences of entities. To identify causal text, we split the corpus into PreModafinil and PostModafinil and used n-gram analysis. To evaluate sentiment, we calculated the polarity of each post between −1 (negative) and +1 (positive). NLP results were mapped to qualitative results.

Results: Posters had used Modafinil for 33 different primary conditions. Eight themes were identified: the reason for taking (condition or symptom), impact of symptoms, acquisition, dosage, side effects, other interventions tried or compared to, effectiveness of Modafinil, and quality of life outcomes. Posters reported perceived effectiveness as follows: 68% positive, 12% mixed, 18% negative. Our classification was consistent with poster ratings. Of the most frequent 100 keywords/keyterms identified by term extraction 88/100 keywords and 84/100 keyterms mapped directly to the eight themes. Seven keyterms indicated negation and temporal states. Sentiment was as follows 72% positive sentiment 4% neutral 24% negative. Matching of sentiment between the qualitative and NLP methods was accurate in 64.2% of posts. If we allow for one category difference matching was accurate in 85% of posts.

Conclusions: User generated patient experience is a rich resource for evaluating real world effectiveness, understanding patient perspectives, and identifying research gaps.

Edited by:
Goran Nenadic,
The University of Manchester,
United Kingdom

Reviewed by:
Elizabeth Ford,
Brighton and Sussex Medical School,
United Kingdom
Elvira Perez Vallejos,
University of Nottingham,
United Kingdom

***Correspondence:**
Julia Walsh
julia.walsh@warwick.ac.uk

Specialty section:
This article was submitted to Health Informatics, a section of the journal Frontiers in Digital Health

Received: *24 August 2020*
Accepted: *27 January 2021*
Published: *17 February 2021*

Citation:
Walsh J, Cave J and Griffiths F (2021) Spontaneously Generated Online Patient Experience of Modafinil: A Qualitative and NLP Analysis. Front. Digit. Health 3:598431. doi: 10.3389/fdgth.2021.598431

Both methods successfully identified the entities and topics contained in the posts. In contrast to current evidence, posters with a wide range of other conditions found Modafinil effective. Perceived causality and effectiveness were identified by both methods demonstrating the potential to augment existing knowledge.

Keywords: social media, natural language processing, effectiveness, causality, patient experience, evidence-based medicine, sentiment analysis, qualitative/NLP comparison

INTRODUCTION

Increasing numbers of people use social media and other online spaces as either a first or second line health information (1) and exchange resource (2, 3) with estimates suggesting the volume of online health related data will have grown by 300% between 2017 and 2020 (4). This unstructured freeform textual data contains a mass of contextually grounded detail about the perceptions and health concerns of those who post online. It has potential to add to clinical understanding, either by adding to knowledge where existing evidence is inconclusive (5), or in aiding understanding of real-world usage (6), although the methods for analyzing it are still at an early stage of development (7–13).

Although evidence based medicine (EBM) has been instrumental in raising healthcare standards and developing clinical knowledge, it has acknowledged weaknesses (14–16), including a divide between patient priorities and the research agenda (15–20) and a structural reliance on evidence from RCTs and systematic reviews (17, 18, 20, 21). Spontaneously generated online patient experience (SGOPE) is a data resource which could help address these weaknesses. However, the lack of established methodologies to analyze it inhibits its use (22–25). Natural language processing (NLP) refers to the use of computational techniques and algorithms that aim to interpret the semantic meaning from large volumes of unstructured text (26). A rapidly developing area (27), it is being used to explore health related social media usage (28–32), detecting drug or device related adverse events from user generated content (33, 34), generating new understanding about treatment switching and adherence behavior (35, 36) and as a surveillance tool for infectious disease outbreaks (37, 38) and suicide risk (12) although little work has been carried out into its use for assessing effectiveness (35).

This study was undertaken in preparation for a larger study of SGOPE data on Modafinil using NLP. Our aim was to understand the data in depth in order to develop relevant NPL analysis for the subsequent study.

Study objectives were to

- Qualitatively explore context, health conditions, and symptoms where Modafinil is used, its perceived effectiveness and impact, and identify indications of causation of effect and outcomes.
- Use NLP and corpus linguistics to identify topics, create an ontology of entities, relationships, and causal text, and evaluate overall sentiment toward perceived effectiveness of Modafinil.
- Evaluate the ability of NLP methods to identify the qualitative findings.

Why Modafinil?

Sudden onset cognitive dysfunction and fatigue are debilitating, and distressing symptoms seen in a variety of conditions and clinical presentations. Modafinil is an out of patent oral wakefulness-promoting drug, first developed in the late 1990s, shown to be relatively safe, and with low abuse potential (39). Currently indicated only for narcolepsy in the UK (40, 41), its US FDA status enables clinicians to prescribe it "off label" to improve cognition or fatigue symptoms in many other conditions. Around 90% of its prescribed US usage is "off label" (42). Modafinil has been considered a potential therapy for a range of conditions (43), including ADHD (44), multiple sclerosis (45, 46), premature ejaculation (47), depression (48), Parkinson's disease (49), chemotherapy related fatigue (50, 51), traumatic brain injury (52), and cocaine dependence (53). Findings have been mixed, with systematic reviews generally inconclusive, showing either insufficient (52, 54–56) or low quality evidence (56–58). Previous studies have commented on the lack of research into either long term (39) or "as required" use (59). However, despite the lack of conclusive trial based evidence there appears to be a substantial amount of online discussion suggesting that there are people for whom it has made a significant difference to their symptoms and quality of life (60).

METHODS

Study Design

Qualitative inductive thematic analysis (61), and basic NLP analysis, of spontaneously generated online patient experience data (SGOPE) (see **Figure 1**). We compared the results of the NLP analysis with those from the qualitative analysis.

Data Selection and Preparation

In January 2017, using google searches, we identified websites containing publicly available text about the experience of Modafinil use. We defined publicly available as where the data identified was available to view by anyone without any form of login, password or registration. We selected sites containing single comment "User review" posts so the type of text was similar from the different sites enabling comparison across the data sources. The final selection included: AskAPatient (62), Drugs.com (63), and WebMD (64) which provided short

Abbreviations: EBM, Evidence based medicine; EHR, Electronic health records; NLP, Natural language processing; QOL, Quality of life; RCT, Randomized controlled trial; SGOPE, Spontaneously generated online patient experience; UGC, User generated content.

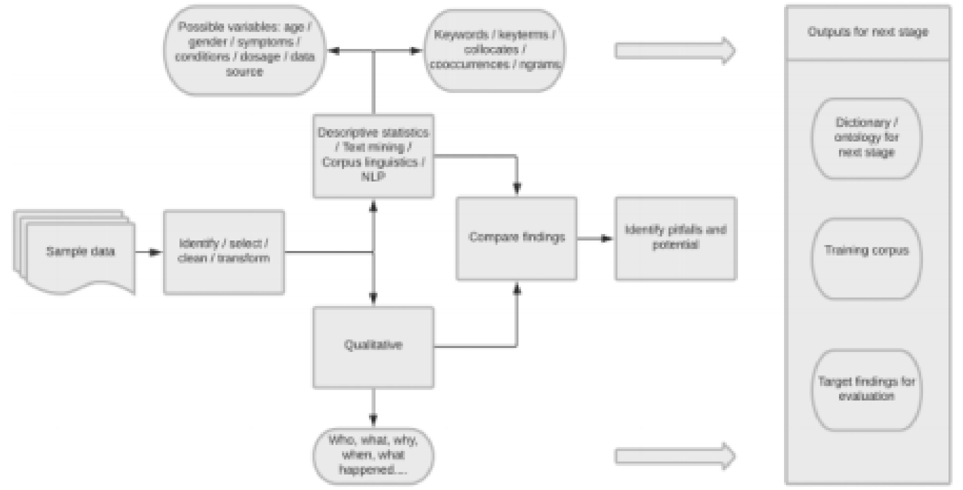


FIGURE 1 | Overview of study design.

accounts of condition-based experiences, and Erowid (65) and ModUp where the posts were longer with greater detail of symptoms, side-effects, and self-experimentation. Online spaces can be transient and unfortunately the ModUp site no longer exists online, but all the others are still visible. From the sites we identified posts made between 1st Jan 2002 and 17th Jan 2017, and searched for individual posts about Modafinil (or variant names Provigil, Armodafinil, Nuvigil) using the site search engine. We then used random number generation to select 260 posts from across the five sites for further analysis. This volume of data was likely to be sufficient to reach data saturation for the qualitative analysis and be sufficient for linguistic analysis.

Each site had its own data structure with a variety of fields. Age and gender self-definition were optional on each of the sites. We standardized the data using the following steps:

- Standardizing field names across sources.
- Translating/encoding coded values: e.g., M/F or male/female.
- Standardizing numerical ratings scores for experience of Modafinil. Erowid and ModUp had no numerical rating; AskAPatient had a rating from 1 to 5 and drugs.com from 1 to 10 for effectiveness of Modafinil, and WebMD had ratings for effectiveness, ease of use, and satisfaction, each from 1 to 5. For the latter, the average of the three scores was calculated. We standardized all ratings to a value of between 1 and 10.
- Ages and duration of taking Modafinil, where given as an identifiable field, were grouped into ranges, and standardized across the sources.
- Posting date simplified to PostYear.

All poster identification was removed, and a unique code allocated to each post. To generate initial descriptive statistics we calculated post lengths, before coding and quantifying any included gender, age groups, duration of taking Modafinil, and numeric ratings.

