

UNDERSTANDING THE PLIGHT OF COVID-19 LONG HAULERS THROUGH
COMPUTATIONAL ANALYSIS OF YOUTUBE CONTENT

by

Alexis Jordan

A thesis submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Master of Science in
Computer Science

Charlotte

2022

Approved by:

Dr. Albert Park

Dr. Yaorong Ge

Dr. Mohamed Shehab

Dr. Siddharth Krishnan

ABSTRACT

ALEXIS JORDAN. Understanding the plight of Covid-19 Long Haulers through Computational Analysis of YouTube Content. (Under the direction of DR. ALBERT PARK)

The coronavirus pandemic is a pervasive event that has had a drastic impact on countries globally. In the United States alone there have been over 98 million Covid-19 cases and over 1 million deaths. One consequence of Covid-19 infection has been PASC or Coronavirus Post-Acute Sequelae. People with this syndrome, colloquially called Long Haulers experience a wide range of symptoms, up to 22, that impact every system of the body. The root cause of PASC as well as treatment continues to mystify doctors. Many of those suffering from PASC (otherwise called Long Covid) have used social media to turn to each other for support and guidance. Social media surveillance is the use of social media data to be used as a rich source of information on health-related issues and attitudes (Melton). Due to the pervasiveness of social media apps and the comfort with which people engage in health-related discourse online, researchers have employed text mining techniques in order to gain insights about patient experiences.

In this study, I wanted to gain a better understanding of Long Hauler experiences as well as how information about Long Haulers is received. I chose YouTube videos as the data source due to YouTube's unique nature. YouTube is a platform in which creators can make and share videos with their audiences. Discourse then continues in the comment section between viewers and the creators of the videos as well. I gathered data from three different types of content creators: Medical Sources, News Sources and Long Haulers. Medical sources included content creators who identified as doctors, health insurance companies and medical schools. News sources included content from various types of news stations ranging from local to international. Lastly, Long Haulers represented first person accounts of those suffering from PASC themselves. I used Biterm, a topic model created specifically for short texts, to analyze video transcripts and all top-level comments in the comment section. Ultimately, I organized resulting topics into 20 themes across all of the sources. These themes included: Explanations in Layman's Terms, Show Housekeeping,

Biological Explanations, Sharing Patient Experiences, Negative Experiences, Experts Weighing In, Handling the Long Haul, Taking Treatment into Own Hands, Changes to Daily Life, Choose Homeopathy over Pharmaceuticals, Ingesting Shared Information, Seeking More from Shared Content, Misinformation, Skepticism, Sharing Long Covid Experiences, Complete Distrust of Information, Fears of Hidden Dishonest, Nihilistic Entity, Interacting with Content Creators, Disillusionment with Traditional Medicine and Distrust with Health Care System. Results of this thesis could help public health agencies, policymakers, organizations and health researchers to understand symptomatology and experiences related to Long Covid. It also helps these agencies to understand how information concerning Long Covid is being received.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Dr. Albert Park, and Dr. Yaorong Ge for continued help, guidance, and advice throughout this process. It has been a pleasure to work with you all. I genuinely appreciate your encouragement throughout the years.

I would also like to thank my committee members, Dr. Mohamed Shehab and Dr. Siddharth Krishnan for your time and suggestions that have guided me to complete my thesis. Special thanks goes to Dr. Jing Yang for offering support to students in our graduate program who are passionate about research. This is extended to Ms. Carmalita Govan who has also provided guidance throughout this academic process.

DEDICATION

I would like to thank my friends and family for their continued support during my studies. You all provided me with encouragement to keep learning, growing, and pushing forward.

TABLE OF CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	x
LIST OF ABBREVIATIONS	xi
CHAPTER 1: Introduction	1
CHAPTER 2: Theoretical and Conceptual Background	4
2.1 Social Media Surveillance	4
2.2 Health on YouTube	6
CHAPTER 3: Methods	9
3.1 Social Media Site and Datasets	9
3.2 Topic Modeling	9
3.3 Data Collection	11
3.4 Data Preprocessing	12
3.5 Model Selection	13
3.6 Manual Analysis	15
CHAPTER 4: Results	16
4.1 Captions	16
4.1.1 Medical Source	16
4.1.2 News Source	18
4.1.3 Long Haulers	21
4.2 Comments	25
4.2.1 Medical Source	25
4.2.2 News Source	27
4.2.3 Long Haulers	30
4.3 Additional Analysis	31

CHAPTER 5: Discussion and Implications	33
5.1 Discussion	33
5.2 Public Health Implications	35
5.3 Limitations and Future Research	35
REFERENCES	37
APPENDIX A: Detailed Dataset Description	39

LIST OF TABLES

TABLE 2.1: Dataset Description	12
TABLE 2.2: Pros and Cons of Text Mining Methods	15
TABLE 4.1: Themes and Example Transcripts from Medical Source Videos	16
TABLE 4.2: Themes and Example Transcripts from News Source Videos	18
TABLE 4.3: Themes and Example Transcripts from Long Hauler Videos	22
TABLE 4.4: Themes and Example Comments from Medical Source Videos	25
TABLE 4.5: Themes and Example Comments from News Source Videos	27
TABLE 4.6: Themes and Example Comments from Long Hauler Videos	30
TABLE A.1: Detailed Dataset Description	39

LIST OF FIGURES

Figure 1.1 Sadi Nagamutu on 60 Minutes	1
Figure 2.1 BERTopic Results for Medical Source	14
Figure 2.2 BERTopic Results for News Source	14

LIST OF ABBREVIATIONS

PASC is an acronym for Post-Acute Sequelae of Covid-19

SARS-CoV-2 is an acronym for Severe Acute Respiratory Syndrome Coronavirus 2

ME is an acronym for Myalgic Encephalomyelitis

CFS is an acronym for Chronic Fatigue Syndrome

POTS is an acronym for Postural Orthostatic Tachycardia Syndrome

CHAPTER 1: INTRODUCTION

“It’s like a... like a viral tornado that goes in you and kind of just messes you up,” Sadi Nagamutu says in between labored breaths. “It leaves the rubble that the tornado leaves behind,” Anderson Cooper offers. Nagamutu nods in agreement. “It leaves the damage behind.” [1].

Such begins the account of 44-year-old fitness instructor Nagamutu’s battle with Long Covid [1]. In the comments section of the video, one user writes “I had to pause this video at 2:20. I broke down in tears because I feel like I’m not alone. I have the same thing.” At the time of recording Nagamutu had been struggling with Long Covid for 8 months. By this time, it had completely disrupted her life. She notes that she went from being a trainer, to not being able to lift grocery bags and walk at the same time. Judging by the comments left under the 60 minutes video, Nagamutu is not alone.



Figure 1.1: Sadi Nagamutu on 60 Minutes

The coronavirus pandemic has drastically changed the lives of many. Ever increasing amounts of SARS-CoV-2 hospitalizations and deaths were presented by news broadcasters across the world. Given the large number of fatalities, the consequences of SARS-CoV-2 infection appeared to be clear. However, there is another phenomenon that has gradually received more recognition and reaps major ramifications for COVID-19 survivors. It is the case of those with coronavirus post-acute sequelae (PASC) who experience wide ranges of symptoms, even after their initial infection has long cleared. These patients are colloquially called *Long Haulers* [2].

It is not uncommon to see persistent symptoms in those infected with respiratory infections [2]. An antecedent of the current SARS-CoV-2 outbreak, Severe Acute Respiratory Syndrome (SARS), first appeared in 2002. This was also caused by a coronavirus. In a Hong Kong study of 55 patients, researchers found that some infected with SARS experienced impaired lung function up to two years after they were initially hospitalized [2]. However, counting about 8000 infections globally, the breadth of SARS pales in comparison to current COVID-19 cases.

PASC, also called Long Covid, has affected the quality of life of many, such that the syndrome is now approved for inclusion and protections under the Americans with Disabilities Act in the United States [3]. The CDC recognizes those who have received a negative COVID-19 since initial infection but still have symptoms at least 4 weeks after this negative test, to be Long Haulers [4]. High ratios of those who have been infected with COVID-19 have persisting symptoms that last months after initial infection. In a study on Italian adults ranging from 19 to 84, 125 out of 143 Italian patients experienced COVID-19 related symptoms on average two months after initial infection [2]; however, it is important to note that all the patients had been hospitalized with COVID-19 for an average of two weeks.

There have been a wide range of Long Covid symptoms observed that affect many areas of the body. Previous surveys have found that fatigue was the most common symptom amongst 1500 respondents [1]. Muscle and body aches, shortness of breath and difficulty concentrating were also noted as common symptoms. Many patients also reported impaired memory, brain fog and tachycardia.

There have been several theories as to the origins of Long Covid. Psychiatrist Mady Hornig hypothesizes that SARS-CoV-2 infection may reactivate latent Epstein-Barr virus (EBV) triggering fatigue [2]. According to Hornig, autonomic nervous system dysregulation is a possible explanation for Long Haulers' tachycardia, extreme fatigue and other persistent symptoms. Neurologist Mitchell Miglis suspects that "POTS (Postural Orthostatic Tachycardia Syndrome), ME/CFS and persistent COVID-19 may be different names for the same disorder, and patients' diagnoses depend on their physician's sub-specialty" [2, p. 1383].

Long Covid has real implications in people's everyday lives. In their study, Wisk et al. sought to understand the effects of COVID-19 on Quality of Life. Data was gathered from DKI Jakarta hospitalized COVID-19 patients who had recovered within the past month. They distributed a quality of life questionnaire called WHOQOL-BREF. In WHOQOL-BREF, quality of life is a multidimensional construct [3]. Wisk et al. discovered that 30.2% of respondents had Long Covid. They also found that long haul symptoms affected nearly all domains of Quality of Life as defined in the WHOQOL-BREF [3]. In the QOL survey, the lowest score was in the physical domain and the highest score was in the social domain. They found that patients with Long Covid were twice as likely to have a poor quality of life [3].

As we have seen in the previously mentioned literature, around 30% of those infected with COVID-19 experience long haul symptoms. To date, the total COVID-19 case number in the United States is just over 98.4 million [5]. That means that roughly 29.5 million people could be experiencing long haul symptoms. The studies mentioned earlier demonstrate manual efforts to understand the experience of Long Haulers. Given the limitations of manually performed studies, it is likely that there are some insights that have yet to be uncovered due to the large number of Long Haulers in the United States alone.

I proposed a study that uses computational methods to gather insights via Long Hauler discussions on social media, particularly YouTube. Due to its ubiquitous nature, social media offers access to conversations and content from millions worldwide. As of April 2022, a report from Statista estimates that YouTube has 247 million users, whilst India leads all countries with 467 million users [6]. Social media data has been used as a rich source of health-related information [7]. Long Hauler-related YouTube video comments and video transcripts, also called captions, provided me with additional insight into Long Covid symptomatology and the challenges that Long Haulers face. This approach was also likely quicker and able to capture a wider distribution than manual studies. To my knowledge, this was the first study to examine YouTube video transcripts and comments related to Long Covid experiences.

CHAPTER 2: THEORETICAL AND CONCEPTUAL BACKGROUND

2.1 Social Media Surveillance

In recent history, social media data has been seen as a rich source of information of health-related experiences and attitudes. This is due to not only the pervasiveness of social media apps, but also the freedom with which people engage in discourse on various topics. Topic modeling has been used for several text mining applications in social media content analysis, content characterization, event tracking, and content recommendation [8]. With the amassment of health-related social media data came the opportunity to use natural language processing techniques to conduct surveillance of posts [7].

Mutanga et al. used an LDA to study COVID-19 related tweets in South Africa. At the time, there was no vaccine or viral therapy for the disease. The authors aimed to understand issues that the South African public were experiencing due to the pandemic [11]. From their analysis, they discovered that conversations revolved around alcohol consumption, staying at home, vaccine conspiracy theories, police brutality, statistics tracing, and 5G [11]. They deemed one of the major findings to be people's reluctance to take future vaccines due to conspiracies. They ultimately felt that their findings would be useful to the government and policymakers in managing pandemic outbreaks and addressing public concerns [11].

Thompson et al. discussed the plight of Long Haulers in the context of Uncertainty Management Theory. Uncertainty Management theory encompasses individual's responses to and experience with uncertainty [12]. The previous sections have alluded to the layers of uncertainty that Long Haulers face. To date, Long Covid continues to mystify scientists and medical practitioners. Public discourse and explanations on the topic have been elusive and ever changing. Patients have been experiencing a variety of life altering symptoms without many answers, thus they must manage both their illness and their uncertainty.

With the convenience of social media as a crowdsourcing platform, where users constantly and readily exchange information, Thompson et al. have found that it has become common for individuals to share their symptomatology and diagnoses with their peers. Crowdsourcing is a rapidly growing, low-cost

strategy for openly sharing information that can aid in reducing uncertainty. In such communities, members typically seek and share information regarding diagnosis and the disease spectrum [12]. Past research has focused on the intersection of crowdsourcing and COVID-19. However, such studies have been limited in size.

In their study, Melton et al. explored public sentiment and discourse on coronavirus vaccines on Reddit [7]. In order to achieve this, they used a lexical based sentiment analysis from Gensim in order to uncover public sentiment in the vaccine related posts. After they performed some basic text normalization by removing stopwords and lemmatization, they generated 50 different LDA models. They chose their ideal number of topics based on the coherence scores of the 50 different models [7]. Ultimately, Melton et al found that 56.68% of the posts were positive and 27.69% of the posts were negative, whilst 15.63% were neutral. Their resulting topics covered a broader discussion of vaccine, safety concerns, efficacy, and potential side effects [7]. They also found that results for individual months agreed with the combined analysis. However, the number of topics tended to be smaller than 5 as each month's corpus was smaller than the aggregated dataset.

Trihandini et al. conducted a study of COVID-19 Long Hauler posts on Reddit. They used the Mean Extraction Method to conduct text analysis. This technique identifies topics by finding words that group together mathematically. MEM has been used by researchers in the past to gather insights about communication patterns, relationship dynamics, health behaviors and psychological health [13]. MEM works by first calculating frequencies of words in a body of work. Any low frequency words and stop words are removed. N-grams are divided by total word count to find the percentage of use score. This is fed into a Principal Component Analysis algorithm that the researchers used to extract common word clusters. The 16 topics derived from this process included symptoms, diagnostic concerns, broad health concerns, chronicity, support, identity and anxiety. Symptoms were the largest grouping. These symptoms included: shortness of breath, brain fog, fatigue, loss of smell and taste, cardiac pain, interrupted sleep and blood pressure [13].

In their study, Basch et al. conducted a study to describe the content of news coverage of COVID-19 long haul symptoms posted on the internet. The first 100 English language articles and/or videos were selected from Google News. Google News is a news aggregator service offered by Google. The content from the videos and news articles were coded for information on COVID-19 long haul symptoms [14]. They found that 41% of news reports mentioned the duration of the symptoms. The duration tended to range between 1 month to over a year. Tiredness and fatigue were the most mentioned symptom, occurring in 74% of the news content. Next was difficulty breathing and shortness of breath, and difficulty thinking or concentrating, which occurred in 62% of the observed news content and 50% respectively. Basch et al. found no statistically significant differences between the types of sources observed: be it video, written news reports or some combination of the two [14].

2.2 Health on YouTube

Like many other social media, YouTube is a social media platform that motivates users to create, publish and comment on posts. YouTube is unique in that the majority of the content on the platform is videos which range from 60 seconds to several hours long. Essentially, they provide creators with a unique mode of communication. In addition to watching content, creators and viewers can engage in discourse within the comment section on a variety of subjects. There have been several studies in which researchers analyzed YouTube comments in order to understand public sentiment on various health-related matters. These studies have used either manual or NLP-based approaches.

For example, in their 2021 study McLellan et al. thematically coded the comments of YouTube videos that were meant to de-stigmatize mental illness [15]. They collected 12,842 comments from 20 YouTube videos on the subject, however, only 985 videos were randomly selected for coding [15]. In the end, they uncovered five central themes which were: community building, personal experiences of mental illnesses, personal experiences of stigma, debates on mental illness validity, and providing explanations for mental illness. In all, they discovered both positive and negative outcomes related to the viewing of these videos [15].

Similarly, Syed-Abdul et al. employed the help of three physicians to investigate the nature of misinformation dissemination in anorexia-related videos. The physicians first reviewed 140 videos that were anorexia related [16]. These videos were then divided into two categories: informative and pro-anorexia. Informative videos were those that offered insight on the consequences of anorexia as well as strategies for recovery. In contrast, pro-anorexia videos promoted anorexia as a beauty standard and encouraged viewers to engage by sharing methods with which they could become anorexic [16]. Ultimately, they found that while the pro-anorexia videos were less common, they were favored three times more than anorexic videos [16].

In contrast, Oyeboode et al. employed NLP techniques to analyze social media comments [17]. It should be noted that their data source was not limited to solely YouTube. In fact, they collected over 47 million COVID-19 related comments from Twitter, Facebook and three online discussion forums as well [17]. Instead of manual coding, the authors employed NLP techniques to extract meaningful phrases from over 1 million selected comments. As part of their discovery of thematic topics, they assigned sentiment polarity to comments based on lexicon-techniques [17]. They ended up grouping negative and positive key phrases into broader themes.

Overarching negative themes that they uncovered were: health related issues, psychosocial issues and social issues. Examples of health related issues included: increased mortality, health concerns, struggling health systems and fitness issues. Examples of psychosocial issues included: frustrations due to life disruptions, panic shopping and expression of fear [17]. Social issues included harassment, domestic violence and wrong societal attitudes. Positive themes that they uncovered included: public awareness, encouragement, gratitude, cleaner environment, online learning, charity, spiritual support and innovative research [16].

Jacques et al. studied COVID-19 related content from YouTube [18]. Their dataset included 100 of the most viewed videos that were identified with the keywords “Covid-19 long haul”. They assessed video origin, engagement, and content related to COVID-19 long-haul. The majority of the videos in their dataset were uploaded by news outlets. This was followed by consumers, health professionals and lastly

entertainment television. Metadata for each video was studied. Prior to analysis, the authors generated a list of content categories which were: the length of time symptoms persisted, tiredness and/or fatigue, brain fog or memory loss, ear ringing, sleep disturbance, stroke, fear and/or worry, hair loss, headache/sinus pain, loss of smell or taste, dizziness, heart palpitations, chest pain, difficulty breathing and/or shortness of breath, lack of mobility, joint or muscle pain, depression or anxiety, fever, diarrhea, symptoms getting worse after physical or mental activities, multisystem inflammatory syndrome, post-intensive care syndrome (PICS), post-traumatic stress disorder, difficulty getting help, available treatment, symptoms being worse in women, symptoms being worse in older populations, job loss/inability to work, other related life issues related to long haul symptoms, waves of symptoms, relief after vaccine, other symptoms [18].

Jacques et al. used IBM SPSS for statistical analysis. In their significance test, $p < 0.05$ was used to determine statistical significance. The number of likes and length of YouTube videos were analyzed using one-way ANOVA [18]. The most commonly mentioned symptoms in Jacques et al.'s study were fatigue, difficulty breathing, joint or muscle pain and difficulty thinking or concentrating [18].

CHAPTER 3: METHODS

3.1 Social Media Site and Datasets

YouTube is a ubiquitous social media platform that has been adopted by users worldwide. It is completely free to post videos to the site. As such, many have access to upload, share and participate in user generated content. Additionally, YouTube has been adopted by organizations and professionals from various fields as a means to share relevant and important information. News stations, ranging from local to international, regularly post to YouTube. Professionals such as doctors, lawyers and architects also openly share information regarding their respective professions and research on YouTube as well. Because of this, I deemed YouTube to be an adequate source from which to gather videos about Long Covid. I sought to gather information from a range of sources to use in my analysis.

3.2 Topic Modeling

A variety of Natural Language Processing techniques can be used to help researchers and practitioners gather insights from written, often conversational data. One such popular method is Topic Modeling. Topic Modeling is the discovery of thematic relationships and patterns within a body of text, often called a corpus, using natural language models.

Topic models have been used to uncover the latent semantic structure in a text corpus [8]. The first such effort of this task is latent semantic analysis (LSA). LSA uses singular value decomposition in the document-term matrix. This in turn reveals associative word patterns [8]. There is a probabilistic version of LSA that is an improvement on the original LSA. Probabilistic Latent Semantic Analysis (PLSA) represents documents as a mixture of topics and represents topics as a probability distribution over words. Latent Dirichlet Allocation (LDA) extends PLSA in that it adds Dirichlet priors to the topic distributions [7]. There are various models that can be used for Topic Modeling, however, Latent Dirichlet Allocation (LDA) remains a popular choice [7].

LDA models utilize a probabilistic unsupervised classification technique [8]. The model uses an iterative approach to discover meaningful topics within a text. This approach is based on evaluating the co-

occurrence of words in documents and represents each document as a mix of topics with related weights. The basic idea of LDA models is that individual topics will feature similar diction and will display a sparse Dirichlet distribution [7]. In LDAs, the number of topics is first defined by the user. Then, each word will be randomly assigned to a topic. The algorithm calculates the proportion of words assigned to a document: $P(\text{topic } T \mid \text{document } D)$ [7]. Then it calculates the proportion of words assigned to a topic: $P(\text{word } W \mid \text{topic } T)$. For each topic, the product of these two proportions is found. Then, the products are compared to each other until the algorithm converges [7].

LDA has been widely used in studies with a large set of documents. For example, in their study, Jelodaer et al. used an LDA to extract semantic topics from YouTube user comments [9]. The scope of their study was to understand user sentiment and discourse with the ultimate goal of enhancing the quality of videos shown and meeting the needs of the viewers. They defined their set of documents as YouTube comments and the words within the comments were the topics. The discrete topic distributions were drawn from a symmetric Dirichlet distribution.

Yan et al. proposed an alternative for LDA topic models that was developed specifically for shorter bodies of text [8]. They argued that it may be difficult to apply conventional topic models like LDA and PLSA to shorter bodies of texts, like tweets or instant messages, because they suffer from data sparsity. Their proposed method, the Biterm Topic Model, learns topics by modeling the generation of word co-occurrence patterns in the whole corpus [8]. They propose that this method helps to counter the sparse word co-occurrence pattern problem that occurs when evaluating at the document level.

Yan et al. found that their model generated more coherent and prominent topics when compared in an experiment with three conventional topic modeling methods: LDA, LDA-U and Mixture of Unigrams [9]. They also found that Biterm outperformed the LDA on normal-length texts, which suggests that Biterm Topic Models (BTM) can be used in a broader range of scenarios.

BTM models word-occurrence patterns based on biterms. A biterm is an unordered word-pair co-occurring in a window of text that contains meaningful word co-occurrences. Each document is considered to be an individual context unit. Any two distinct words in a short text document is considered to be a biterm

[8]. All biterns in a corpus are aggregated for topic learning [8]. This allows BTM to leverage global word co-occurrence patterns as well to uncover latent topics [8].

In contrast to traditional LDA, BERTopic ditches the bag of words representation of topics because they disregard semantic relationships between words [19]. Instead, it uses embeddings to generate contextual word and sentence vector representations. Documents are embedded using pre-trained, transformer-based language models. These embeddings are then clustered and then topic representations are created using TF-IDF [19].

3.3 Data Collection

I used Google's API (`googleapiclient.discovery`) to capture video comments and descriptive information. Data from the top 50 videos were collected as a result of searching each of the following search terms: "Covid Long haulers", "Covid-19 Long haulers", "Long Covid", "Long Haul Covid", "PASC Covid", "Post-Covid Symptoms" and "Post-Covid Syndrome." I used the python package *YouTubeTranscriptAPI* to capture video transcripts from the videos related to the comments produced in my first search. The video ID was used to achieve this. Resulting videos were in the date range August 3, 2020 and October 29, 2021. After removal of duplicates and irrelevant videos, I collected 152 videos total. Additional information regarding these videos, such as number of comments, views and likes can be found in Appendix A. It should be noted that the comments collected in my data gathering process only reflect the top-level comments. In essence, this means that any replies to the original comment are not captured.

After this process, I arranged the videos into subgroups based on the video source. In this case, source refers to the content creator who posted the video. The natural groupings were: News Source, Medical Source and Long Haulers. News Source videos were those that were uploaded by news entities. These entities ranged from local to national to international news stations. News Source videos represented 78 of the collected videos. Medical Source videos were those that were posted by medical experts such as doctors, health insurance companies and medical schools. I collected 49 such videos. Lastly, Long Hauler

videos represented first-person accounts of those who considered themselves to have Long Covid. Ultimately, 25 of the 152 videos were from Long Haulers.