Ethical Considerations

The ethical issues surrounding the use of SGOPE data for research purposes are complex and continue to evolve (66, 67). Making a clear distinction between public and private spaces online can be difficult (68, 69). SGOPE can be classified as publicly available data (70) but as it was originally collated by the online sites and contains detail of individuals it does not fit the narrower definition of open data which can be freely used, re-used, and distributed by anyone (71). At the time of the design of this study there was a lack of clear guidance from UK Research Councils or other organizations (68, 72). In our methods we tried to minimize the potential for any form of harm.

There has been significant recent debate around expectations of privacy (73, 74). It is impossible to know the motivation, or expectation of privacy of each poster in publishing their content, but posters writing on sites that are password protected or restricted to members may have greater expectations that their privacy will be protected. Concerns exist that individuals could be identified from the posts they make, and that they may consequently suffer harm from some unforeseen use of the data. Potential harms range from unwanted commercial marketing use to profiling that could negatively impact future insurance or career choices (75). However, some studies looking at user attitudes found that social media users were generally positive toward their posts being used for research provided that they were protected from harm and that the research had potential benefit (73). There are examples of social media communities

TABLE 1 | Using categories to identify causal text and perceived effectiveness.

Sequence	Post categories	Text describing
Base state	PreModafinil	Symptoms + context
Action:	Acquisition/dosage	Took/did/prescribed
Consequence	PostModafinil	Effect on symptoms/side effects/context/QOL

deliberately formed in open online spaces to enable individuals to come together to form a voice that is heard by health systems (76, 77).

No IP address or other geographical data was collected, all forms of usernames were removed, and the dates of the post reduced to a year value to minimize any risk of reidentification (69). Use of this type of data is covered under the doctrine of fair use (78, 79). However, we successfully arranged a data sharing agreement with AskApatient and unsuccessfully sought to put one in place with ModUp. Erowid position themselves as working with academics and medical experts and state that they generally agree to research use. However, we received no response from our repeated requests. All of the sites included invited posters to submit experience reports for publication on the respective platform. Content from drugs.com (80) and WebMD (81) carried clear messages to posters that posts were publicly viewable and could be read, collected, and used by others.

Qualitative Analysis

Following familiarization with the posts, the data was coded and the codes merged into themes. We used MaxQDA software (82), using an iterative process of code identification and review as we progressed through the data. The coding and theme generation was done by JW, with discussion and input from FG & JC. For each theme we counted the number of posts in which they appeared.

Evaluating Effectiveness

We categorized text within each post into one of four broad categories, PreModafinil, Acquisition, Dosage, and PostModafinil. These categories align with the base state, action, and consequence sequence required to indicate a possible perceived causal effect (83, 84) (**Table 1**). We compared the coded sections of each post across the sequence categories to identify the poster's own view of whether taking Modafinil was linked to an causal belief and identifiable outcome.

We classified each post for perceived effectiveness (positive, mixed, negative, neutral, unclear) (**Table 2**). We assessed each post in isolation; balancing the positive and negative aspects of language used, reported benefits and side effects, and reference to the continued use or cessation. Fifty posts were initially independently classified by two team members and discrepancies discussed. JW then classified the remaining 210 posts.

For posts which had associated numerical ratings we categorized ratings of 0–3 as negative, 4–7 as mixed, and 8–10 as positive. Using chi squared test we compared our manually assessed classification with the poster's rating.

NLP

The narrative fields were extracted to create a corpus. Due to the small size of this exploratory dataset, we used a corpus linguistics tool, SketchEngine (85) for the structural analysis of the text. Typical NLP projects return best results from very large datasets, while corpus linguistics can be used on smaller data sets of the size also amenable to qualitative analysis. Corpus linguistics and NLP share some similar analysis techniques (86). Pre-processing for both NLP and corpus linguistics begins by dividing the text into tokens representing the smallest possible linguistic unit. Each token was assigned a part-of-speech (POS) tag from the English TreeTagger POS tagset with Sketch Engine modifications (87). We used stemming and lemmatization to assign inflected words to the same term, reducing the number of inflectional forms of a word and reducing variants to a common base (88, 89).

We used case independent word frequency and term extraction. Similar to TF-IDF of NLP, term extraction identifies the terms most specific to the text by calculating term frequency in the text compared to frequency of the same term in the reference corpus. For our reference corpus we used the English Web corpus 2013 (enTenTen13) (90), a corpus of 19 billion words collected from online sources. We extracted the top 500 specific keywords and terms. The top 100 of each indicated the most prevalent topics. The least frequent were used to identify instances of spelling variations or non-words; these were added to the domain specific dictionary intended for use in the next stage of the project.

Entity Identification

To identify relevant entities, we used the following POS tokens tagged as nouns:

- Drug Names—both name variations of Modafinil and other drugs; those taken previously, concurrently, or subsequently in addition to some that may have no relevance to the post.
- Condition Names—identifiable condition names were categorized from term extraction analysis. Sleep related disorders were classified in line with the ICSD3 classification systems (91).
- Symptoms—symptoms of interest in this study relating to fatigue or cognitive issues. Initial dictionary entries were created from common synonyms, with further additions identified from the previous analysis.
- Action—the action of taking Modafinil has two main components: amount and frequency. Terms and phrases to identify both were found within the posts and included in the dictionary.
- Side Effects—term extraction was particularly useful in identifying side effects that the poster described, as patients often use a wide range of terms to describe them that may not map easily to recognizable medical terms.

Relationship Identification

We used three methods to identify the relationships between entities in order to understand the semantic meaning of the text:

1. POS tagging of verbs occurring between entities to indicate simple relationships;

TABLE 2 | Examples of sentiment grading.

Grade	Explanation	Example
Positive	Overall positive. Ranged from overwhelmingly positive to indicating that benefits outweigh the disadvantages	"Significantly improved my quality-of-life with the only side effect being minor occasional headache" (2400)
Mixed	Both positive and negative effects were reported; unclear as to which sentiment prevailed	"It is a lifeline to me, but the side effects are many and do suck. Though i can honestly say, i don't find it addictive." (1117)
Negative	Predominantly negative, usually regarding side effects.	"… feel really strange shaking overall out of sorts I am not falling asleep at work but feel so weird I'm wired that I need to find something else" (1123)
Neutral	No response or side effects noticed.	"…didn't notice any effect whatsoever, not even a side effect. "(330)

2. Collocation analysis (92) to reveal patterns and meanings that may not be apparent from frequency lists or manual reading of the texts;
3. Co-occurrence analysis: this assumes that if two entities co-exist within so many words that there is an underlying relationship between them. Unlike collocations, the relevant words need not be adjacent to each other, but occur within the same unit of text. Co-occurrences can highlight relationships indicating a causal link such as a side effect, outcome event, or demonstrate a negated drug event—one which denies a causal relationship between the drug and the event.

To identify possible causal text, we split the corpus into to sub corpora based on the text categories PreModafinil and PostModafinil (see section Qualitative Analysis above) and used n-gram analysis on each, looking for phrases between 3 and 5 words long that occurred at least five times in the corpora. Where an ngram was ambiguous we examined the co-location and co-occurrence analysis to assist categorization.

Sentiment Analysis Using NLP

To evaluate sentiment we used the Python "TextBlob" package (93) to calculate the polarity of each post as a value between −1 (negative) and +1 (positive). Pre-processing included converting text to lower case, removing punctuation, and removal of the default stop words.

Comparing the Two Methods

We manually mapped each of the 100 most frequent key words and terms from the computational corpus analysis to the themes that emerged from the qualitative analysis. Where a word/term was ambiguous or related to negation, time or scale we placed them in a separate group.

To compare NLP sentiment analysis to the qualitative categorization of positive, mixed, neutral, or negative we used two comparison scales. The first classifying a "mixed" result as being in the range ±0.01 (**Table 6**) and the second widening the "mixed" range to ±0.05 (**Table 7**). In both cases a polarity value of 0 was mapped to Neutral.

We mapped each of the 3–5 word length ngrams to the themes from the qualitative analysis. Where an ngram could apply to more than a single theme, we used the collocation and co-occurrence techniques in order to map it to the theme or group for which it was most prevalent.

We compared the NLP sentiment analysis with the qualitative analysis results for perceived effectiveness of Modafinil as follows: comparison of totals for each type of perceived effectiveness/sentiment; comparison of analysis of individual posts. The accuracy of the post level comparison was assessed using a confusion matrix.

RESULTS

The dataset included posts with a total length of 72,427 words (average 279; minimum 15; maximum 2,384). Posts from AskAPatient (30–417 words), Drugs.com (15–204), WebMD (29–358), Erowid (44–2,384), and ModUp (125–1,030).

Of the posters, 158/260 (61%) identified their gender and 156/260 (60%) included their age, either as an integer or as being within a range. From the two sites with 100% gender identification, there were 65% female posters on AskAPatient and 22% on Erowid. The defined age-groups ranged from under 18 to over 75, with the largest age-group being 45–54 years.

The quantifiable length of time that posters stated they had been taking Modafinil was included in 184/260 (70%) of posts. Of these 34 (18.5%) had taken it for 7 days or less, 31 (17%) 8–31 days, 61 (33.1%) for between 2 and 12 months and 58 (31.5%) for longer than 1 year.

Qualitative Analysis

We identified eight themes which we describe below.