As previously mentioned, associated comments were gathered as well. There were 2859 comments total. There were 1258 comments associated with the Medical Source videos. 1078 comments were from News Source videos. 509 comments were from Long Haulers.

Table 2.1: Dataset Description

	Medical Source	News Source	Long Haulers
Videos	78	49	25
Video Length (mean, median)	12:32; 23:57	4:16; 5:04	20:34; 3:11
No. Comments	1258	1078	509
Likes (mean, median)	1363; 5500	749; 56.5	344; 39
Views (mean, median)	72,711; 361,000	63303; 3800	10,529; 297
Subscribers	2,105,960; 1,030,500	3,104,657; 6,334,000	53,668; 125,000

3.4 Data Preprocessing

I handled caption and comment data separately as they existed in different forms. My caption data was captured in json format while my comment data was captured in a tabular format. I manually went through the list of videos and separated each into three categories based on the content creator: Long Hauler, Medical Source and News Source. For each YouTube video, a unique video ID is generated. These video IDs were used in this categorization effort.

Comments were grouped by their Comment ID, so that each comment was being considered individually. The next step was to normalize the comments. I passed the comments through Gensim's `simple_preprocess()` function. This function lowercases and tokenizes the input. The output of the function

is Unicode strings. Punctuation and numbers were removed in addition. Stop words as well as non-English words were additionally removed.

Processing of the captions was slightly different. Because captions were much longer, chronological batches of 50 consecutive words were fed into the model. In their paper, Schwarz et al. argued that text documents that would work well for LDA to be of 50-100 words long [20]. It is because of this that I felt that grouping each video transcript into smaller documents with 50 words each to be sufficient. Once the model generated the topics, these 50 words were traced back to the original text data.

3.5 Model Selection

A large part of my process involved selecting a model. As part of my preliminary experimentation, I tested and compared Gensim's LDA Topic Model, BERTopic and Biterm. In my tests with Gensim's LDA model, I experimented with hyperparameters such as number of topics, chunk size and number of passes. Additionally, I used the `filter_extremes()` function from Gensim to filter out words that did not occur frequently enough in the corpus. The resultant condition was that the word must occur at least twice in the corpus and no more than 65 times. These parameters were arrived to after continual monitoring while generating various models. The best performing model, based on coherence score, performed reasonably well on caption data, however it failed to produce reasonably coherent topics on the shorter, comment data.

Due to its ability to handle shorter texts, I decided to try Biterm on comment data. For each source, I tuned the topic number hyperparameter individually. I employed an iterative process of changing the topic number, running the model, and manually analyzing the resulting topics. This process continued until the topics were reasonably converged based on my manual analysis. Overall, I was satisfied with the results of the Biterm model and decided to try it on the caption data as well. The captions were chunked into 50-word sections so that multiple topics could be uncovered for the same video. I chose 50 words because Schwarz et al. defined 50-100 word documents to be of an optimal size for LDA models [20] I employed the same process of iteratively changing the number of topics and then evaluating the results until the topics were reasonably converged.

Lastly, I tried BERTopic on the longer, caption data. BERTopic abstracts away the ability to choose your own topics. Instead, it determines the optimal number of topics on its own. This results in a hyperparameter tuning process that is more objective and less biased. The trade-off, however, is that there is less control. On the caption data, the resulting topics for the Medical Source category were generally fine, though there was overlap. This is visualized in the heatmap in figure 2.

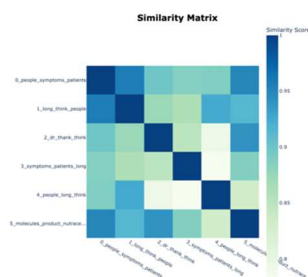


Figure 2.1: BERTopic Results for Medical Source

For the News Source however, only three topics were generated and there was considerable overlap between them. This can be seen in the figure below.

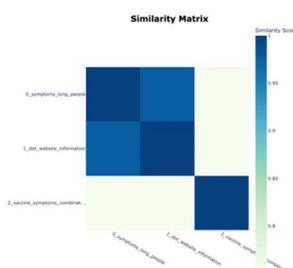


Figure 2.2: BERTopic Results for News Source

Finally, BERTopic generated no topics for the final source, Long Haulers.

After this experimentation, I ultimately went with Biterm as it performed the best on both caption and comment data. To further emphasize the difference between the choices, I've created the table below. Ultimately, this resulted in me creating six models: one each for comment data for each of the three sources and one each for caption data for each of the three sources.

Table 1.2: Pros and Cons of Three Text Mining Methods

Gensim's LDA	Biterm	BERTopic
Pros: Widely used topic model, with a lot of hyperparameter control.	Pros: Optimized for shorter texts	Pros: Less bias in hyperparameter tuning as the ability to choose the number of topics has been removed.
Cons: Not optimized for shorter texts. Performance is much better on longer text data.	Cons: May take considerable tuning to reach desired results.	Cons: Lack of control

3.6 Manual Analysis

After satisfactorily developing the six total models and examining the resultant topics, the next task was to manually analyze the topics that were produced and group them into overarching themes. In order to do this, the top 50 words associated with each topic as well as representative comments or captions were used. I looked for commonalities amongst the comments and words to devise the subject of the topics. Once I had created labels for each topic, I examined them in conjunction with each other to uncover overall themes. These themes and topics are what were ultimately analyzed to make sense of the caption and comment data.

CHAPTER 4: RESULTS

In this section, I describe how the various sources of shared Long Covid information as well as how this information was received by viewers in the video comments. I first discuss the different video transcript i.e. caption topics that were uncovered during the topic modeling process. Then, I discuss how viewers responded to the video sources in the comments section. As previously mentioned, caption videos were divided into three categories based on their origin: Medical Source, News, and Long haulers.

4.1 Captions

4.1.1 Medical Source

Table 1.2 summarizes the 3 themes and 7 topics of Long Covid-related captions observed from Medical Source videos.

Table 4.1: Medical Source Caption Results

Theme	Topic Label	Keywords	Selected sample captions
Explanations in layman's terms	Symptomatology	<i>symptoms, long, fatigue, common, brain, pain, loss, breath, chest, shortness, smell, body, fog, taste, breathing, cough</i>	Cognitive impairments things like word finding difficulty, short-term memory loss, difficulty with multitasking, poor concentration as well as anxiety and PTSD especially in patients who have been hospitalized. The less common ones but still things that we're seeing are loss of sense of taste and smell that is persistent anosmia and digusa...
	Symptom Etiology	<i>syndrome, severe, illness, chronic, different</i>	This is a pretty important complaint because it affects the quality of life for so many of our patients and as I said earlier all of these symptoms, the headache, the sleep disturbance, the brain fog, they often tend to run together and sometimes it's hard to say as to what is leading to what other symptom. It's sort of like the chicken and the egg analogy. Is it because somebody has poor sleep, is that what leads to headaches because we do know what headaches can be triggered when the sleep is poor.
	Symptom Management	<i>vitamin, time, day, sleep, work, different, need, help, right</i>	Anecdotally physicians are trying those psychostimulants as well to see if it helps, we've also heard the mentioned insomnia by Dr. Reggie so it's very common um in long-covered sufferers and I think the first treatment for that insomnia is really sleep hygiene so that's things like um turning off devices a half an hour before bed time, making sure you go to bed at the same time with a relaxing bedtime ritual, waking up at the same time every day, shutting devices off..

Show Housekeeping	Introducing the show/and or guest and validating guest's credentials as a reliable source.	dr, going, thank, time, want, talk, need, help, work, research, medical, information	and he has organized several conferences given many lectures and has done live surgeries as demonstrations in several international conferences and forums... and nhs hospitals in UK. Completed his senior rotation from the Saint Barth and Royal London Hospital in Uk.. He has spent over 14 years in the UK practicing. He has been a board member of the arthroplasty society in Asia. Chairman of the fellowship committee of the Indian Arthroplasty Association ...
	Encouraging the audience to keep in touch	question, talk, help, data, group, better, information	Um as I said there's a conversation on Twitter hashtag covered science um and uh all that remains then is for me to thank everyone that's submitted questions. I hope I got through as many as I could.
Biological Explanations	Immunophenotyping	<i>ccr5, data, number, antigen, cd16, cd14, interleukin, dotted, chord, monocytic, interstitial</i>	But if we dive a little bit deeper into the structure of these molecules we can see that we have simple organic dyes such as feed c and alexa floor 647 you can see the structure here and we are expanding this portfolio with our proprietary spark dyes and we are also expanding our portfolio with a protein with fluid force that are based on proteins such as pe acp or percip our tandems based on these proteins are called fire and in addition to that we also have an expanding or growing portfolio of fluorescent complex molecules such as the brilliant violet dyes and the ketovia dyeswe can see here that you know we can apply these variants in multiple applications such as immunophenotyping cell sorting and also to study cell physiology
	Explaining Mechanics of Immune Responses	<i>disease, percent, inflammation, severe, heart, illness, brain, course, study, viral</i>	One of the similarities was that this protein rantes which is elevated in active coveted was also elevated in long haulers the thing that makes a big difference between long haulers and active covid are two cytokines called interleukin 2 and interferon gamma what's what's really interesting is interleukin 2 and interferon gamma are two cytokines that are intimately involved in antiviral immune responses and they're low in active because it's an emerging infection our immune system presumably has not seen that virus before and can and can only at least begin to initiate an innate immune response as opposed to an adaptive immune response so so long covid actually has an immune response with high il2 high interferon gamma that looks very much like a typical antiviral immune response

Explanations in Layman's Terms. The first theme, Explanations in Layman's Terms, covers three topics: “*Symptomatology*”, “*Symptom Etiology*” and “*Symptom Management*”. The first topic, Symptomatology, covers video transcripts in which the speaker is explaining the symptoms associated with covid. As noted in previous sections, there are a wide range of symptoms associated with Long Covid. Some Medical Source content creators dedicated entire videos to just a few symptoms, or a particular health system as was the case with a video from UAB Medicine titled “Hair Loss & Hives: Post-COVID

Symptoms in Dermatology”: “*when you go through something stressful and you have a telogen effluvium, most of your hairs can enter the resting phase at the same time.*”

Show Housekeeping. Show Housekeeping was another prevalent theme in Medical Source video transcripts. Associated topics were: “*Introducing the Show/Guest and Validating Guest’s Credentials as a Reliable Source*” and “*Encouraging the Audience to Keep in Touch*”. As the name suggests, these videos routinely introduced each of the medical experts on the show and expounded on their credentials. This could potentially be due to the idea that many information consumers can be critical of the source of their information. Expounding on the guest speaker’s credentials could help to build credibility and trust between the video publisher and the audience. The next topic dealt with Encouraging the Audience to Keep in Touch.” Some Medical Source content creators link other social media platforms where they can continue the long hauler conversation with engagers.

Biological Explanations. Biological Explanations featured two distinct topics: “*Immunophenotyping*” and “*Explaining the Mechanics of Immune Responses*”. Immunophenotyping is the process of identifying cells based on antigens or markers [23]. In a video posted by BioLegend called “*Diagnosis Treatment and Immunopathogenesis of Chronic COVID-19 (PASC)*”, the speakers discussed using “*proprietary spark dyes*” which can be used for immunophenotyping [23]. Additionally, these videos were concerned with Explaining the Mechanics of Immune Responses. Here a biological perspective of disease etiology is offered, with less use of layman terminology.

4.1.2 News Source

Table 1.3 summarizes the 4 themes and 8 topics of Long Covid-related captions observed from News Source videos.

Table 4.2: News Source Caption Results

Theme	Topic Label	Keywords	Selected Sample Transcripts
-------	-------------	----------	-----------------------------

Sharing Patient Experiences	Symptoms	<i>patients, symptoms, life, understand, hair, feel, medical, sick, protein, heartbeat, health, doctors</i>	She did an ironman competition earlier this year, the 36-year old uprooted her life to work as a medical missionary and ya that is where she got covid-19. I was young, i do not have health conditions i thought, two weeks and this will be behind me. reporter : she was medically evacuated back to the united states. Five months later, she is still short of breath. Doing therapy three times a week. It often feels like this body is not mine. That the things that i want to do i can't do. Report: a year and a half into the pandemic, we are learning about long hauler covid.
	Treatment	<i>need, better, understand, doctors, months, trying, research, care, answers</i>	That's really reassuring to patients that like, wow, it's not in my head, there's other people going through this. This person can help me and even though there's not a magic pill yet, to cure a long COVID, at least we can try to aggressively manage the symptoms, connect them with other patients, other resources, and try to help in whatever way we can. No it's really similar to many aspects of medicine. If there is no diagnostic test or your CAT scan is negative, we say there's nothing wrong.
Negative Experiences	Not being believed by others and doctors	<i>symptoms, covid, virus, physician, dr, feeling, need, progress, says</i>	And potentially down the line lead to treatments because that's all we want is effective treatments to end this nightmare but also it says to those doctors that deny the existence of long covey that this thing of course it's really look at the science. A glimmer of hope in an ordeal so far without end jonah al jazeera at london very pleased to welcome dr deepti gurusani to al jazeera now clinical epidemiologist and a senior lecturer at queen mary university of london on skype from cambridge today doctor thanks for your time.
	Explaining the impact of Long Covid on lives	<i>started, need, progress, end, taken, coming, time, medical, virus, smell, health, watch, feeling</i>	Differently, less like the flu and more like a condition that can have lasting repercussions. They're the scenes that have given us hope in the deadly fight against the coronavirus. The moment the sockets of the sick get to go home. But for many it's not the end, it's just the beginning of a long and perilous road to recovery.
Experts weighing in	Etiology of the disease	<i>effects, infection, different, virus, actually, research, seen, syndrome, persistent, fatigue</i>	Longhaulers, symptoms and others don't. Today chris hrapsky talked with an expert whose theory on this is gaining attention. Mast cells are the first responders of your immune system when an infection occurs in under a second these cells and stuff like histamine to other cells to say, hey, wake up, something's wrong here. In some people these mass cells go haywire and overreact like central dispatch calling in the swat team for a coffee spill at starbucks and this is called mast cell activation syndrome.
	Experts explaining Long Covid	<i>struggle, lingering, illness, health, syndrome, persistent,</i>	But you'll see as we'll talk about, different studies use different thresholds, which makes it really challenging to compare apples to apples when

		<i>group, body, covid19, physical, related</i>	we're reading the scientific literature or even the popular press. The other thing that makes it really challenging, is that symptoms are not necessarily always correlated or equal to organ dysfunction that we can measure. What that means is that, you have people who may be really severely ill, even critically ill on death's door, on a ventilator, paralyzed, flipped over onto their bellies in the prone position, and a couple of weeks after, sometimes i'll see them in clinic and there'll be remarkably unscathed walking around doing well.
Handling the Long Haul	Managing Symptoms	<i>test, hair, brain, doctor, fatigue, pain, disease, talk, common, exercise, need, home, levels</i>	Sure that's what's been like in terms of they've been wonderful i mean they've sent it took a bit but listen they listened they sent an occupational therapist to see what they could do in the house so our washroom has been retrofitted with a brand new high toilet because he had issues getting on and off the toilet um they retrofitted the shower with a telephone shower and gave us the shower seat so that um you know i can shout i can shower i can do things you know feeling like myself.
	Handling Cardiac/Chest Problems Specifically	<i>oxygen, need, lung, blood, chest, pain, infection, shortness, ventilator, loss, pulmonary, complications, attack. Disease, breath</i>	And it will take three to six months for this myocarditis to settle and that will present with just tachycardia or palpitations so what i suggest is that those of our patients who are having tachycardia it's not a bad idea to get themselves screened by their physicians or cardiologists so that at least we are clear that a patient does not have baseline pulmonary embolism this is one part second reason for tachycardia is that many of the patients are hypoxic.

Sharing Patient Experiences. “Symptoms” and “Treatments” are two topics that belong to the Sharing Patient Experiences theme. The Symptoms related video transcripts dealt with interviewees sharing their daily symptoms in order to give perspective to audiences. Interviewees experienced a wide range of symptoms. These symptoms appeared to have a significant impact on daily life. One interviewee noted that she would fall due to elevated heart rate that worsened doing routine tasks like “*just walking from here to the kitchen*”. Guests were also concerned with find some sort of treatment that could mitigate Long Covid symptoms. Patients seemed to have managed expectations regarding treatment but exhibited some level of hope, “*there's not a magic pill yet, to cure long Covid... at least we can try to aggressively manage the symptoms.*”

Negative Experiences. The Negative Experience theme featured two related topics. The first topic was “*Not being believed by others and Doctors.*” This was a particularly common topic throughout the text. Interviewees shared their experiences of being ignored or not believed. These long haulers sought and could not find affirmation, “*no one really understands me.*” The next topic dealt with “*Explaining the impact of Long Covid*” on Lives. Long Haulers and news reporters introduced Long Covid in general terms as well as the people that it has impacted. Long Haulers of varying age ranges explained “*nearly seven months later and I’m still unwell and I am still a broken woman.*”

Experts Weighing In. The Experts Weighing In theme had two topics: “*Etiology of the Disease*” and “*Experts explaining Long Covid.*” Similar to the case in Medical Source videos, experts took two approaches when speaking about Long Covid. The first approach, as evidenced in Etiology of the Disease explained things from a strict biological perspective, “*Mast cells are the first responders of your immune system when an infection occurs.*” In contrast, in Experts Explaining Long Covid, more commonly used colloquial language was used to explain Long Covid, “*different studies use different thresholds, which makes it really challenging to compare apples to apples.*”

Handling the Long Haul. The last theme for this New Source had two topics as well: “*Managing Symptoms*” and “*Handling Cardiac/Chest Problems Specifically.*” Managing Symptoms dealt mainly with Long Haulers finding their own ways to manage their illness. Additionally, Cardiac and Chest problems were often discussed. They are fairly common symptoms that were addressed by experts and patients alike. Experts offered symptom management advice, “*and it will take three to six months for this myocarditis to settle.*”

4.1.3 Long Haulers

Table 1.4 summarizes the 3 themes and 8 topics of Long-Covid-related captions observed from Long Hauler videos.

Table 4.3: Long Hauler Caption Results

Theme	Topic Label	Keywords	Selected Sample Transcripts
Taking Treatment into Own Hands	Alternate Remedies	<i>new, health, try, better, care, fungus, changing, trusted, broken</i>	The you know the the idea that these toxins are really what's hurting me makes so much sense you know i can't blame all of this what i'm going through now on covid kovid and i believe this wholeheartedly this is not me pretending to sound like some kind of fake spiritual guru coved was my wake-up call to fix my gut and ultimately fix my health i was declining i was already declining before covid i was getting weak i could not keep up my um my aerobics i couldn't keep up my jogging i was going downhill and then covey came and opened my eyes to the fact that i am my body is unhealthy and um so now i have to work on changing that and not necessarily getting back to the person that i was because actually the person the person that i was previously before colbit was a diseased person
	Dealing with Uncertainty	<i>changing, declining, shitty, new, life, work</i>	Please just bear with me through this video this is a update on coronavirus and my personal experience on it haven't personally seen a update video like this like months in advance this is my story right like this is this is what i have to live with for an indefinite period of time so my very good family friend she runs her own practice she's an md and she said you know like nobody should want to get covered because nobody knows the lasting effects of covid
	Not Being Listened to by Doctors	<i>Biases, trusted, chore, dr, doctors, feel, medicine, care</i>	Especially female patients and patients of color the benefit of the doubt there is so much research on patients reporting doctors not believing them or not treating them with the same level of compassion it's unbelievable and as a health expert i knew these alarming statistics going into it but i didn't think it would happen to me let's be real for a moment we all have biases but doctors same as with teachers judges police officers they need to train themselves to drop those biases at the door when they enter their workplace doctors who practice medicine while still holding on to their biases or are simply burned out are violating the trust of their patients and trust is a key element to the patient physician relationship it's time for all of the medical community not just some to move forward
Changes to Daily Life	Insomnia	<i>helped, started, pills, prevent, restless, waking, blockers</i>	I am allowed to take a maximum amount of the sleeping aids and they don't work i just get a calming feeling along with my multitude of symptoms i think along with the drenching sweats and the fevers that just won't stop because my husband has to cover me in ice sometimes because even with medication the fever doesn't stop climbing and he tries to keep me out of a hospital

			visit
	How Symptoms Interrupt Activities	<i>Day, symptoms, time, feel, bad, need, breath, overgrowth, taste, chronic, fever, life, nuts, sacrifice</i>	Didn't realize i was having uh there was having intolerances to food until maybe september or august i had to stop eating eggs i recognize that eggs weren't agreeing with me anymore and i eat three eggs every day i was eating three eggs every day like that was you know that was a breakfast staple for me there wasn't a day where i didn't skip eggs.
	How Symptoms Present Themselves	<i>Experience, fever, health, day, highly, discovering, seizures, entry, permanent</i>	So like whenever i would get near like the oven or the stove or like the air fryer or take a shower or try to exercise like whenever my internal body temperature would rise my face would go bright red it would get swollen id' get like weird patches it was super strange and i didn't know what was going on and like at this stage since i had no clue what was happening i was super aggravated because i like couldn't even work out i could like barely shower and also when i would shower i would get like weird rashes and weird patches on my legs too and i was like this doesn't seem like the hot tub but i don't have any other clue what this like flare-up strange situation was hap coming from so like i was trying everything like i mentioned the rag on my face all the time morning night throughout the day i was also putting manuka honey all over my face to see if it would like calm my skin
Choosing Homeopathy over Pharmaceuticals	Use of CBD and THC	<i>gummies, high, work, try, started, need, help</i>	container um and that is the cbd and the gummies that i take to sleep at night and then in the foreground the three little bottles are the peptides that i take um and i inject those um every morning okay so as far as pharmaceuticals you saw that i'm only taking three different medications um i don't like pharmaceuticals i have never really liked them um even over-the-counter drugs i don't really like like even when i was healthy i was resistant to taking you know like advil whenever i had a headache just ... toxicity into my body that my liver then has to deal with or my kidneys have to deal with um so i just try to keep things as natural
	Turning Down Over the Counter Medicine	<i>deficiency, vitamin, blood, different, taking, bad</i>	and um that's happened before so she put me on vitamin d medication and from there all of my other signs my other tests came back pretty normal so she decided let's just take this vitamin d medication see if your numbers for the vitamin d comes up and see if your headaches stop so that's what she thinks it is so for me i was like no i've had a vitamin d deficiency before and i never had headaches like this so naturally i assume that it's still coronavirus so she encouraged me to take over the count of medication which i don't do i've never done it i don't do it i don't believe in it i don't have a tylenol deficiency i don't have an aspirin deficiency i'm not ibuprofen deficient so i don't think i should take that i feel

Taking Treatment into own Hands. Taking Treatment into own Hands is a theme that has been alluded to in other video sources. It did not surprise me that it showed up here as well. The three related topics were: “*Alternate Remedies*”, “*Dealing with Uncertainties*”, and “*Not Being Listened to by Doctors*.” Alternate Remedies dealt with Long haulers sharing alternative medicine that they used and recommending alternative medicine to others. Uncertainty Management was a common topic as well, Long Haulers noted that they were dealing with symptoms for “*an indefinite period of time*.” Based on their experiences, they had an understanding that doctors were mystified by Long Covid and thus treatments were not certain or foolproof. This leads to the last topic which was “*Not Being Listened to by Doctors*.” A recurrent topic thus far in the study, this dealt with patients not feeling listened to and supported by members of the healthcare system. One particularly popular account of this was shared by Margot Gage Witvilet in a video titled “I’ve had COVID-19 for a year. Here’s what I’ve learned.” She shared her experience as a woman, person of color who felt particularly unfair treatment, “*there is so much research on patients reporting doctors not believing them or treating them with the same level of compassion*.” Long Haulers called for doctors to hold themselves accountable when confronting their own biases, if not long haulers suggested they were “*violating the trust of their patients and trust is a key element to the patient physician relationship*.”