Reason for Taking Modafinil

All posts were concerned with finding a solution for symptoms of fatigue, sleep and or cognitive dysfunction. Although Modafinil is only indicated for a single condition within the UK, 33 different health conditions were mentioned within this small sample of 260 posts. The most frequent were central disorders of hypersomnolence (mentioned in 26% of posts), depression (22%), sleep related breathing disorders (16%), general fatigue (9%), CFS/ME (7.5%), ADHD/ADD (6%), and MS (6%). Other conditions included cancer, traumatic brain injury, diabetes, epilepsy, fibromyalgia, autoimmune conditions, pain, IBS, hepatitis C, or post stroke fatigue. Multi-morbidity was a regular feature. While many posts referred to a single diagnosed condition, 23% referred to two concurrent conditions, 3% to three and 1.5% to four.

Impact of Symptoms

Almost all posts contained detail of how these fatigue or cognitive symptoms affect their lives, emotionally, socially, and practically. Responses to their conditions included fear, desperation, hopelessness, resignation, embarrassment, and guilt:

> *Life was miserable. I was being treated for depression and had even considered suicide. There was no way out of this rut.* [422]

> *I had resigned myself to life handicapped with fatigue, and I felt really hopeless about it* [321]

Frustration was a common theme, often at their own inability to engage with "normal" life.

> *I couldn't stand being this form of myself any longer—it's not me* [424]

Symptoms were described as having considerable impact on family and social relationships, putting a strain on marriages, partnerships and affecting parenting:

> *My husband gets sick and tired of me being tired all the time and particularly hates it when I have to have a nap* [503]

> *Before Nuvigal I couldn't keep my eyes open and live my normal life with 3 boys! Now, after Nuvigal I can actually play with my kids and be a normal mother.* [2348]

The loss, or anticipated loss of a job featured in 47 (18%) of the posts and 18 (7%) posters detailed their fear of driving, either because they had experienced falling asleep at the wheel or were concerned that they would.

Effectiveness of Modafinil

Posts were classified as follows: 68% positive, 18% mixed, and 12% negative; four posts were neutral (see **Table 2**). A total of 181 posts had the potential to include a numeric rating of the effectiveness of Modafinil of which 178 posters completed the rating. The average value (after standardization) was 7.5/10. We found no significant difference between the posters numeric rating and our assessment (χ 3.3419, $p = 0.3$).

There was considerable variation in the proportion of posters reporting positive effect of Modafinil across the different sites: positive values ranged from 46 to 100%, mixed from 0 to 27%, and negative from 0 to 25% (see **Table 3**).

Impact of Effectiveness on QOL

A recurring topic among those finding Modafinil effective, was how it allowed them to return to what they felt was their personal "normal" state rather than enhancing their abilities in any way.

> *This stuff is pretty amazing, i can actually have a normal day rather than fighting just to get through one. It's not what i feel but what i don't feel which is the constant fatigue, without that life has returned to "normal."* [1388]

TABLE 3 | Manual assessment of perceived effectiveness across data sources (% age).

	AAP ($n = 79$)	DCUR ($n = 53$)	Erowid ($n = 41$)	ModUp ($n = 38$)	WebMd ($n = 49$)	All sources ($n = 260$)
Positive	46%	72%	78%	100%	61%	67%
Mixed	27%	15%	12%	0%	27%	18%
Negative	25%	13%	7%	0%	10%	13%
No effect	1%	0%	2%	0%	0%	1%

Dosage

Of the 141 (55%) posts included text relating to Modafinil dosage the reported dosage taken ranged from 25 g to 1,200 mg per day in one extreme case. Although clinical guidelines usually suggest 200–400 mg daily (94), there are indications that a lower dose was found to more effective for some posters, with 17 reporting taking 100 mg/day. Tolerance was described as an issue for some, with 51 (20%) posters commenting on an apparent reduced effectiveness after weeks or months of regular daily use. Some posters reported that stopping taking Modafinil for a few days before resuming a daily dose appeared to restore its effectiveness

> *After a week or so, effects not as strong and can make you feel paradoxically very tired. Take 2–3 days off, and it will resume working.* [2344]

whereas others felt it was better to take it only when they felt that they would most benefit from it:

> *I did notice however that I have to take breaks from it for it to remain effective. I now only take it if I have a full day planned and have to go out, otherwise I stay at home and take a nap.* [502]

The posts also illustrated how users have experimented to find a dosage pattern that they find effective (**Table 4**). Almost half the posts contained text detailing the variations in frequency they had tried and those they found most effective. Comments also included the cause/effect results of experimentation of increasing or lowering the dose, taking before or after meals, with or without alcohol and how that impacted on the side effects and effectiveness

> *I found if i took 50 mg every couple of days, and then 100 mg on busy days, it kept the headaches/migraines at bay.* [1117]

Side Effects

Of the 260 posts, 128 (49%) specifically mentioned one or more side effects they considered related to the use of Modafinil. Thirty-four posts (13%) stated that they did not suffer any side effects at all, while the remaining 98 (38%) did not mention any specific side effect. Across the sample the most commonly reported side effects were headaches (57), mental health/mood

TABLE 4 | Qualitative analysis: dosage frequency.

	Documents	Percentage	Percentage (valid)
Daily	53	20.38	44.92
As required	30	11.54	25.42
Twice daily	21	8.08	17.8
M-F	13	5	11.02
Unclear (variable)	10	3.85	8.47
Every other day	2	0.77	1.69
Posts with code(s)	118	45.38	100
Unspecified	142	54.62	–
Total	260	100	

related (43), appetite (30), gastric (18), urinary (16), oral (16), skin (15), cardiovascular (11), jittery (10), and difficulty sleeping (10). Other side effects including difficulty sleeping, muscular, vision effects, motor function, weight gain, tinnitus, shortness of breath, magnified pain, neuropathy, lupus flare up, swollen tongue, weight loss, and increased libido were mentioned by <10 posters. The impact of side effects varied, 12 posts described them as minimal, while 13 felt they were temporary, passing within a few days. Nine posters stated that they had stopped taking Modafinil; eight due to side effects and one because of an interaction with an MAOI antidepressant.

Acquisition of Modafinil

Detail of how the poster found out about or acquired Modafinil was present in 136/260 (52%) posts, with 82 (31%) stating they were prescribed Modafinil by a clinician, while 54 (21%) discovered it through either their own research or via word of mouth. Difficulties in obtaining it, either within the NHS where its use is restricted to narcolepsy, or in the US where insurance companies often will not cover the cost despite clinicians prescribing it, were mentioned by 37/177 (21%) of those finding Modafinil beneficial. Self-purchasing from online sources was reported by 35 (13%) of posters:

> *Now because they say Modafinil is not a bi-polar medicine they refuse to pay for it. I will not be able to afford the $650 a month. Without it I wake with nightmares. It's very sad insurance says they know better than a group of doctors and 10 years of success using a prescription* [2098]

Other Interventions

Almost all posts included details of previously prescribed or tried interventions including self-help or lifestyle changes, and any interventions taken in combination with Modafinil. Posts often include comparative descriptors both of effect and/or side effects of the alternative intervention or combination.

> *I find modafinil it more effective than caffeine although the initial effects seemed to wear off after about 8 hours or so. There are definitely less side effects than with other prescription stimulants such as phentermine or ritalin.* [2016]

Causality

Among the 260 posts, we manually identified text relating to the perceptions of the poster's experience both pre and post Modafinil in 209 (80%). Of these, 258 (99%) contained text relating to the effect of taking Modafinil. Identification of causal text was helped by the reported rapid onset of any effect, with many posters who believe it to have an effect, either positive or negative, noticing changes within an hour of taking it.

Comparing Qualitative and Corpus Results

Of the 100 highest frequency keywords 88 mapped directly to qualitative themes, seven related to negation or scale and 5 could not be classified. Of the 100 highest frequency key terms, 84 mapped directly to the qualitative themes, seven referred to negation and temporal aspects, and nine could not be classified (**Table 5**).

Sentiment Analysis

The NLP TextBlob package returns sentiment polarity as a value between −1 (negative) and +1 (positive). Of the 260 posts 188 (72%) indicated positive sentiment, 10 (4%) neutral and 62 (24%) negative. The range of polarity values of posts was from −0.26 to 0.4. **Tables 6, 7** show the results of comparing the classification of each method for each post. Matching was accurate in 64% of posts. If we allow for one category difference matching was accurate in 85% of posts.

The 3–5-word ngram analysis on both the pre-Modafanil (35) and post-Modafanil (106) text generated ngrams classified into the eight themes and 6 categories reported in **Table 8**.

As with the keywords and keyterms we found that many of these ngrams correlated with and mapped onto the themes that emerged from the qualitative analysis. Others related specifically to temporal, sequential, negation, or confirmation text that could be used to identify phrases inferring causality. The frequently occurring ngram *"I have found that"* seen in nine posts was used to describe ways of taking the drug to maximize the effectiveness. Examples of generic ngrams and the context in which they were used are given in **Table 9**.

We were able to match ngrams to the expression of causal analysis identified by the qualitative analysis (**Table 10**).

DISCUSSION

Within this exploratory study of the unstructured narrative post content, both methods successfully demonstrated how the majority of posters with a wide range of conditions found Modafinil effective in reducing fatigue or cognitive symptoms.

In performing the human based qualitative study first, those findings acted as an informal benchmark for the automated NLP study. The eight themes generated reflected the main aspects of patient experiences of an intervention. It also explored the detailed context that was often included within the poster's evaluation, including the reasons for starting or stopping using it, comparisons with other medications that they may have tried or moved onto, side effects and tangible or intangible effects on their quality of life.

TABLE 5 | 100 highest frequency keywords and keyterms by topic.