Changes to Daily Life. Next, long haulers discussed the impact of the Long Haul on Daily Life. Associated topics included: “*Insomnia*”, “*How Symptoms Interrupt Activities*”, and “*How Symptoms Present Themselves*.” Long haulers discussed how insomnia impacted their lives. Symptoms impeded their ability to exercise, eat foods they regularly eat and even take showers. Lastly, Long Haulers talked about how the symptoms initially presented themselves.

Choosing Homeopathy over Pharmaceuticals. The two related videos were “*Use of CBD and THC*” for treatment and “*Turning Down Over the Counter Medicine*”. One long hauler looked to THC gummies to cure insomnia in part because “*I don’t like pharmaceuticals, I have never really liked them*.”

4.2 Comments

4.2.1 Medical Source

Table 1.5 summarizes the 3 themes and 9 topics related to Medical Source Comments

Table 4.4 Medical Source Comments Results

Themes	Topic Labels	Keywords	Selected Sample Comments
Ingesting Shared Information	Gratitude for Sharing Science-Based Content	<i>help, many, hope, much, gratitude, respect</i>	Thanks Dr Hansen this Type of science based knowledge empowers us all. Stay safe up there! I just wish many would listen to the science instead of so called "news analysts"
	Health Literacy Issues	<i>information, find, search, assess, help, know</i>	Is there a paragraph in simple English somewhere summing this up?
Seeking More from Shared Content	Seeking Additional Info	<i>know, would, also, worsening, demanding, conditions, assess</i>	Why does this reference unique treatments when he mentions no treatment
	Offering Additional Info	<i>say, funding, impairment, exacerbation, specially, conditions, science, environment, said</i>	I saw advice on NSW Health (Australia) website yesterday to close windows and blinds, use air conditioning and fans. This is during unusually hot weather, but ventilation is important, regardless of ambient conditions.
	Asking Specific, Biological Questions	<i>deficiency, positive, fatigue, brain, negative, taste, science, cardinal, autonomic</i>	What marker is behind vasodilation?
	Desire to Improve Shareability	<i>need, help, people, getting, find</i>	□□□ Please we want to translate it in Arabic
Misinformation	Criticizing the News	<i>convince, know, random, effectiveness, vivid, adverse, mentally, determination</i>	This is the 1st thing I have seen that explains anything besides the news trying to sensationalize and leave out detail on everything. Thank you! keep up the good work.
	Criticizing Doctors	<i>random, never, speak, effectiveness, worried</i>	Thank you for delivering the news and facts. However, I have found your comments or interpretation about some facts are not accurate, sometimes false. Please don't misinterpret the information you have if you are not hundred percent sure since the majority of audience here has no medical or immunology background.

	Offering Alternate Analysis	<i>deficiency, neck, cytokine, magnetic, increase, antigen, sun, lift</i>	long covid is a leaky gut syndrome. A doctor in pakistan is treating people long covid by prescribing them proton pump inhobitor like pantaprazole, eatung honey and yougurt in morning, eating raw onion and garlic with rice veggies, avoiding gluten, sugar, fast food, using olive oil etc. And people are healing in about 1 month of doing this
--	-----------------------------	---	---

Ingesting Shared Information. For the first theme, Ingesting Shared Information, there are two associated topics: “*Gratitude for Sharing Science-Based Content*”, and “*Health Literacy Issues*”. The first topic, Gratitude for Sharing Science-Based Content, includes comments where users are calling out differences between Medical Source and News Source information. They note that science-based comments are more reliable than “*so called ‘news analysts’*”. In contrast to the praise for Science-Based content was issues with health literacy. Health Literacy has been defined by the CDC as the “degree to which individuals have the ability to find, understand and use information... to form health-related decisions.” [22]. As the topic alludes, some commenters had difficulties understanding the information presented by doctors. They excused themselves for “*dumbing this down*”, before offering their own interpretation of the communicated information in layman’s terms.

Seeking More from Shared Content. The topics of this theme included: “*Seeking Additional Info*”, “*Asking Specific, Biological Questions*” and “*Desire to Improve Shareability*”. The comments section of Medical Source videos often had a question, answer component. In the Seeking Additional Info topic, commenters routinely asked additional questions to improve interpretability. This could include accessing additional materials provided by these sources, “*has the website referenced by Dr. Patterson been launched yet?*” Another genre of questions asked were of a more biological nature, as noted in the Asking Specific, Biological Questions topic. For example, one commenter asked for advice on “*using Leronlimab to calm this inflammation.*” Lastly, there was a Desire to Improve Shareability. This often included seeking the video to be translated to various other languages and to include data from other countries. One commenter asked “*when do you include Europe?*”

Misinformation. Misinformation topics included “*Criticizing the News*”, “*Criticizing Doctors*”, and “*Offering Alternate Analysis*”. Comments related to Criticizing the News often included both praise for the sharing of scientific information and condemnation of other sources, most notably news stations. Commenters stated “*all we need is scientific information, regularly given by credible sources.*” In contrast, doctors were at times criticized for not sharing relevant or credible information themselves. They were urged by one commenter to not “*misinterpret information*” because it may negatively persuade viewers.

4.2.2 News Source

Table 1.6 summarizes the 4 themes and 11 topics related to News Source comments.

Table 4.5: News Source Comments Results

<i>Theme</i>	<i>Topic Label</i>	<i>Keywords</i>	<i>Selected Sample Comment</i>
<i>Skepticism</i>	<i>Adverse Political, Government Entity Intervention</i>	<i>perception, problematic, china,</i>	Is there a cure for liberalism yet?
	<i>Perceived Overinflated Dangers of Covid</i>	<i>enter, let, suboptimal, normal, get, perception, problematic</i>	So dangerous yet big box stores are open
	<i>Alternate Causes for Long Covid</i>	<i>cause, still, know, develop, bacteria, activation, gut, glucose</i>	“Long-haulers” may actually be suffering from effects of prolonged mask-wearing. See TLAV channel on YT
	<i>Lack of News Credibility</i>	<i>perception, problematic, neither, know, develop</i>	That women seem a little off her rocker to me
<i>Sharing Long Covid Experiences</i>	<i>Recounting the Long Haul</i>	<i>breathe, gone, fatigued, temple, racing, coughing, everyday, night, intense</i>	April 4 will be 1 year for me hang in there long haulers.
	<i>Relating to Content Shown</i>	<i>shocked, like, feel, know, see</i>	This is how it felt trying to deal with chronic Lyme! Youre fine fine fine, but we are not

<i>Complete Distrust of Information</i>	<i>Anti-Vaccine</i>	<i>plan, harm, frightened, strict, affect, react, vaccine, safe, stop, retaliation</i>	Not buying it, Nope Enjoy your experimental jab
	<i>Relying on Faith</i>	<i>must, unto, hath, shall, ye, believeth, sin</i>	<i>For God so loved the world, that he gave his only begotten Son, that whosoever believeth in him should not perish, but have everlasting life.</i>
	<i>Calling out Perceived Misinformation</i>	<i>react, plan, stop, remember, know, harm</i>	LMAO. This propaganda is so low brow and pathetic
<i>Fears of Hidden Dishonest, Nihilistic Entity</i>	<i>Vaccines Cause Harm</i>	<i>worry, kill, dreadful, wild, useless, exist, vaccine</i>	It's all placebo , they're walking zombies
	<i>Plot to Harm Long haulers</i>	<i>wound, kill, dreadful, wild, useless, exist, death</i>	They're just trying to kill all the long haulers when all you need is some ivy meastrin... Fixed me up in 12 hours

Skepticism. In this sense, skepticism represents distrust in the information presented. I found that there were four root causes of this skepticism, which were relayed in the four topics: “*Adverse Political, Government Entity Intervention*”, “*Perceived Overinflated Dangers of Covid*”, “*Alternate Causes for Long Covid*” and “*Lack of News Credibility.*” Comments in the Adverse Political, Government Entity Intervention reflected misinformation shared by political parties during the Covid pandemic. Many commenters focused on China’s perceived culpability as the genesis of the coronavirus pandemic. As such, they felt that China should be penalized, “*bill china*”. Others perpetrated an idea spread by some politicians that the Covid death count was overinflated, “*I feel for all those gunshot deaths in Colorado labeled as covid.*” In Perceived Overinflated Dangers of Covid, commenters suspected that the dangers of covid were not as great as shared. They offered evidence for why they felt this was true, citing that although the public is told that covid is dangerous, companies are allowed to operate as usual. Alternate Causes for Covid is a topic wherein commenters offer different etiology of Long Covid. Some of these comments are reminiscent

of ideas promoted by some politicians, such as the idea that wearing masks is not helpful. One commenter, blamed mask wearing itself to be a cause of long Covid. In Lack of News Credibility, there were multiple points on which the news was criticized. Commenters outwardly wondered about the political affiliations of speakers in the videos, “*curious to know if she’s a Trump voter?*” Some questioned the credibility of speakers outright, “*this guy doesn’t sound like he knows what he is talking about.*”

Sharing Long Haul Experiences. This theme consisted of two topics: “*Recounting the Long Haul*” and “*Relating to Content Shown.*” Recounting the Long Haul involved commenters sharing specific experiences about their Long Covid journey. These comments are mostly related to the duration of the commenter’s long haul journey and/or the symptoms that they experience. The second topic was Relating to Content Shown. Some long hauler commenters saw themselves in speakers and interviewees in the News Source videos. They shared this openly with comments like “*God, I so relate*” and “*I knew I wasn’t going insane!*”

Complete Distrust of Information. In this theme, commenters took a more firm stance on their distrust of the information presented to them. In Anti-Vaccine, commenters were vocal about their intentions to not get vaccinated. They urged other viewers to not “*take this poison*”, when referring to the vaccine. Additionally, in response to the idea that the vaccine may improve Long Covid symptoms, commenters doubled down on their disbelief, “*scam scam scam. lies lies lies.*” In Relying on Faith, commenters may have been leaning on their faith as guidance in times of turmoil and discord. It could potentially provide a guiding light, when other paths are deemed inadvisable. Quoted Bible verses included themes concerning sin, forgiveness, liars, healing and relying on God instead of man. Lastly, in Calling out Perceived Misinformation, many commenters deemed news reports to be propaganda and lies.

Fears of Hidden Dishonest, Nihilistic Entity. This theme included two topics: “*Vaccines Causing Harm*” and “*Plot to Harm Long haulers.*” The Vaccines Causing Harm topic continued with the idea that vaccines are not helpful. However, comments in this topic took a darker turn. Some suggested that the Covid vaccine could turn people into zombies. Another quoted a movie in which a character says, “*...always trying to kill somebody, why don’t you try lovin somebody?*” Related to this is the idea that there is a Plot

to Harm Long haulers. One commenter noted, “*they’re just trying to kill all the long haulers,*” before promoting Ivermectin. Another suggested that a selection event was occurring and “*a sub-population proves less adapted to living as compared with the rest.*”

4.2.3 Long Haulers

Table 1.7 summarizes the 3 themes and 7 topics related to Long Hauler Comments.

Table 4.6: Long Hauler Comments Results

<i>Themes</i>	<i>Topic Label</i>	<i>Keywords</i>	<i>Selected Sample Comment</i>
<i>Interacting with Content Creators</i>	<i>Asking Follow Up Questions</i>	<i>rash, headache, nausea, heart, pain, brain, sick, sore, feel</i>	<i>Does the vaccine cure lingering symptoms?</i>
	<i>Inviting Conversation</i>	<i>trouble, mysterious, impossible, curious, constantly, say, better, think, know, im, check, usually, help, people</i>	<i>Anybody here can't workout properly after covid?</i>
	<i>Connecting Offline</i>	<i>would, since, hope, call, check, sharing</i>	<i>Hello! Wow this gives me more confidence! Is there any way I can talk to you plz and message you?</i>
<i>Disillusionment with Traditional Medicine</i>	<i>Homeopathic Medicine</i>	<i>fasting, recommend, buy, mindfulness, prescribe, recovery, treatment, better, different, new</i>	<i>Is this type of long term reaction, possibly an indication that it's turning into an autoimmune disorder? Are you open to or able to afford to see a practitioner of natural medicine (which can generally be in conjunction & complementary to what an md may prescribe or advise)? MD's are solely trained to prescribe medications to mitigate symptoms, it sounds like you could use more than solely mitigation of symptoms. My thoughts are with you as you struggle through this. <3</i>
	<i>Criticism of Traditional Doctors</i>	<i>testing, notice, medical, doctors, pressure, make, understand, support</i>	<i>If you really want to wake TRADITIONAL doctors up then separate them from their financial pharmaceutical interest</i>
<i>Distrust with Health Care System</i>	<i>Marginalized Groups receiving unfair treatment</i>	<i>patient, matter, stand, fair, caring, empathy, anxious, support, doctor, hard</i>	<i>This is a lot more common for POC in healthcare. Sadly enough, racial and ethnic minorities receive lower-quality health care.</i>
	<i>Calling for Change in Medical Schools</i>	<i>fair, caring, empathy, college, medical</i>	<i>This presentation needs to be seen in all medical schools</i>

Interacting with Content Creators. The topics included in this theme were: “*Asking Follow Up Questions*”, “*Inviting Conversation*”, and “*Connecting Offline*”. In *Asking Follow Up Questions*, commenters ask questions related to the long haul as well as the content creator’s experiences. For example, one commenter followed up on a creator’s covid journal vlog, by asking about the creator’s hair. She wondered if her hair had been affected by “*any underlying issues.*” Additionally, commenters also interacted with others within the comment section to Invite Conversation about shared symptoms. Lastly, in *Connecting Offline*, users sought to build community with others by connecting offline to further discuss symptoms and experiences.

Disillusionment with Traditional Medicine. This theme included two topics: “*Homeopathic Medicine*” and “*Criticism of Traditional Doctors*”. In *Homeopathic Medicine*, commenters suggested treatments alternative to those commonly suggested by doctors. One commenter suggested supplements such as zinc to help treat symptoms. In *Criticism of Traditional Doctors*, I focus here only on traditional approaches taken by doctors. The morals of traditional medicine were questioned given the money that doctors stood to gain from pharmaceutical companies. The effectiveness of traditional medicine was also called into question. One commenter noted “*I have found the allopathic medical community do not have much to help people actually heal.*”

Distrust with Health Care System. This theme consisted of three topics: “*Marginalized Groups Receiving Unfair Treatment*” and “*Calling for Change in Medical Schools*”. In *Marginalized Groups Receiving Unfair Treatment*, commenters noted that women and people of color were often treated unfairly, “*sadly enough racial and ethnic minorities receive lower-quality health care.*” Another commenter shared, “*Doctors are to black women what cops are to black men.*” In addition to this, *Calling for Change in Medical Schools* was another relevant topic. One commenter noted that “*empathy and compassion training plus holistic and nutrition training*” should be provided to medical students.

4.3 Additional Analysis

A video published by 60 Minutes titled “‘Post-acute COVID-19 Syndrome’: COVID ‘long-haulers’ suffering symptoms months after initial diagnosis” had many interactions within the comment section in comparison to the other videos. The most liked comment was “Condolences to the family and friends of Dr. Mary Fowkes. What a loss, especially of one doing critical work during this pandemic. Rest in peace, Dr. Fowkes.” which received 2520 likes by other users as well as 63 replies. Another well liked-comment was “This is what living with chronic pain and an autoimmune disease is like...24/7.” which garnered 1289 likes from other users. Thirdly, the comment “The first two women described my daily symptoms, but I’m a 73 year old stroke victim with a heart condition. The virus aged them.” received 895 likes.

From the long-hauler category, the video from TEDx Talks “I’ve had COVID-19 for a year. Here’s what I’ve learned | Margot Gage Witvliet | TEDxMileHigh. The most liked comment was “Our medical schools need to add empathy and compassion training plus holistic and nutrition training to them.” which received 4127 likes. Next was “_”Doctors didn't start treating me with compassion until my drug test came back clean”_ This is so wrong for so many reasons! Glad this young woman made it through, despite being treated horribly.” which received 2277 likes, third was “I hate when doctors assume that your symptoms are as a result of anxiety.” which received 2272 likes.

CHAPTER 5: DISCUSSION AND IMPLICATIONS

5.1 Discussion

In both captions and comments, women and POC (people of color) were cited as being less believed by doctors and having more perilous relationships with the healthcare system. This notion has been corroborated by a recent article in NBC news. In [24] journalists documented the stories of several black women suffering with long covid. In one account, one woman was brushed off by physicians and labeled as aggressive, although she had had by that time lost 30 pounds and sight in her right eye due to long covid. According to (NBC News - Black Covid), black people make up the majority of long Covid hospitalizations and researchers think that this will ultimately shrink black life expectancy by 5 years. Margot Witvliet, an epidemiologist, whose video was part of my study, was also featured in [24]. She initiated a Facebook support group where users could discuss their illness in a “safe space.”

As can be imagined, disillusionment with the healthcare system and disappointment with doctors is not limited to the black experience. Long time sufferers of CFS (chronic fatigue syndrome), fibromyalgia, and ME have drawn similarities between not only the symptoms that long haulers experience, but their subsequent treatment as well. Many felt that doctors were ignoring them, or insinuating that they were hypochondriacs, though they knew that this was not the case. This pushed some to seek alternative medicines and take their health into their own hands.

As of August 2022, there is no official test to diagnose Long Covid [24]. However, commonly sought symptoms are prolonged fatigue, cognitive dysfunction, shortness of breath, and lingering pulmonary symptoms [24]. These symptoms commonly appeared in discussion of symptoms, both in caption data and comment data.

Skepticism and Fear of propaganda regularly appeared in News related content. Here, theories suggested by politicians, such as the use of Ivermectin to cure covid abounded. Many criticized the credibility of the news source and its supposed neutrality. News stations and reporters at times were accused of pushing liberal agendas and of pushing fear mongering propaganda. These sentiments were alluded to in

response to Medical Source videos, where commenters thanked Doctors for sharing scientific backed information, as opposed to what is commonly shared from the news media. The YouTube platform itself was not immune to criticism. It was accused of distorting like ratios in favor of viewpoints that correspond to certain political agendas.

The sharing of symptoms and treatments were a common topic amongst all three sources. The idea that long haulers suffered from a wide range of symptoms alluded to by articles in the introduction and literature review section was corroborated by both commenters and video transcriptions alike. Symptoms were shared for a range of purposes. At times, it was to share general knowledge. At other times, it was used as a means to connect with others enduring similar situations.

Health Literacy was a theme that appeared in Medical Source related videos. As noted previously, content from these sources tended to take two tones. In the first, information was delivered in layman's terms which would likely be easy for the average person to understand. In the second, biological explanations of the disease were presented in which less commonly used words were used. Commenters noted that at times, they had issues understanding the content. Health Literacy has been defined by the CDC as the "degree to which individuals have the ability to find, understand and use information... to form health-related decisions." [22]. Issues with health literacy can impede one's ability to properly advocate for themselves and to understand what their options are.

Misinformation and Source Credibility were major themes in both Medical Source and News Source videos and comments. Rather surprisingly, some commenters felt that doctors were not sharing correct information or misinterpreting the information that they had received. This is despite the fact that in many Medical Source videos, there is ample time spent expounding on the credentials of guest speakers, perhaps in an attempt to boost credibility before information is shared. In contrast, some commenters shared the opposite; they appreciated the scientific approach taken by doctors as opposed to news sensationalism. In some ways, this sentiment carried over into my analysis of News Source content. New channels were heavily criticized by commenters for sharing misinformation. Many of the ideas shared by commenters reflect those of recent politicians. In these views, blame for the spread of Covid and Covid-based restrictions

is shifted onto China and liberal politicians. The fierce opposition to the vaccine by commenters has also been reverberated by some politicians as well. In an unexpected and darker turn, some commenters were fearful that those administering the vaccine and those treating long haulers had sinister motives to kill. This sheds light on the idea that, although many debunked sentiments of politicians are being repeated, there is a genuine fear felt by the general public when it comes to vaccines, the healthcare system and some members of the government.

5.2 Public Health Implications

To my knowledge, this was the first study to examine YouTube video transcripts and comments related to Long Covid experiences. This study provides multiple public health implications. For starters, it can help the medical community to understand patient experiences when interacting with those within the healthcare system. Many patients spoke of not being heard or taken seriously. This has ultimately pushed many to seek help amongst peers and from alternative sources. Whilst these activities are not usually detrimental to patient health, it is important for trusting relationships between patients and doctors to exist. Knowing patient sentiments could be a start in bridging the gap in hopes of creating better lines of communication between all parties.

Additionally, I was able to see the types of responses that are voiced by viewers to the various sources of content. Many of those voices expressed skepticism and fears of propaganda. Misinformation and disinformation are their own illnesses that have been plaguing countries. This study supports the idea that these issues should be addressed so that the general public can make informed decisions regarding healthcare moving forward.

5.3 Limitations and Future Research

One limitation is that we cannot assume that the comments present underneath the videos I studied are representative of all viewers. Many viewers do not comment on videos, thus their opinions are not captured. Additionally, I only captured top level comments. This means that any comments posted in reply to these

top level comments have not been captured. Because of this, there should be caution taken to not use this work to make sweeping generalizations.

An additional limitation is that sarcasm cannot be detected using LDA. Sarcasm is commonly used in human speech, however, models have not been taught to recognize it. Because of this, some comments may be received differently than originally intended.

Future research could be focused on longitudinal studies of long haulers. It could be helpful to see how long hauler experiences change over time. Additionally, more research efforts have been focused on Long Covid since the beginning of this study. Long hauler sentiments towards the healthcare system and doctors could potentially have changed over time. Additionally, as more information has surfaced and more Covid infections have likely led to more Long Covid cases, there may have been a change in the level of skepticism and distrust when it comes to Long Hauler content. This would be explored in longitudinal studies as well.

REFERENCES

- [1] 60minutes, “‘post-acute covid-19 syndrome’: COVID ‘long-haulers’ suffering symptoms months after initial diagnosis,” *YouTube*, 24-Nov-2020. [Online]. Available: <https://www.youtube.com/watch?v=0gLmMPOHDwM>. [Accessed: 08-Nov-2022].
- [2] R. Rubin, “As their numbers grow, COVID-19 ‘long haulers’ stump experts,” *JAMA*, vol. 324, no. 14, p. 1381, 2020. doi:10.1001/jama.2020.17709
- [3] L. E. Wisk, G. Nichol, and J. G. Elmore, “Toward unbiased evaluation of postacute sequelae of SARS-COV-2 infection: Challenges and solutions for the Long Haul Ahead,” *Annals of Internal Medicine*, vol. 175, no. 5, pp. 740–743, 2022.
- [4] “Long Covid or post-covid conditions,” *Centers for Disease Control and Prevention*. [Online]. Available: <https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html>. [Accessed: 08-Nov-2022].
- [5] M. W. Tenforde, S. S. Kim, C. J. Lindsell, E. Billig Rose, N. I. Shapiro, D. C. Files, K. W. Gibbs, H. L. Erickson, J. S. Steingrub, H. A. Smithline, M. N. Gong, M. S. Aboodi, M. C. Exline, D. J. Henning, J. G. Wilson, A. Khan, N. Qadir, S. M. Brown, I. D. Peltan, T. W. Rice, D. N. Hager, A. A. Ginde, W. B. Stubblefield, M. M. Patel, W. H. Self, L. R. Feldstein, K. W. Hart, R. McClellan, L. Dorough, N. Dzuris, E. P. Griggs, A. M. Kassem, P. L. Marcet, C. E. Ogokeh, C. N. Sciarratta, A. Siddula, E. R. Smith, and M. J. Wu, “Symptom duration and risk factors for delayed return to usual health among outpatients with covid-19 in a Multistate Health Care Systems Network — United States, March–June 2020,” *MMWR. Morbidity and Mortality Weekly Report*, vol. 69, no. 30, pp. 993–998, 2020.
- [6] “CDC Covid Data tracker,” *Centers for Disease Control and Prevention*. [Online]. Available: <https://covid.cdc.gov/covid-data-tracker/?ftag=MSF0951a18>. [Accessed: 08-Dec-2022].
- [7] Published by L. Ceci and N. 14, “YouTube users by country 2022,” *Statista*, 14-Nov-2022. [Online]. Available: <https://www.statista.com/statistics/280685/number-of-monthly-unique-youtube-users/>. [Accessed: 08-Dec-2022].
- [8] C. A. Melton, O. A. Olusanya, N. Ammar, and A. Shaban-Nejad, “Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence,” *Journal of Infection and Public Health*, vol. 14, no. 10, pp. 1505–1512, 2021.
- [9] X. Yan, J. Guo, Y. Lan, and X. Cheng, “A bitern topic model for short texts,” *Proceedings of the 22nd international conference on World Wide Web - WWW '13*, 2013.
- [10] H. Jelodar, Y. Wang, M. Rabbani, S. B. Ahmadi, L. Boukela, R. Zhao, and R. S. Larik, “A NLP framework based on meaningful latent-topic detection and sentiment analysis via fuzzy lattice reasoning on YouTube comments,” *Multimedia Tools and Applications*, vol. 80, no. 3, pp. 4155–4181, 2020.
- [11] M. Varghese and V. S. Anoop, “Deep learning-based sentiment analysis on covid-19 news videos,” *Lecture Notes in Networks and Systems*, pp. 229–238, 2022.
- [12] M. B. Mutanga and A. Abayomi, “Tweeting on covid-19 pandemic in South Africa: LDA-based topic modelling approach,” *African Journal of Science, Technology, Innovation and Development*, vol. 14, no. 1, pp. 163–172, 2020.
- [13] C. M. Thompson, K. B. Rhidenour, K. G. Blackburn, A. K. Barrett, and S. Babu, “Using crowdsourced medicine to manage uncertainty on reddit: The case of covid-19 long-haulers,” *Patient Education and Counseling*, vol. 105, no. 2, pp. 322–330, 2022.
- [14] I. Trihandini, M. Muhtar, D. A. Sakti, and C. P. Erlianti, “Effects of long-haul COVID on the health-related quality of life among recovered hospitalized patients,” 2022.
- [15] C. H. Basch, E. Park, B. Kollia, and N. Quinones, “Online news coverage of COVID-19 long haul symptoms,” *Journal of Community Health*, vol. 47, no. 2, pp. 306–310, 2021.