Theme/type of text	Keyword	Keyterm
Drug	Modafinil; provigil; nuvigil; armodafinil; modalert; nootropic; modafanil; modafinal; modvigil; nuvigal; modavigil; moda;	200 mg provigil;
Condition	Narcolepsy; hypersomnia; apnea; idiopathic; fatigue; fibromyalgia; cfs; insomnia;	Sleep apnea; daytime sleepiness; sleep cycle; chronic fatigue; excessive sleepiness; sleep disorder; excessive daytime sleepiness; extreme fatigue; severe sleep apnea; obstructive sleep apnea; shift work; nerve entrapment;
Symptom	Sleepiness; sleepy; tiredness; drowsiness; asleep; fatigue; drowsy; procrastination; sleep; lethargy; nap; spaciness; procrastinate; doze; irritable; tired; exhaustion	Head fog; daytime sleepiness; sleep cycle; chronic fatigue; excessive sleepiness; term memory; day time; short term memory; anxious state; afternoon fatigue; constant fatigue; brain fog;
Acquisition	Reddit; mymodafinil;	Prescription drug; sleep study;
Dosage	mg; dose; tolerance; pill;	200 mg dose; 200 mg pill; full dose; first dose; second dose; empty stomach; 100 mg dose; second pill; 200 mg provigil; 1st week; drink plenty; daily dose;
Side effect	Jittery; jitter; headache; hallucination; irritability; impulsiveness; irritable; itchiness; appetite; nausea; grouchy; clench; bpm;	Side effect; dry mouth; smelly urine; heart rate; mild anxiety; slight headache; unpleasant side; bad side; heart beat; negative side; anxious state; jittery feeling
Other drug/intervention	Stimulant; piracetam; ritalin; cpap; caffeine; amphetamine; ephedrine; adderall; adrafinil; phenylephrine; bupropion; ssri; med; pseudoephedrine; caffene; methylphenidate; fluoxetine; cocaine;	Taking bupropion
Effect	Wakefulness; awake; alertness; euphoria; enhancer; psychoactive; nighter; schoolwork; talkative; lifesaver; palpitation; impulsiveness; amped; chatty;	Term memory; cognitive enhancer; normal sleep; productive day; mental acuity; normal sleep schedule; mental clarity; positive impact; short term memory;
Outcome		New person; normal sleep schedule; miracle drug; wonder drug;
Negation	Didn't; wasn't; hasn't; couldn't; wouldn't; hadn't	I didn't; i wasn't;
Temporal		1st week; first dose; second dose; hour period; entire day;
Scale	Hyper	
Ungrouped	Comedown; sleepless; midterm, cephalon; had	Sleep deprivation; side note; trouble sleeping; placebo effect; college student; enhancing drug; study aid; year old male; work day;

TABLE 6 | Sentiment analysis confusion matrix [±0.01].

	NLP						
Qual	**≥0.01**	**+ = -0.01 to <0.01**	**0**	**≤-0.01**	**Total**		
Positive	145	2	4	23	174		Agreed evaluation
Mixed	28	3	3	13	47		1 category different
Neutral	0	0	1	3	4		2 categories different
Negative	11	4	2	18	35		Completely opposite evaluation
Total	184	9	10	57	260		
Accuracy	0.642						

green - Agreed evaluation, beige - 1 category different, plum - 2 categories different, and red - completely opposite evaluation.

The sample size was too small to realistically expect good results from the NLP analysis, but by using the corpus linguistics tool which used some methods found in a full NLP approach we were able to demonstrate how an NLP methodology could be used on a much larger scale to both extract topics/themes, expressions of perceived causality and evaluate effectiveness from unstructured text.

As with a recent paper comparing grounded theory with topic modeling on survey data (95), our NLP based methods successfully identified many of the qualitative findings, demonstrating how this form of data has the potential to identify effectiveness and the topics discussed within the posts. In terms of sentiment analysis, the results highlight some of the current issues with NLP methods. Although both methods show a majority of posters finding it effective for them, the confusion matrices (**Tables 6, 7**) highlighted some of the issues with applying generic sentiment analysis tools to health-related data. Rule based methods that determine sentiment are based on a lexicon of prelabelled words and the accuracy of the results is heavily dependent on the data that the model was trained on and the words that are considered important to that model. The majority of the existing generic NLP sentiment analysis tools were trained on either film, restaurant, or Amazon product reviews as these represent some of the largest shared annotated sentiment resources (11). Looking at some of the posts with opposing categorizations (**Table 11**), demonstrates how many of

TABLE 7 | Sentiment analysis confusion matrix [±0.05].

	NLP						
Qual	**≥0.05**	**+ =-0.05 to -0.05**	**0**	**≤-0.05**	**Total**		
Positive	130	26	4	14	174		Agreed evaluation
Mixed	24	9	3	11	47		1 category different
Neutral	0	0	1	3	4		2 categories different
Negative	9	11	2	13	35		Completely opposite evaluation
Total	163	46	10	41	260		
Accuracy	0.588						

green - Agreed evaluation, beige - 1 category different, plum - 2 categories different, and red - completely opposite evaluation.

TABLE 8 | PreModafinil and PostModafinil 3–5-word ngrams grouped by theme.

Theme/category	PreModafinil	PostModafinil
Reasons	I was diagnosed with; obstructive sleep apnea; chronic fatigue syndrome; I have been; sleep apnea and; I suffer from; at the age of;	
Symptoms	I have been; that I was; I found myself; I have to; I suffer from; for the last; I was a; at the age of; and I was; I used to; I was still; I wake up; I had to	
Other interventions	I started taking;	I was on;
Acquisition	I went to; I was prescribed; I decided to	
Dosage		Early in the morning; I don't take it; I have been taking; I started taking; I take it; I don't take; I have found that; I took it; I have to; in the morning; on days that I; if I don't; when I don't; to take it
Side effects		With no side effects; I didn't notice; and I was; don't have; the next day; I don't feel; the first time I took; the side effects;
Effectiveness		I am able to; I don't feel; to be able to; get out of bed; I didn't notice; first time I took; I began to; and I was; don't have; the next day; go to sleep; I feel like I; I have found that; I was able to; I don't think; I have not; I felt like I; I used to; if I don't; that I could;
Outcome		To go to; to be able to; was able to; that I could; I felt like I;
Temporal	All the time; during the day; through the day; for the last; in the morning;	A few days; first time I took; during the day; for a few days; in the morning and; through the day; the first time
Sequential	For the last; I used to; I was still;	At the same time; as soon as I; for the first time; for a few days; the next day: I have found that; I used to; on days that I; I had to; if I don't; if I need to; the first time I took; the first time;
Negation	I don't; I didn't;	Don't have; I can not; I didn't feel; I didn't have; I didn't notice; I don't feel; I am not; I did not; I don't have; I don't know; I have not; I do not; I was not; it does not; n't be able to;
Confirmation		I was able to; I felt like I; I want to; I used to; I was on; I had to; it was a;
Ungrouped	A lot of	A lot of; hours of sleep; I need to; I have been; that I could; that I had; that I have; that I was; the rest of the; to be a; to take a
Causal ngram		I began to; I have found that; on days that I; if I don't; when I don't;

the concepts that posters describe in their evaluations include stopwords or words that may not be evaluated as expressing sentiment. Improved accuracy will require the development or use of a domain specific model.

Compared to Current Evidence

These findings of overall effectiveness contrast strongly to the existing current RCT and systematic review evidence, which are generally used to determine treatment pathway options for clinicians (96). Although various RCTs have looked at Modafinil as a potential therapy across a range of conditions, findings have been mixed, and the systematic reviews generally conclude that the evidence is either inconclusive or of insufficient quality (44–47, 49, 50, 52, 53). This contradiction may have implications on both on patient care and the efficiency of healthcare provision, either through the patient not receiving an intervention that may be effective, or by receiving one that is ineffective (97, 98).

How SGOPE Can Complement RCTs in Generating Evidence

Our results demonstrate how SGOPE can help address some of the identified issues with a research driven agenda (15)

TABLE 9 | Example ngrams in context.

ngram—I have been (PreModafinil)—categorized as "Reason for taking"
I have been battling MS Fatigue to the point of almost thinking of quitting my job, but desperately need the money.
Depression I have been working for many years with one combination after another of medications for bi polar disorder.
Before Nuvigil I have been suffering for the past 3 years or so with marked fatigue.
For the last few years I have been taking medicines to calm me down and ease my stress levels.
With that said, recently I have been back and forth to the doctor for 4 months now.
I have been fighting this for as long as my memory will take me.
I consider my own experiences to be significant in that I have been on the SSRI cipralex (celexa, escitalopram) since age 19, having experienced bouts of diagnosed major depression in my late teens.
I have been taking Provigil for about 9 months now after my sleep disorder kept me awake for 8 days even after being on a 6 mg dose of lorazepam to sleep at night for several years.
In addition to Provigil I have been on Effexor XR at 150 mg/day for my mild depression.
I have a severe lack of motivation and I have been diagnosed with ADHD.
I have been diagnosed with and suffering from Idiopathic Hypersomnia for the last 6 years.
I have been through 2 sleep studies and I wasn't quite a match for the CPAP machine but according to my doctor at the Mayo Clinic, there is obviously something wrong with how tired I am and how easily I can fall asleep.
ngram—I used to (PreModafinil)—categorized as "Symptoms"
I used to fight sleep all day at work, it got to the point where I was staying home because I just couldn't stay awake.
After the buzz wore off, my life became normal, which was a great improvement over the constant feelings of lethargy and helplessness I used to feel.
As someone who works online, as a writer and retailer, I used to find myself researching an article 1 min, and somehow snapping out of a haze a few hours later.
I used to drink coffee for this kind of thing, but tolerance builds up quickly and by the end of exams I'd be drinking a few cups a day and it made me feel no good.
I used to be a PhD student who was heavily dependent on Adderall for a cognitive and motivational boost.
ngram—I have found that—(PostModafinil)—categorized as "Dosage"
I have found that if I don't take it on the weekends that it works better.
Overall, I have found that effects of Modafinil, for me at least, are extremely subtle and almost unnoticeable until I start to think back and examine the things that I have done on a given day.
I have found that a very effective remedy is to take a couple of co-codamol tablets, which each contain 8 mg of codeine.
I have found that my own lack of worry or guilt in situations like these prevents people from becoming suspicious—nobody batted an eyelid.
The one concern I have is that I have found that cutting the dose (as I did once for several days when I didn't place the online order in time) seems to have a dramatic negative effect.
I have found that Modafinil gives me a very clear mind for problem solving.
I have found that if I skip the workout, I don't have as much energy throughout the day.
I have found that if I eat a lighter lunch, the dip is not as bad.
I have found that overall mod just works best on its own.
ngram—the first time (PostModafinil)-categorized as "Effectiveness," "Temporal," and "Sequential"
The first day on Nuvigil I felt like I had never felt before: My mind felt awake for the first time in what seems like forever.
I feel my age for the first time ever!
I have done some research and found that taking a "drug holiday" or going a day or two out of the week without it will help it stay just as potent as the first time I used it.
Like many others, the first time I took it was great!
The first time around, I nearly drove my family crazy with my talking and myself crazy trying to keep my rapid thoughts to myself.
The first time I took it, I did not have the headaches, smelly urine, post nasal drip, fuzzy vision, or muscle aches.
All I can say is that I'm now a 4.0 college student and I feel like I'm actually awake for the first time in my life.
The first time I took this (prescribed for obstructive sleep apnea) I thought "wow, this is the answer!"
At first I was skeptical that it had been the Modafinil that had caused the happiness because I tend to go through short bursts of depression and happiness and I assumed that I had just been on a good day the first time I took it but looking back I haven't had any significantly bad days while I was on Modafinil.
I am dramatically more productive at work and for the first time in my like I feel capable of planning for the future.
I will never forget how deep my mind sank that week, it was the first time I'd ever felt truly depressed—not even extended family deaths or the comedown from 220 mg of pure MDMA was as bad as how I was feeling that morning.
When I took it for an exam the first time it was amazing. I took 400mg at around 8 and stayed up the entire night studying with no problems.
I was captivated by his work, which really was excellent (we both received 1st for our efforts), but for the first time that day I was ever so slightly distracted while I was reading.
The first time, I seriously wondered about the efficacy of the things...
I think it important to note that I have never taken it every day and usually never take more than 200 mg. 600 mgs, which I took for the first time today, really has me jacked.

(Continued)

TABLE 9 | Continued

The first time I tried Modafinil the effect was immediate.
Right from the first time I took Modafinil not only did I feel better, but I could tell how much more I was worried about my work, the amount of detail I would put into my projects even shocked myself.
The first time I took modafinil I understood what all the hype was about.

TABLE 10 | Examples of causation reason and consequence.

Document	Causation: reason	Causation: consequence
1,063	***I used to*** fight sleep all day at work,	Once I started Nuvigil I have not had this problem at all, ***I feel like I*** have more energy, and my mental alertness has improved 100%.
1,090	I've tried 44 anti-depressants and only got about 25% relief	Nuvigil obliterated my depression and eased my anxiety by about 60%
1,065	Prior to provigil, I would regularly fall asleep at work or in meetings. I was afraid to drive alone for more than an hour for fear that I would fall asleep driving (I had many close calls!).	This drug has made a significant improvement in my quality of life.
1,065	I ***had felt*** (just weeks earlier) that I could not go on any further.	***I can keep*** my job and haven't wrecked any cars.
1,136	Constant feelings of lethargy and helplessness ***I used to feel.***	This drug has returned my pre-MS life to me. ***I can fully f***unction on the job
1,207	I forgot to take it 1 day	And could not stay awake and could not stop eating, just like prior to starting Provigil.

and complement RCTs. One of the possible reasons for the inconclusive trial evidence to date is the heterogeneity of effect that can occur within trials (99). Trials generally exclude participants with multiple comorbidities as these may act as confounders when measuring effectiveness (97) whereas many of the posters have two or more co-existing conditions, and may use combinations of interventions, or react to a single intervention in different ways.

Systematic reviews show how trials report either the effects of a single dose or a regular daily dose for a limited time (48, 100–102), whereas our findings include much greater variety of usage patterns. Our results illustrate how some posters have varied dosage patterns and amounts to find the optimal dosage regime for them, with some finding that lower doses than those usually prescribed were more effective. The data also demonstrated the existence of a possible tolerance effect but included the suggestion that taking occasional breaks or taking as required appeared to be a viable method of retaining effectiveness over time. Identified side effects generally reflected those already known (94), however the retrospective nature of the posts enabled the discovery of the temporary nature of some common side effects, a factor that will not be reflected in single dose trials.

Identifying Causal Inference

Studies have begun to look at the lexical and grammatical features of causal statements in free text (84) and some work has been done using NLP to identify pharmacological adverse events from social media (33, 103, 104) suggesting that negative effectiveness can be shown from this type of data. Identifying causal text requires showing temporality; the effect occurring after the cause. Dividing the corpus into pre and post intervention by tagging the tense of tokens facilitated this classification, while ngrams and other POS tags helped us identify sequential events.

One of the issues of identifying causality in any kind of study has always been in differentiating between correlation and causation (105). Identified patterns and correlations can indicate that "something is happening" but not necessarily explain "why" (106, 107) as it does not differentiate between the causes of patterns, whether they are true, coincidental or as a result of bias. Increasing the volume and range of data may achieve a higher degree of precision and external validity (108) and while summarizing and visualization may be useful in analyzing SGOPE datasets, correlation is not the same as causation and on its own it is unlikely to be robust enough to add to an evidence base.

In our study, strength is demonstrated by how almost all posters reported an effect, either positive, negative or mixed. By using multiple data sources and including patients with a wide range of conditions we have shown consistency of findings across populations. The reported rapid onset of effect shows specificity and a biological gradient, with the cause/effect sequencing showing temporality.

The purpose of our research is not to provide a statistical proof of effectiveness across the whole patient population, but to generate a better understanding of the patient experience of using Modafinil, by exploring individual patient's perspective of whether or not it is effective for them. Causal dispositionalism is an alternative theory to the non-reductionist approach to causation, which may be relevant to this type of data. This takes a more nuanced view of how the characteristics or dispositions of both the intervention and the individual combine to affect the effectiveness (109). Rather than taking a statistically based population level view, marginal cases, and outliers are used as

TABLE 11 | Example posts with conflicting sentiment analysis results.

Manual grade negative—NLP grade positive	Manual grade positive—NLP grade negative
*First day was great (started **ar** 150 dose) then falling asleep during day. Increased to 250, didn't fall asleep during day but very nervous and couldn't sleep at night. Going to breakup dosage to see if that helps Side Effects Itching, **can't** sleep at night* [1146]	*I have sleep apnea and also Multiple Sclerosis. I use the 100mg tab, but not on a daily basis. I use this when I feel tired from my MS. I recommend this for MS patients that tend to have no energy when they wake up in the mornings. It doesn't seem to affect my symptoms of MS; i.e., the tingling in my feet or legs. It just gives me energy to get **throught** the day, when I need to do what I need to do. I see my Neurologists for the medication prescription.* [5037]

a starting point for further investigation of potential predicates (110). However, no matter how accurately causal text is identified, the possibility of a placebo effect, recognized as a powerful factor in a patient's assessment of effectiveness both in and out of trials (111–113) means that it is impossible to tell how much of the sentiment toward effects, either positive or negative, is due to such an effect rather than the Modafinil itself.

Strengths and Limitations

Using content purely from the public domain is both a strength and a limitation. Although the easiest to access, it may not contain the richest patient experience data, which may be posted on sites requiring a "login." However, using public domain data enables future replication. Validity is increased by using a diverse range of data sources. Each site comprises posts from a "community" of people who feel comfortable there, potentially leading to an element of emotional contagion between the posters (114, 115). This clustering of individuals can lead to a confirmation bias as consensus has been shown to have a positive impact on the perceived effectiveness of treatment (116). Using multiple sites can mitigate this type of contagion while the scale of the data being analyzed should negate the problems of an individual post being incorrectly classified or missed. Although there will always be an element of the unknown about the motivations and authenticity of such posts, analyzing them on a large scale rather than just a small subsection, can negate the impact of those individuals or organizations who might try to create an inaccurate impression, while techniques are continually being developed to identify spam or non-genuine posts.

As the content is generated entirely by the poster, SGOPE relies on the poster's self-description of their condition, which may include self-diagnosis, rather than that of a clinician. Reporting of symptoms and outcomes may not be as accurate or complete as it could be although this limitation can apply to any form of self-reported data, whether in a trial, clinical encounter, or online. Self-reported data, especially on hard to measure factors such as fatigue and cognition is subjective, but generally reflects the normative value of the patient. The natural, non-clinical language used within unstructured text can contain valuable information that may remain unexplored in a clinical or research setting (117), but it can also contain many spelling or grammatical errors as well as slang terms or colloquialisms that are problematic even for NLP methods created for electronic health records (EHRs) (118).

Future Research

The next study in the project will be a fully NLP based analysis of a much larger dataset of patient experiences of Modafinil use. Having identified some of the possibilities and potential pitfalls, we will use these findings to develop methods that can be subsequently generalized to evaluate other interventions from unstructured text.