- [16] A. McLellan, K. Schmidt-Waselenchuk, K. Duerksen, and E. Woodin, "Talking back to mental health stigma: An exploration of YouTube comments on anti-stigma videos," *Computers in Human Behavior*, vol. 131, p. 107214, 2022.
- [17] S. Syed-Abdul, L. Fernandez-Luque, W.-S. Jian, Y.-C. Li, S. Crain, M.-H. Hsu, Y.-C. Wang, D. Khandregzen, E. Chuluunbaatar, P. A. Nguyen, and D.-M. Liou, "Misleading health-related information promoted through video-based social media: Anorexia on YouTube," *Journal of Medical Internet Research*, vol. 15, no. 2, 2013.
- [18] O. Oyeboode, C. Ndulue, A. Adib, D. Mulchandani, B. Suruliraj, F. A. Orji, C. T. Chambers, S. Meier, and R. Orji, "Health, psychosocial, and social issues emanating from the COVID-19 pandemic based on social media comments: Text mining and thematic analysis approach," *JMIR Medical Informatics*, vol. 9, no. 4, 2021.
- [19] M. Grootendorst, "Neural topic modeling with a class-based TF-IDF procedure," arXiv: 2203.05794 [cs.CL], Mar. 2022
- [20] C. Schwarz, "Ldagibbs: A command for topic modeling in stata using latent Dirichlet allocation," *The Stata Journal: Promoting communications on statistics and Stata*, vol. 18, no. 1, pp. 101–117, 2018.
- [21] E. T. Jacques, C. H. Basch, E. Park, B. Kollia, and E. Barry, "Long Haul COVID-19 videos on YouTube: Implications for health communication," *Journal of Community Health*, vol. 47, no. 4, pp. 610–615, 2022.
- [22] "What is health literacy?," *Centers for Disease Control and Prevention*, 02-Feb-2022. [Online]. Available: <https://www.cdc.gov/healthliteracy/learn/index.html>. [Accessed: 08-Dec-2022].
- [23] "Immunophenotyping," *Stanford Health Care (SHC) - Stanford Medical Center*, 12-Sep-2017. [Online]. Available: <https://stanfordhealthcare.org/medical-tests/i/immunophenotyping.html#:~:text=Immunophenotyping%20is%20a%20test%20in,the%20surface%20of%20the%20cell>. [Accessed: 08-Nov-2022].
- [24] "Black covid long-haulers say doctors dismissed their symptoms, so now they're relying on one another for support," *NBCNews.com*. [Online]. Available: <https://www.nbcnews.com/news/nbcblk/black-covid-long-haulers-felt-invisible-health-care-system-formed-supply-rcna44468>. [Accessed: 08-Nov-2022].

APPENDIX A: Detailed Dataset Description

TABLE A.1: Detailed Dataset Description

Video Title	Duration	Publish Date	Likes	Views (as of 12-1-22)	# Comments (as of 12-1-22)	# Subscribers
'Pray For A Miracle': Covid Long-Haulers Fight Debilitating Symptoms A Year Later NBC News	6:06	March 11, 2021	1.5K	117K	827	7.25M
'Post-acute COVID-19 Syndrome': COVID "long-haulers" suffering symptoms months after initial diagnosis.	13:22	Nov 24, 2020	33K	2.5M	10, 307	1.77M
Long Term Effects of COVID on the Body Long Haulers Long COVID	9:15	Oct 8, 2020	9.8K	192K	1, 456	1.03M
COVID-19 'long haulers'	6:46	Nov 18, 2020	890	84K	794	13.9M
COVID long hauler claim vaccine helped ease symptoms GMA	2:57	March 24, 2021	127	14K	95	13.9M
COVID Symptoms – Long Hauler	15:44	March 25, 2021	648	11K	213	601K
COVID Long-Haulers Are Fighting For Their Future	4:25	Sept 30, 2021	280	5.9K	84	645K
Vaccine impacts of COVID long haulers	3:28	April 14, 2021	17	2,427	17	778K
COVID long haulers share experiences and how to get help	5:59	Jul 23, 2021	13	803	3	242K
COVID-19 'long haulers' struggle with debilitating symptoms long after infection	2:51	Nov 17, 2020	31	2.6K	23	508K

COVID long-haulers call for better research into their conditions	8:11	Sept 1, 2021	71	2.9K	38	4.66M
Children are among the COVID-19 'long haulers'	4:47	Feb 25, 2021	344	17K	123	13.9M
COVID Long Haulers – Discussion with Dr. Bruce Patterson	1:11:38	Dec 15, 2020	1.6K	68K	456	601K
Hope for COVID Long-Haulers	9:44	April 28, 2021	2.1K	105K	1,069	159K
My Long Hauler Covid-19 Recovery (Brain Fog, Candida Overgrowth, Leaky Gut	53:20	Sept 10, 2021	30	652	25	26
Long COVID/ Long Hauler Recovery Tips	21:31	May 23, 2021	256	4.4K	162	364
8 Things that has Helped Me Treat My Long Hauler Symptoms How I'm Doing	28:52	Mar 15, 2021	64	1K	48	8.7K
COVID – LONG HAULER (recovery – POTS) MARAVIROC + STATIN	8:27	August 26, 2021	123	4.7K	112	334
Tips for people recovering from post covid POTS	9:44	Sept 21, 2021	82	2.9K	67	334
How to recover from Long Covid My long covid story Long covid recovery DNRS success story	16:17	April 18, 2021	409	15K	310	599
Am I a COVID Long Hauler?	15:27	Feb 17, 2021	6	105	2	17
I Caught Covid: 111 Days Later I'm Still Sick. Journal: #1	<i>video is now private</i>	<i>video is now private</i>	<i>video is now private</i>	<i>video is now private</i>	<i>video is now private</i>	<i>video is now private</i>
COVID Chronicles: My Long Hauler Story	15:34	Oct 29, 2021	76	1K	38	116
I Have COVID – Day 25 Long Hauler?	3:11	Dec 9, 2020	39	297	1	125K

I am suffering with Covid-19 long hauler symptoms	12:06	May 20, 2021	1	61	3	1.3K
COVID Long Hauler Update after 10 months	4:09	Feb 1, 2021	12	457	4	325
Long Covid Journey – Coronavirus Long hauler	27:41	Nov 10, 2020	12	312	7	891
Lung rehab: 7 th month covid long Hauler recover	5:10	May 31, 2021	10	187	8	5.69K
Returning to Work After being Home 3Months with Covid/Long Hauler after being Fully Vaccinated	10:43	May 31, 2021	9	127	4	1.75K
Thoughts on 2020, Recovering from COVID, long hauler symptoms and what's next	5:14	Dec 31, 2020	120	1.5K	79	4.04K
COVID-19 Long Hauler // Update On Michelle And Where We Have Been	44:30	Feb 13, 2021	33	200	11	8.37K
My meds as a COVID long-hauler	Video is now private	Video is now private	Video is now private	Video is now private	Video is now private	Video is now private
COVID-19 Long-Haulers: Patients Share Lasting Impact of Virus TODAY	4:23	Dec 23, 2020	235	41K	404	3.71M
News research helps COVID-19 long haulers	3:09	March 8, 2021	20	2.1K	12	340K
Long Haulers Still Battle COVID-19 Effects	2:31	Jan 22, 2021	94	9.3K	82	29.6K
Covid Long-Haulers Desperate to Find Treatment NBC News NOW	2:33	March 6, 2021	132	10K	106	7.25M
COVID-19 long haulers update	17:29	Sept 28, 2021	33	2.6K	24	8.44K
COVID-19 vaccine may relieve long-	5:25	April 6, 2021	46	3.5K	34	71.9K

haulers of relentless symptoms						
Dr. Steven Sheris on unique treatment for COVID long-hauler COVID-19 Update for April 5, 2021	13:24	April 5, 2021	334	31.7K	N/A	25.2K
Massive Medical Bills Plague Some Covid Long-Haulers	6:42	June 7, 2021	176	12K	178	7.25M
In-Depth: Suicide among COVID long-haulers	3:23	July 2, 2021	51	2.7K	39	340K
COVID-19 Long-Haulers Deal With Lingering Virus Symptoms	2:10	March 4, 2021	534	188K	905	177K
The life of COVID-19 long-haulers ABC News	3:52	May 28, 2021	301	19K	259	13.9M
Some long haulers still experience Covid symptoms a year after contracting the virus	3:26	May 26, 2021	132	9.1K	120	2.28M
COVID vaccine could bring relief for long-haulers	5:29	March 24, 2021	102	6.4K	73	13.9M
New survey finds some COVID long haulers make dramatic improvements post-vaccination	4:45	April 19, 2021	354	46.9K	542	2.4M
COVID-19 “long haulers” organize online to study themselves	7:00	August 18, 2020	158	8.2K	61	4.66M
Covid Long-Haulers Fear for Their Future WSJ	6:35	Dec 4, 2020	1.8K	100.3K	554'	3.91M
COVID long haulers	5:45	Feb 11, 2021	0	205	2	224K
Covid-19 Long Haulers: What to Expect From Post COVID Symptoms	2:50	May 6, 2021	297	31.6K	185	9.27K

Study by the Mayo Clinic released about COVID-19 Long-Haulers Syndrome shows surprising findings	3:41	May 13, 2021	37	3.6K	33	674K
Covid “long haulers” face months of symptoms, financial strain	4:00	March 12, 2021	24	2.7K	17	2.28M
COVID long haulers update	22:23	Nov 28, 2020	4K	1,483	102,258	2.57M
COVID-19 long-haulers see improved symptoms with vaccine	2:57	March 18, 2021	35	2.6K	29	39.1K
Treating COVID-19 long haulers	3:04	Jun 17, 2021	56	3.4K	19	652K
Long COVID Treatment, Symptoms, and Recovery (Long Haulers)	44:44	Jun 5, 2021	10K	701K	428	1.46M
New Treatment Shows Promise for COVID Long-Haulers	2:20	July 24, 2021	1.1K	75.6K	794	345K
Northwestern Studies Symptoms of COVID-19 Long Haulers	7:44	March 31, 2021	116	7,360	50	38.8K
Long-haulers: A look at the long-term effects of COVID-19 and what doctors have discovered	3:21	Nov 12, 2020	64	3.9K	16	368K
Exclusive Look at NIH Investigation Into Covid “Long-Haulers” NBC Nightly News	4:38	March 1, 2021	217	34.6K	150	7.25M
Ocala company reveals new COVID-19 drug for virus ‘long haulers’	1:57	Jan 7, 2021	23	2K	6	339K
Important Long Covid Information	1:12:24	July 21, 2021	9.1K	467K	4,110	2.57M
Post Acute Covid 19 Advice and Rehabilitation	19:17	Jan 18, 2021	869	34K	207	19.4K