CONCLUSION

We have demonstrated how SGOPE shows potential for the identification of perceived causation and evaluation of the effectiveness of Modafinil. The findings show that in comparison to the current inconclusive evidence, most posters find Modafinil to be effective in dealing with fatigue and cognitive symptoms across a wider range of conditions. Our study shows the potential for new research methods and data sources to augment existing knowledge. Although the two methods are very different, we demonstrate how computational methods can extract the same main topic areas as qualitative analysis. Although much work is needed to refine the techniques and address the challenges identified, our comparison suggests NLP can be used to look beyond the literal meaning of the words, gaining an understanding of how posters assess the effectiveness of a healthcare intervention and the outcomes that they value, on a much greater scale than is possible from qualitative studies.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found at: https://github.com/jmw999/P1.

ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

AUTHOR CONTRIBUTIONS

JW conceived the study design, conducted the study, and drafted the paper. FG and JC contributed to study design, advised on study conduct, and contributed to editing the paper. All authors contributed to the article and approved the submitted version.

FUNDING

This research was funded by University of Warwick.

ACKNOWLEDGMENTS

Ines Kander undertook independent review of posts.

REFERENCES

1. Mueller J, Jay C, Harper S, Davies A, Vega J, Todd C. Web use for symptom appraisal of physical health conditions: a systematic review. *J Med Internet Res.* (2017) 19:e202. doi: 10.2196/jmir.6755
2. Vicari S, Cappai F. Health activism and the logic of connective action. A case study of rare disease patient organisations. *Inf Commun Soc.* (2016) 19:1653–71. doi: 10.1080/1369118X.2016.1154587
3. Oprescu F, Campo S, Lowe J, Andsager J, Morcuende JA. Online information exchanges for parents of children with a rare health condition: key findings from an online support community. *J Med Internet Res.* (2013) 15:e16. doi: 10.2196/jmir.2423
4. Dewing C, Jones T. *Future of Patient Data.* London: Future Agenda (2018).
5. Vilar S, Friedman C, Hripcsak G. Detection of drug-drug interactions through data mining studies using clinical sources, scientific literature and social media. *Brief Bioinform.* (2018) 19:863–77. doi: 10.1093/bib/bbx010
6. Yin Z, Sulieman LM, Malin BA. A systematic literature review of machine learning in online personal health data. *J Am Med Inform Assoc.* (2019) 26:561–76. doi: 10.1093/jamia/ocz009
7. Abbe A, Grouin C, Zweigenbaum P, Falissard B. Text mining applications in psychiatry: a systematic literature review. *Int J Methods Psychiatr Res.* (2016) 25:86–100. doi: 10.1002/mpr.1481
8. Convertino I, Ferraro S, Blandizzi C, Tuccori M. The usefulness of listening social media for pharmacovigilance purposes: a systematic review. *Expert Opin Drug Saf.* (2018) 17:1081–93. doi: 10.1080/14740338.2018.1531847
9. Demner-Fushman D, Elhadad N. Aspiring to unintended consequences of natural language processing: a review of recent developments in clinical and consumer-generated text processing. *Yearb Med Inform.* (2016) 1:224–33. doi: 10.15265/IY-2016-017
10. Dreisbach C, Koleck TA, Bourne PE, Bakken S. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. *Int J Med Inform.* (2019) 125:37–46. doi: 10.1016/j.ijmedinf.2019.02.008
11. Zunic A, Corcoran P, Spasic I. Sentiment analysis in health and well-being: systematic review. *JMIR Med Inform.* (2020) 8:e16023. doi: 10.2196/16023
12. Lopez-Castroman J, Moulahi B, Aze J, Bringay S, Deninotti J, Guillaume S, et al. Mining social networks to improve suicide prevention: a scoping review. *J Neurosci Res.* (2019) 98:616–25. doi: 10.1002/jnr.24404
13. Ru B, Yao L. A literature review of social media-based data mining for health outcomes research. In: Bian J, Guo Y, He Z, Hu X, editors. *Social Web and Health Research: Benefits, Limitations, and Best Practices.* Cham: Springer International Publishing (2019). p. 1–14.
14. Altman DG. The scandal of poor medical research. *BMJ.* (1994) 308:283–4. doi: 10.1136/bmj.308.6924.283
15. Ioannidis JPA. Why most clinical research is not useful. *PLoS Med.* (2016) 13:e1002049. doi: 10.1371/journal.pmed.1002049
16. Greenhalgh T, Snow R, Ryan S, Rees S, Salisbury H. Six 'biases' against patients and carers in evidence-based medicine. *BMC Med.* (2015) 13:200. doi: 10.1186/s12916-015-0437-x
17. Mader LB, Harris T, Kläger S, Wilkinson IB, Hiemstra TF. Inverting the patient involvement paradigm: defining patient led research. *Res Involv Engag.* (2018) 4:21. doi: 10.1186/s40900-018-0104-4
18. Bensing J. Bridging the gap the separate worlds of evidence-based medicine and patient-centered medicine. *Patient Educ Couns.* (2000) 39:17–25. doi: 10.1016/S0738-3991(99)00087-7
19. Alvaro N, Miyao Y, Collier N. TwiMed: twitter and pubmed comparable corpus of drugs, diseases, symptoms, and their relations. *JMIR Public Health Surveillance.* (2017) 3:e24. doi: 10.2196/publichealth.6396
20. James Lind Alliance. *About the James Lind Alliance.* James Lind Alliance (2019). Available online at: http://www.jla.nihr.ac.uk/about-the-james-lind-alliance/ (accessed May 29, 2019).
21. Kones R, Rumana U, Merino J. Exclusion of 'nonRCT evidence' in guidelines for chronic diseases - is it always appropriate? The look AHEAD study. *Curr Med Res Opin.* (2014) 30:2009–19. doi: 10.1185/03007995.2014.925438
22. Abbasi A, Adjeroh D, Dredze M, Paul MJ, Zahedi FM, Zhao H, et al. Social media analytics for smart health. *IEEE Intell Syst.* (2014) 29:60–4. doi: 10.1109/MIS.2014.29
23. Powell J, Boylan A-M, Greaves F. Harnessing patient feedback data: a challenge for policy and service improvement. *Digit Health.* (2015) 1:2055207615617910. doi: 10.1177/2055207615617910
24. Brooker P, Barnett J, Cribbin T. Doing social media analytics. *Big Data Soc.* (2016) 3:2053951716658060. doi: 10.1177/2053951716658060
25. Smith J, Bartlett J, Buck D, Honeyman M. *Online Support: Investigating the Role of Public Online Forums in Mental Health.* DEMOS/The Kings Fund (2017). Available online at: https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/Online_Support_Investigating_role_public_online_forums_mental_health.pdf (accessed October 31, 2017).
26. Gonzalez-Hernandez G, Sarker A, O'Connor K, Savova G. Capturing the patient's perspective: a review of advances in natural language processing of health-related text. *Yearb Med Inform.* (2017) 26:214–27. doi: 10.1055/s-0037-1606506
27. Dol J, Tutelman PR, Chambers CT, Barwick M, Drake EK, Parker JA, et al. Health researchers' use of social media: scoping review. *J Med Internet Res.* (2019) 21:e13687. doi: 10.2196/13687
28. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. *Am J Public Health.* (2017) 107:e1–8. doi: 10.2105/AJPH.2016.303512
29. Fox S. *The Social Life of Health Information.* Pew Research Center (2014). Available online at: http://www.pewresearch.org/fact-tank/2014/01/15/the-social-life-of-health-information/ (accessed August 12, 2017).
30. Ravoire S, Lang M, Perrin E, Participants of Giens XXXII Round Table No. 6. Advantages and limitations of online communities of patients for research on health products. *Therapie.* (2017) 72:135–43. doi: 10.1016/j.therap.2016.11.058
31. Martinez B, Dailey F, Almario CV, Keller MS, Desai M, Dupuy T, et al. Patient understanding of the risks and benefits of biologic therapies in inflammatory bowel disease: insights from a large-scale analysis of social media platforms. *Inflamm Bowel Dis.* (2017) 23:1057–64. doi: 10.1097/MIB.0000000000001110
32. Wilson V. Patient use of the internet for diabetes information. *Nurs Times.* (2013) 109:18–20.
33. Sarker A, Ginn R, Nikfarjam A, O'Connor K, Smith K, Jayaraman S, et al. Utilizing social media data for pharmacovigilance: a review. *J Biomed Inform.* (2015) 54:202–12. doi: 10.1016/j.jbi.2015.02.004
34. Tricco AC, Zarin W, Lillie E, Jeblee S, Warren R, Khan PA, et al. Utility of social media and crowd-intelligence data for pharmacovigilance: a scoping review. *BMC Med Inform Decis Mak.* (2018) 18:38. doi: 10.1186/s12911-018-0621-y
35. Kalf RR, Makady A, Ten Ham RM, Meijboom K, Goettsch WG, IMI-GetReal Workpackage 1. Use of social media in the assessment of relative effectiveness: explorative review with examples from oncology. *JMIR Cancer.* (2018) 4:e11. doi: 10.2196/cancer.7952
36. Risson V, Saini D, Bonzani I, Huisman A, Olson M. Patterns of treatment switching in multiple sclerosis therapies in US patients active on social media: application of social media content analysis to health outcomes research. *J Med Internet Res.* (2016) 18:e62. doi: 10.2196/jmir.5409
37. Al-garadi MA, Khan MS, Varathan KD, Mujtaba G, Al-Kabsi AM. Using online social networks to track a pandemic: a systematic review. *J Biomed Inform.* (2016) 62:1–11. doi: 10.1016/j.jbi.2016.05.005
38. Fung ICH, Duke CH, Finch KC, Snook KR, Tseng PL, Hernandez AC, et al. Ebola virus disease and social media: a systematic review. *Am J Infect Control.* (2016) 44:1660–71. doi: 10.1016/j.ajic.2016.05.011

39. www.medicines.org.uk. *Modafinil Provigil 200mg Tablets - Summary of Product Characteristics (SPC)*. medicines.org.uk. electronic Medicines Compendium - eMC (2016). Available online at: http://www.webcitation.org/query?url=https%3A%2F%2Fwww.medicines.org.uk%2Femc%2Fproduct%2F4320%2Fsmpc&date=2019-05-29 (accessed November 21, 2016).
40. Joint Formulary Committee. *Modafinil. British National Formulary*. NICE (2017). Available online at: https://bnf.nice.org.uk/drug/modafinil.html (accessed July 4, 2017).
41. Medicines, Healthcare products Regulatory Agency. *Modafinil: Restricted Use Recommended*. GOVUK. GOV.UK (2014). Available online at: https://www.gov.uk/drug-safety-update/modafinil-restricted-use-recommended (accessed July 7, 2014).
42. Frost J, Okun S, Vaughan T, Heywood J, Wicks P. Patient-reported outcomes as a source of evidence in off-label prescribing: analysis of data from PatientsLikeMe. *J Med Internet Res*. (2011) 13:e6. doi: 10.2196/jmir.1643
43. Gajewski M, Weinhouse G. The use of modafinil in the intensive care unit. *J Intensive Care Med*. (2016) 31:142–5. doi: 10.1177/0885066615571899
44. Castells X, Ramos-Quiroga JA, Bosch R, Nogueira M, Casas M. Amphetamines for Attention Deficit Hyperactivity Disorder (ADHD) in adults. *Cochrane Database Syst Rev*. (2011) CD007813. doi: 10.1002/14651858.CD007813.pub2
45. Möller F, Poettgen J, Broemel F, Neuhaus A, Daumer M, Heesen C. HAGIL (Hamburg Vigil Study): a randomized placebo-controlled double-blind study with modafinil for treatment of fatigue in patients with multiple sclerosis. *Mult Scler*. (2011) 17:1002–9. doi: 10.1177/1352458511402410
46. Stankoff B, Waubant E, Confavreux C, Edan G, Debouverie M, Rumbach L, et al. Modafinil for fatigue in MS: a randomized placebo-controlled double-blind study. *Neurology*. (2005) 64:1139–43. doi: 10.1212/01.WNL.0000158272.27070.6A
47. Tuken M, Kiremit MC, Serefoglu EC. On-demand modafinil improves ejaculation time and patient-reported outcomes in men with lifelong premature ejaculation. *Urology*. (2016) 94:139–42. doi: 10.1016/j.urology.2016.04.036
48. Goss AJ, Kaser M, Costafreda SG, Sahakian BJ, Fu CH, Goss AJ, et al. Modafinil augmentation therapy in unipolar and bipolar depression: a systematic review and meta-analysis of randomized controlled trials. *J Clin Psychiatry*. (2013) 74:1101–7. doi: 10.4088/JCP.13r08560
49. Ondo WG, Fayle R, Atassi F, Jankovic J. Modafinil for daytime somnolence in Parkinson's disease: double blind, placebo controlled parallel trial. *J Neurol Neurosurg Psychiatry*. (2005) 76:1636–9. doi: 10.1136/jnnp.2005.065870
50. Cooper MR, Bird HM, Steinberg M, Cooper MR, Bird HM, Steinberg M. Efficacy and safety of modafinil in the treatment of cancer-related fatigue. *Ann Pharmacother*. (2009) 43:721–5. doi: 10.1345/aph.1L532
51. Spathis A, Fife K, Blackhall F, Dutton S, Bahadori R, Wharton R, et al. Modafinil for the treatment of fatigue in lung cancer: results of a placebo-controlled, double-blind, randomized trial. *J Clin Oncol*. (2014) 32:1882–8. doi: 10.1200/JCO.2013.54.4346
52. Dougall D, Poole N, Agrawal N. Pharmacotherapy for chronic cognitive impairment in traumatic brain injury. *Cochrane Database Syst Rev*. (2015) CD009221. doi: 10.1002/14651858.CD009221.pub2
53. Castells X, Cunill R, Pérez-Mañá C, Vidal X, Capellà D. Psychostimulant drugs for cocaine dependence. *Cochrane Database Syst Rev*. (2016) 9:CD007380. doi: 10.1002/14651858.CD007380.pub4
54. Day J, Zienius K, Gehring K, Grosshans D, Taphoorn M, Grant R, et al. Interventions for preventing and ameliorating cognitive deficits in adults treated with cranial irradiation. *Cochrane Database Syst Rev*. (2014) 2014:CD011335. doi: 10.1002/14651858.CD011335
55. Elbers RG, Verhoef J, van Wegen EEH, Berendse HW, Kwakkel G. Interventions for fatigue in Parkinson's disease. *Cochrane Database Syst Rev*. (2015) 2015:CD010925. doi: 10.1002/14651858.CD010925.pub2
56. Peuckmann V, Elsner F, Krumm N, Trottenberg P, Radbruch L. Pharmacological treatments for fatigue associated with palliative care. *Cochrane Database Syst Rev*. (2010) CD006788. doi: 10.1002/14651858.cd006788.pub2
57. Koopman FS, Beelen A, Gilhus NE, de Visser M, Nollet F. Treatment for postpolio syndrome. *Cochrane Database Syst Rev*. (2015) CD007818. doi: 10.1002/14651858.CD007818.pub3
58. Liira J, Verbeek JH, Costa G, Driscoll TR, Sallinen M, Isotalo LK, et al. Pharmacological interventions for sleepiness and sleep disturbances caused by shift work. *Cochrane Database Syst Rev*. (2014) CD009776. doi: 10.1002/14651858.CD009776.pub2
59. Lyon J. Chess study revives debate over cognition-enhancing drugs. *JAMA*. (2017) 318:784–6. doi: 10.1001/jama.2017.8114
60. Frost JH. The case for using social media to aggregate patient experiences with off-label prescriptions. *Expert Rev Pharmacoecon*. (2011) 11:371–3. doi: 10.1586/erp.11.43
61. Braun V, Clarke V. Using thematic analysis in psychology. *Qual Res Psychol*. (2006) 3:77–101. doi: 10.1191/1478088706qp063oa
62. *Ask a Patient - Drug Reviews by Patients*. AskAPatient (2016). Available online at: https://www.askapatient.com/ (accessed December 10, 2016).
63. *Modafinil Reviews & Ratings at Drugs.com*. (2017). Available online at: https://www.drugs.com/comments/modafinil/ (accessed January 16, 2017).
64. *Modafinil Oral Reviews and User Ratings: Effectiveness, Ease of Use, and Satisfaction*. WebMD (2017). Available online at: https://www.webmd.com/drugs/drugreview-16962-modafinil-oral.aspx?drugid=16962&drugname=modafinil-oral (accessed January 16, 2017).
65. *Erowid Experience Vaults: Modafinil Reports in category: Medical Use*. Erowid (2016). Available online at: https://www.erowid.org/experiences/subs/exp_Modafinil_Medical_Use.shtml (accessed September 30, 2016).
66. Chiauzzi E, Wicks P. Digital trespass: ethical and terms-of-use violations by researchers accessing data from an online patient community. *J Med Internet*. (2019) 21:e11985. doi: 10.2196/11985
67. Nicholas J, Onie S, Larsen ME. Ethics and privacy in social media research for mental health. *Curr Psychiatry Rep*. (2020) 22:84. doi: 10.1007/s11920-020-01205-9
68. Markham A, Buchanan E. *Ethical Decision Making and Internet Research 2.0*. AOIR (2012).
69. The British Psychological Society. *Ethics Guidelines for Internet-Mediated Research*. Report No. INF206. British Psychological Society (2017). Available online at: https://www.bps.org.uk/sites/bps.org.uk/files/Policy/Policy%20-%20Files/Ethics%20Guidelines%20for%20Internet-mediated%20Research%20%282017%29.pdf
70. Cooper AK, Coetzee S. On the ethics of using publicly-available data. In: Hattingh M, Matthee M, Smuts H, Pappas I, Dwivedi Y, and Mäntymäki M. editors. *Responsible Design, Implementation and Use of Information and Communication Technology*. Cham: Springer International Publishing (2020). p. 159–71. doi: 10.1007/978-3-030-45002-1_14
71. *What Is Open Data?* Open Knowledge Foundation (2020). Available online at: https://opendatahandbook.org/guide/en/what-is-open-data/ (accessed January 24, 2021).
72. Taylor J, Pagliari C. Mining social media data: how are research sponsors and researchers addressing the ethical challenges? *Res Ethics*. (2018) 14:1747016117738559. doi: 10.1177/1747016117738559
73. Golder S, Scantlebury A, Christmas H. Understanding public attitudes toward researchers using social media for detecting and monitoring adverse events data: multi methods study. *J Med Internet Res*. (2019) 21:e7081. doi: 10.2196/jmir.7081
74. Golder S, Ahmed S, Norman G, Booth A. Attitudes toward the ethics of research using social media: a systematic review. *J Med Internet Res*. (2017) 19:e195. doi: 10.2196/jmir.7082
75. Beninger K. Social media users' views on the ethics of social media research. In: *The SAGE Handbook of Social Media Research Methods*. Sage (2017). p. 57–73. Available online at: https://books.google.co.uk/books?hl=en&lr=&id=9oewDQAAQBAJ&oi=fnd&pg=PA57&dq=social$+$media$+$research$+$ethics&ots=eOIN34xWAP&sig=j8mbM9L9dT-mWgc-iDn4rAFFOE0
76. Cumberledge J. First Do No Harm96 The report of the Independent Medicines and Medical Devices Safety Review [Internet]. In: *The Independent Medicine & Medical Devices Safety Review*. (2020). Available online at: https://www.immdsreview.org.uk/Report.html
77. Garchitorena M. *Meet the Women Who Forced the FDA to Listen About this Controversial Birth Control Device*. (2016). Available online at: http://www.refinery29.uk/2016/04/107550/essure-problems-fda-lawsuit (accessed May 22, 2017).
78. U.S. Copyright Office. *More Information on Fair Use*. copyright.gov (2020). Available online at: https://www.copyright.gov/fair-use/more-info.html (accessed January 14, 2021).
79. British Library. *What Is Fair Use? Fair Dealing Copyright Explained British Library*. The British Library (2020). Available online at: https://www.bl.