Long Covid Recovery Guidance						
Diagnosis, Treatment and Immunopathogenesis of Chronic COVID-19 (PASC)	42:46	Jun 16, 2021	43	2.8K	5	4.84K
Trying to Understand Long COVID in Kids	2:55	Sept 23, 2021	15	1K	3	53.5K
COVID 19 Update: PASC featuring Dr. Bruce Patterson, Dr. Ram Yongendra and Jenny Hobbs	1:26:40	April 12, 2021	67	1.9K	12	849
What Do We Know about Long COVID	1:13:29	July 12, 2021	811	57.5K	23	1.19M
Coronavirus complications DW Documentary	28:24	Dec 22, 2020	10K	867K	2,070	4.44M
COVID Long Haulers and the Connection to ME/CFS by Dr. Bateman	55:32	Sept 13, 2021	295	9.2K	295	8.92K
Research Webinar: Long COVID	1:00:24	July 28, 2021	15	1K	3	2.63K
Long COVID research update: Blood abnormalities could cause Long COVID COVID-19 Special	12:04	Jun 30, 2021	1.3K	150K	478	4.32M
Researchers find potential cause of long-lasting COVID-19 symptoms	2:33	Sept 9, 2021	18	1.9K	6	49.2K
COVID-19 Long-term Symptoms: Infectious Disease Perspective	17:10	Nov 5, 2020	71	5K	27	12.3K
COVID-19 Webinar: Long COVID	1:12:09	March 3, 2021	11	728	9	689
Chronic Fatigue Syndrome (CFS) and Post-Acute Sequelae of SARS-CoV-2 Covid19 infection (PASC)	11:02	May 23, 2021	564	15K	120	466K

Post-COVID-19: The Long Haul of the Pandemic, UT Health Austin Webinar, September 27, 2021	47:22	Sept 30, 2021	337	25K	99	1.71K
Webinar on LONG COVID CONDITIONS – PASC @COVID19	1:55:33	Oct 23, 2021	29	1K	2	280
IncellDX Long COVID 19 (PASC) Research Video with Dr Bruce Patterson Covid 19 Long Haulers	3:49	July 16, 2021	627	29K	195	23.2K
Post Acute SARS CoV 2 PASC or Long Covid	6:42	August 28, 2021	16	145	2	155
Post-acute sequelae of SARS-CoV-2 infection or PASC is the new name: COVID 19 FAQ 6533-656	1:48	Feb 25, 2021	35	1,511	1	485K
Living with “Long-Haul COVID-19”	28:32	June 16, 2021	280	181	20,288	10.6K
Facebook Live: Post COVID-19 recovery	47:49	March 12, 2021	3	347	3	62.6K
Pasc: Coronavirus (And The Work On My Album)	<i>Video no longer available</i>	<i>Video no longer available</i>	<i>Video no longer available</i>	<i>Video no longer available</i>	<i>Video no longer available</i>	<i>Video no longer available</i>
What is “long-COVID” / Post-acute COVID-19 / Chronic COVID-19 or post-COVID syndrome	3:03	May 21, 2021	191	13K	66	804K
COVID-19 Recovery and Help for Long-Haulers	44:58	May 17, 2021	6	931	3	32.3K
Post Covid Symptoms	13:54	Mar 2, 2021	642	56K	20	23.4K
Post Covid Syndrome: Heart and lungs	8:11	Mar 21, 2021	6.3K	576K	2,472	2.65M
Mayo Clinic doctor explains what is	7:09	Dec 9, 2020	672	65K	356	2.29M

known about 'post-Covid syndrome'						
Post COVID Symptoms Dr. Ravi Shekhar Jha	3:05	Dec 2, 2020	3.2K	292K	1,251	168K
Post-Covid complications: Symptoms and treatment explained by AIIMS Chief	3:39	May 25, 2021	1.4K	143K	794	204K
Post-COVID Symptoms Dr. Mrinal Sircar	3:47	Dec 16, 2020	1.1K	125K	768	168K
How to deal with 3 lasting symptoms of 'long covid'	3:36	Dec 17, 2020	840	80K	738	2.11M
Neurologic Complications with Post-COVID Syndrome	44:23	Jun 30, 2021	360	39K	200	16.9K
Post-Covid Complications To Watch Out For Doctors Explain	<i>Video no longer available</i>	<i>Video no longer available</i>	<i>Video no longer available</i>	<i>Video no longer available</i>	<i>Video no longer available</i>	<i>Video no longer available</i>
Post-COVID syndrome: Mental Health	9:53	Mar 31, 2021	1.2K	94K	308	2.65M
What is "long COVID" / Post-acute COVID-19 / Chronic COVID-19 or post-COVID syndrome	3:03	May 21, 2021	191	13K	66	804K
What is Post-COVID Syndrome? What are Post-COVID Symptoms? Phablecare	3:12	May 14, 2021	83	9.6K	155	7.76K
What is Post Covid? Symptoms.. Do's & Don't's	10:14	Jun 20, 2021	1.4K	67K	1,315	44.8K
Covid-19 and Long-term Recovery	4:04	Sept. 16, 2020	3.3K	236K	913	441K
Pulmonary Complications with Post-Covid Syndrome	41:43	Jun 30, 2021	69	6.1K	3	16.9K
Bone & Joint Symptoms post	36:37	Jul 13, 2021	59	24K	10	327K

COVID-19 infection						
What should be actions taken by the patient post Covid-19 Recovery? Apollo Hospitals	3:40	May 18, 2021	267	41K	35	327K
WHO's Science in 5 on COVID-19: Prolonged symptoms after Covid-19-30 December 2020	5:12	Dec 30, 2020	99	17K	14	834K
Post-COVID-19 GI Symptoms: Are We in for the Long Haul?	6:59	Oct 18, 2021	317	20K	228	181K
Expert on Managing Post-Covid Recovery Coronavirus: Fact Vs Myth	6:47	Jun 10, 2021	49	3.7K	13	12.3M
Symptoms to watch out for post COVID recovery	10:18	May 8, 2021	123	11K	6	33.2M
Post COVID Fatigue Syndrome: What can we do about it? Apollo Spectra & COVID #1	3:09	Jun 7, 2021	918	62K	369	86.9K
AIIMS Director Dr Randeep Guleria Speaks About Post-COVID Symptoms and Long-Term Damages on Body	5:36	Oct 16, 2020	539	78K	79	6.67M
Hair Loss & Hives: Post-COVID Symptoms in Dermatology	23:14	Jul 7, 2021	319	32K	231	16.9K
Future uncertain for those struggling with post-COVID symptoms ABC7 Chicago	4:20	Feb 23, 2021	27	1.2K	7	330K
My Lingering Symptoms Since Recovering from Coronavirus Lasting Symptoms of Covid-19 Months Later	11:36	Oct 17, 2020	232	17K	291	40.3K

Post-Covid Care Dr. Abhradip Das	3:34	Jan 21, 2021	47	4K	10	658K
Don't ignore any symptoms of eye infection after COVID 19 Apollo Hospitals	33:07	June 25, 2021	179	57K	108	327K
Is Hair Loss a Post-Covid Symptom?	14:03	June 22, 2021	1.3K	108K	853	12.3M
COVID 19: Lingering and Chronic Symptoms After Recovery from Coronavirus Infection	10:35	Aug 30, 2020	1.6K	133K	1,526	15.3K
POST-COVID Symptoms? Do This! – Dr.Berg	6:20	Feb 10, 2021	25K	854K	1,949	8.98M
Gastro-Intestinal Post COVID-19 Dr. Pinakin Patel	3:30	May 21, 2021	532	88K	1,129	658K
Take Care of Your Heart After Covid Recovery FYI	24:32	May 4, 2021	958	84K	214	12.3M
How to recover from lung damage post COVID-19 infection? Apollo Hospitals	2:38	Jul 9, 2021	505	73K	150	327K
Post-COVID Syndrome	1:49	Nov 22, 2020	183	6.6K	43	626K
Top Doctors Answer FAQs on Long Covid or Post Covid Syndrome News Today with Rajdeep Sardesai	8:49	Jun 4, 2021	12	1.2K	1	6.67M
Post Covid-19 Syndrome Symptoms Treatment and Management Second Pandemic	2:42	Jun 16, 2021	31	3.3K	2	86.9K
'Post-Covid syndrome' may have severe impacts: WHO	1:33	Feb 26, 2021	31	11K	29	479K
Post-COVID Syndrome: A Functional Medicine Approach	37:51	Aug 3, 2020	589	15K	71	626K

New study looks at post-COVID syndrome	2:06	Oct 1, 2021	76	4.8K	13	4.33M
Mnay suffer post-COVID syndrome	0:35	Feb 25, 2021	3	187	1	224K
What is MIS-C which is affecting children post COVID recovery?	1:56	May 31, 2021	33	3.5K	4	34.4M
Study reveals post-COVID syndrome severely damages children's hearts	1:10	Sep 7, 2020	4	1.1K	2	33.2M
Post-Covid Syndrome	12:16	Jan 28, 2021	780	44K	209	1.85M
Post-COVID syndrome: Latest Treatment: Dr. Gupta Lokesh – TV9	12:39	Dec 18, 2020	0	24K	48	9.44M
What is 'post-COVID syndrome'? ANC	3:37	Dec 8, 2020	9	2.3K	5	1.17M
Post COVID Long Haul Syndrome (PCLS) Management	51:10	Jun 16, 2021	1K	21K	152	601K
Post Covid Condition # Long COVID # Post COVID Syndrome	5:15	Aug 5, 2021	17	675	1	63.6K
Dr. Mitchell Miglis on treating post-COVID syndrome patients COVID-19 Update for April 28, 2021	15:03	April 28, 2021	97	12K	<i>Comments turned off</i>	25.3K
Here's what doctors know about Post-Covid Syndrome so far	2:29	Jan 10, 2021	5	667	3	94.7K
Post-COVID Syndrome: What Family Physicians Need to Know	1:00:58	July 21, 2021	27	3K	2	3.07K
Post Acute Covid 19 Exercise and Rehabilitation Long Covid Recovery	19:17	Jan 18, 2021	874	34K	207	19.5K

Diagnosis, Treatment and Immunopathogenesis of Chronic COVID-19 (PASC)	42:46	Jun 16, 2021	43	2.8K	5	4.85K
Research Webinar: Long COVID	1:00:24	Jul 28, 2021	15	1K	3	2.64K
Chronic Fatigue Syndrome (CFS) and Post-Acute Sequelae of SARS-CoV-2 Covid19 Infection (PASC)	11:02	May 23, 2021	564	15K	120	468K
Lungs rehab: 7 th month covid long hauler recovery	5:10	May 31, 2021	10	187	8	5.72K
COVID long haulers claim vaccine helped ease symptoms GMA	2:57	Mar 24, 2021	127	14K	95	13.9M
Matt Breslow's COVID-19 long hauler story	2:44	Sep 23, 2021	292	21K	365	331
Expert Talks About COVID-19 Long Hauler Symptoms	2:59	Apr 16, 2021	29	3.5K	17	29.6K
COVID Long Hauler (Symptoms and Management)	31:40	Feb 22, 2021	2.1K	65K	614	601K
WTOL 11: Can vaccines help COVID-19 'long-haulers'?	2:37	Apr 9, 2021	13	2K	4	807
COVID-19 long haulers share their stories	2:50	Jul 28, 2021	8	1.7K	11	674K
COVID 19 long haulers	4:13	Jun 18, 2021	13	801	3	653K
Ways to help COVID-19 long haulers	6:39	May 17, 2021	8	711	7	165K
COVID long-hauler now covered by disability	2:34	Jul 26, 2021	98	7.3K	34	700K
New research helps COVID-19 long haulers	3:09	Mar 8, 2021	20	2.1K	12	340K
COVID-19 "Long-Haulers" Bring Long-Term Effects	8:00	Aug 12, 2020	1K	104K	200	7.26M

Into The Spotlight NBC News						
----------------------------------	--	--	--	--	--	--

Please bear in mind that some videos have since been made private and these comment counts represent all comments, not just the top-level comments that I captured. Some comments have since been turned off on videos as well.