uk/business-and-ip-centre/articles/fair-use-copyright-explained (accessed January 14, 2021).
80. *Privacy Policy*. drugs.com (2020). Available online at: https://www.drugs.com/support/privacy.html (accessed January 14, 2021).
81. *WebMD Privacy Policy*. WebMD (2020). Available online at: https://www.webmd.com/about-webmd-policies/about-privacy-policy (accessed January 14, 2021).
82. VERBI Software. *MAXQDA*. (2018). Available online at: www.maxqda.com (accessed December 21, 2017).
83. Neeleman A, van de Koot H. *The Linguistic Expression of Causation. The Theta System*. Oxford: Oxford University Press (2012).
84. McAndrew TC, Bongard JC, Danforth CM, Dodds PS, Hines PDH, Bagrow JP. What we write about when we write about causality: features of causal statements across large-scale social discourse. In: *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining Piscataway*. Piscataway, NJ: IEEE Press (2016). p. 519–24.
85. Kilgarriff A, Baisa V, Bušta J, Jakubíček M, Kovar V, Michelfeit J, et al. *SketchEngine*. Sketch Engine (2020). Available online at: https://www.sketchengine.eu/ (accessed May 21, 2020).
86. Lee J. *Writing Linguistic Rules for Natural Language Processing Medium*. Towards Data Science (2019). Available online at: https://towardsdatascience.com/linguistic-rule-writing-for-nlp-ml-64d9af824ee8 (accessed April 28, 2020).
87. SketchEngine.eu. *English Penn Treebank Tag Set With Modifications*. SketchEngine.eu (2018). Available online at: https://www.sketchengine.eu/english-treetagger-pipeline-2/ (accessed April 20, 2018).
88. Kilgarriff A. *Getting to Know Your Corpus. Text, Speech and Dialogue*. Berlin: Springer (2012).
89. Manning C, Raghavan P, Schütze H. Introduction to information retrieval. *Natl Lang Eng*. (2010) 16:100–3. doi: 10.1017/S1351324909005129
90. TenTen Corpus Family | Sketch Engine. *Sketch Engine*. (2015). Available online at: https://www.sketchengine.eu/documentation/tenten-corpora/ (accessed May 21, 2020).
91. Sateia MJ. International classification of sleep disorders-third edition: highlights and modifications. Chest. (2014) 146:1387–94. doi: 10.1378/chest.14-0970
92. Sinclair J. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press (1991).
93. Loria S. *TextBlob*. (2020). Available online at: https://textblob.readthedocs.io/en/dev/index.html (accessed May 21, 2020).
94. Joint Formulary Committee, editor. *BNF 75: March - September 2018: Modafinil*. London: Pharmaceutical Press (2018).
95. Baumer EPS, Mimno D, Guha S, Quan E, Gay GK. Comparing grounded theory and topic modeling: extreme divergence or unlikely convergence? *J Assoc Inform Sci Tech*. (2017) 68:1397–410. doi: 10.1002/asi.23786
96. Schlegl E, Ducournau P, Ruof J. Different weights of the evidence-based medicine triad in regulatory, health technology assessment, and clinical decision making. *Pharmaceut Med*. (2017) 31:213–6. doi: 10.1007/s40290-017-0197-3
97. Sheridan DJ, Julian DG. Achievements and limitations of evidence-based medicine. *J Am Coll Cardiol*. (2016) 68:204–13. doi: 10.1016/j.jacc.2016.03.600
98. Chapman S. *The Problem With Sex: Is Our Reluctance to Talk About it Harming Patients?* Evidently Cochrane (2017). Available online at: http://www.evidentlycochrane.net/problem-with-sex-reluctance-talk-harming-patients/ (accessed April 6, 2017).
99. Shaffer JA, Falzon L, Cheung K, Davidson KW. N-of-1 randomized trials for psychological and health behavior outcomes: a systematic review protocol. *Syst Rev*. (2015) 4:87. doi: 10.1186/s13643-015-0071-x
100. Battleday RM, Brem AK, Battleday RM, Brem A-K. Modafinil for cognitive neuroenhancement in healthy non-sleep-deprived subjects: a systematic review. *Eur Neuropsychopharmacol*. (2015) 25:1865–81. doi: 10.1016/j.euroneuro.2015.07.028
101. Ballon JS, Feifel D, Ballon JS, Feifel D. A systematic review of modafinil: potential clinical uses and mechanisms of action. *J Clin Psychiatry*. (2006) 67:554–66. doi: 10.4088/JCP.v67n0406
102. Sheng P, Hou L, Wang X, Wang X, Huang C, Yu M, et al. Efficacy of modafinil on fatigue and excessive daytime sleepiness associated with neurological disorders: a systematic review and meta-analysis. *PLoS ONE*. (2013) 8:e81802. doi: 10.1371/journal.pone.0081802
103. Segura-Bedmar I, Martínez P. Pharmacovigilance through the development of text mining and natural language processing techniques. *J Biomed Inform*. (2015) 58:288–91. doi: 10.1016/j.jbi.2015.11.001
104. Yang M, Kiang M, Shang W. Filtering big data from social media – building an early warning system for adverse drug reactions. *J Biomed Inform*. (2015) 54:230–40. doi: 10.1016/j.jbi.2015.01.011
105. Bollegala D, Maskell S, Sloane R, Hajne J, Pirmohamed M. Causality patterns for detecting adverse drug reactions from social media: text mining approach. *JMIR Public Health*. (2018) 4:e51. doi: 10.2196/publichealth.8214
106. Mayer-Schönberger V, Cukier K. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. London: Houghton Mifflin Harcour (2013).
107. Prainsack B. Through thick and big: data-rich medicine in the era of personalisation. In: Vollman J, Sandow V, Wäscher S, Schildmann J, editors. *The Ethics of Personalised Medicine*. Farnham: Ashgate (2015). p. 161–72. Available online at: http://www.webcitation.org/query?url=https%3A%2F%2Fs3.amazonaws.com%2Facademia.edu.documents%2F37568485%2FPrainsack_in_Vollmann_et_al._2015.pdf%3FAWSAccessKeyId%3DAKIAIWOWYYGZ2Y53UL3A%26Expires%3D1559133050%26Signature%3DdENibbTnoMSCr37wyROKr0Y9MTI%253D%26response-content-disposition%3Dinline%253B%2520filename%253DThrough_Thick&date=2019-05-29
108. Sim I. Two ways of knowing: big data and evidence-based medicine. *Ann Intern Med*. (2016) 164:562–63. doi: 10.7326/M15-2970
109. Edwards R. *Living With Complexity and Big Data*. Report No. 78. Uppsala Monitoring Centre (2018). p. 36. Available online at: https://view.publitas.com/uppsala-monitoring-centre/uppsala-reports-78/page/28 (accessed November 06, 2018).
110. Anjum RL. What is the guidelines challenge? The cause health perspective. *J Eval Clin Pract*. (2018) 24:1127–31. doi: 10.1111/jep.12950
111. Colagiuri B, Schenk LA, Kessler MD, Dorsey SG, Colloca L. The placebo effect: from concepts to genes. *Neuroscience*. (2015) 307:171–90. doi: 10.1016/j.neuroscience.2015.08.017
112. Beecher HK. The powerful placebo. *J Am Med Assoc*. (1955) 159:1602–6. doi: 10.1001/jama.1955.02960340022006
113. Kienle GS, Kiene H. The powerful placebo effect: fact or fiction? *J Clin Epidemiol*. (1997) 50:1311–8. doi: 10.1016/S0895-4356(97)00203-5
114. Kramer ADI, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. *Proc Natl Acad Sci USA*. (2014) 111:8788–90. doi: 10.1073/pnas.1320040111
115. Coviello L, Sohn Y, Kramer ADI, Marlow C, Franceschetti M, Christakis NA, et al. Detecting emotional contagion in massive social networks. *PLoS ONE*. (2014) 9:e90315. doi: 10.1371/journal.pone.0090315
116. Yan L (lucy), Tan Y. The consensus effect on shared treatment experience in online healthcare communities. *J Manage Inform Syst*. (2016) 34:11–39. doi: 10.2139/ssrn.2603042
117. Rastegar-Mojarad M, Ye Z, Wall D, Murali N, Lin S. Collecting and analyzing patient experiences of health care from social media. *JMIR Res Protoc*. (2015) 4:e78. doi: 10.2196/resprot.3433
118. Dirkson A, Verberne S, Sarker A, Kraaij W. Data-Driven lexical normalization for medical social media. *Multimodal Technol Interact*. (2019) 3:60. doi: 10.3390/mti3030060

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.