

# Understanding COVID-19 Long Haulers: Computational Analysis of YouTube Content

Alexis Jordan, Albert Park

Submitted to: JMIR AI  
on: November 14, 2023

**Disclaimer:** © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

*Table of Contents*

**Original Manuscript..... 5**  
**Supplementary Files..... 36**  
    Figures ..... 37  
        Figure 1..... 38

Preprint  
JMIR Publications

# Understanding COVID-19 Long Haulers: Computational Analysis of YouTube Content

Alexis Jordan<sup>1</sup> MS; Albert Park<sup>1</sup> PhD

<sup>1</sup>Department of Software and Information Systems UNC Charlotte Charlotte US

## Corresponding Author:

Albert Park PhD  
Department of Software and Information Systems  
UNC Charlotte  
9201 University City Blvd, Charlotte, NC 28223  
Woodward 310H  
Charlotte  
US

## Abstract

**Background:** The coronavirus (COVID-19) pandemic had a devastating global impact. In the United States, there were more than 98 million COVID-19 cases and over 1 million resulting deaths. One consequence of COVID-19 infection has been Coronavirus Post-Acute Sequelae (PASC). People with this syndrome, colloquially called Long Haulers, experience symptoms that impact their quality of life. The root cause of PASC and effective treatments remains unknown. Many Long Haulers have turned to social media for support and guidance.

**Objective:** In this study, we sought to gain a better understanding of the Long Hauler experience by investigating what has been discussed and how information about Long Haulers is perceived on social media. We specifically investigated the following: (1) the range of symptoms that are discussed, (2) the ways in which information about Long Haulers is perceived, (3) informational and emotional support that is available to Long Haulers, and (4) discourse between viewers and creators. We selected YouTube as our data source due to its popularity and wide range of audience.

**Methods:** We systematically gathered data from three different types of content creators: Medical sources, News sources, and Long Haulers. To computationally understand the video content and viewers' reaction, we used Biterm, a topic modeling algorithm created specifically for short texts, to analyze snippets of video transcripts and all top-level comments from the comment section. To triangulate our findings about viewers' reaction, we used VADER to conduct sentiment analysis on comments from each type of content creator. We grouped the comments into positive and negative categories and generated topics for these groups using Biterm. We then manually grouped resulting topics into broader themes for the purpose of analysis.

**Results:** We organized the resulting topics into 28 themes across all sources. Examples of Medical source Transcripts themes were Explanations in Layman's Terms and Biological Explanations. Examples of News Transcripts themes were Negative Experiences and Handling the Long Haul. Two Long Hauler Transcripts themes were Taking Treatments into Own Hands and Changes to Daily Life. News sources received a greater share of negative comments. A few themes of these negative comments included Misinformation/Disinformation and Issues with the Healthcare System. Similarly, negative Long Hauler comments were organized into several themes, including Disillusionment of Healthcare System and Requiring More Visibility. In contrast, positive Medical source comments captured themes such as Appreciation of Helpful Content, and Exchange of Helpful Information. Following this theme, one positive theme found in Long Hauler Comments was Community Building.

**Conclusions:** Results of this study 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 develop their communication strategy concerning Long Covid.

(JMIR Preprints 14/11/2023:54501)

DOI: <https://doi.org/10.2196/preprints.54501>

## Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

Please make my preprint PDF available to anyone at any time (recommended).

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.

✓ **Only make the preprint title and abstract visible.**

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <http://www.jmir.org/>

## Original Manuscript

## Abstract

**Background:** The coronavirus (COVID-19) pandemic had a devastating global impact. In the United States, there were more than 98 million COVID-19 cases and over 1 million resulting deaths. One consequence of COVID-19 infection has been Coronavirus Post-Acute Sequelae (PASC). People with this syndrome, colloquially called Long Haulers, experience symptoms that impact their quality of life. The root cause of PASC and effective treatments remains unknown. Many Long Haulers have turned to social media for support and guidance.

**Objective:** In this study, we sought to gain a better understanding of the Long Hauler experience by investigating what has been discussed and how information about Long Haulers is perceived on social media. We specifically investigated the following: (1) the range of symptoms that are discussed, (2) the ways in which information about Long Haulers is perceived, (3) informational and emotional support that is available to Long Haulers, and (4) discourse between viewers and creators. We selected YouTube as our data source due to its popularity and wide range of audience.

**Methods:** We systematically gathered data from three different types of content creators: Medical sources, News sources, and Long Haulers. To computationally understand the video content and viewers' reaction, we used Biterm, a topic modeling algorithm created specifically for short texts, to analyze snippets of video transcripts and all top-level comments from the comment section. To triangulate our findings about viewers' reaction, we used VADER to conduct sentiment analysis on comments from each type of content creator. We grouped the comments into positive and negative categories and generated topics for these groups using Biterm. We then manually grouped resulting topics into broader themes for the purpose of analysis.

**Results:** We organized the resulting topics into 28 themes across all sources. Examples of Medical source Transcripts themes were *Explanations in Layman's Terms* and *Biological Explanations*. Examples of News Transcripts themes were *Negative Experiences* and *Handling the Long Haul*. Two Long Hauler Transcripts themes were *Taking Treatments into Own Hands* and *Changes to Daily Life*. News sources received a greater share of negative comments. A few themes of these negative comments included *Misinformation/Disinformation* and *Issues with the Healthcare System*. Similarly, negative Long Hauler comments were organized into several themes, including *Disillusionment of Healthcare System* and *Requiring More Visibility*. In contrast, positive Medical source comments captured themes such as *Appreciation of Helpful Content*, and *Exchange of Helpful Information*. Following this theme, one positive theme found in Long Hauler Comments was *Community Building*.

**Conclusions:** Results of this study 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 develop their communication strategy concerning Long Covid.

**Keywords:** Long Haulers; Long Covid; COVID-19; YouTube; Topic Modeling; Natural Language Processing

## Introduction

*"It's like a ... like a viral tornado that goes in you and kind of just messes you up,"* S. N. says in between labored breaths [1] This is how the account of 44-year-old fitness instructor S. N.'s battle with Long Covid began during a news interview [1]. In the comments section of the video, one user writes, *"I had to pause the 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, S. N. had been a Long Covid patient for 8 months. By this time, she claims that 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 [1]. It is clear from the comments left under the 60 Minutes video, that S. N. is not alone in experiencing a drastic change in her quality of life.

The coronavirus pandemic has changed the lives of many, though one consequence of it has received less attention [2]. Coronavirus Post-Acute Sequelae (PASC) has been identified as a syndrome affecting patients long after their initial COVID-19 infection has cleared. These patients are colloquially called Long Haulers [3]. High ratios of those who have been infected with COVID-19 have persisting symptoms that last months after initial infection.

Studies have shown that Long Covid has real implications in people's everyday lives. The World Health Organization Quality of Life Brief Version (WHOQOL-BREF), a quality-of-life questionnaire, was administered amongst patients who had been hospitalized with COVID-19 [4]. The results showed that 30.2% of respondents had Long Covid, which affected nearly all domains of Quality of Life as outlined in the WHOQOL-BREF criteria [4]. Moreover, there have been recent links between Long Covid and deteriorating mental health [5].

Long Covid leaves negative economic implications as well. Those with Long Covid are often not in condition to work and thus realize their full earning potential [2]. Approximately 44% of people with Long Covid are completely out of the workplace, whereas 51% had reduced hours at work [2]. This could result in over \$50 billion in lost income annually [2].

Additionally, some patients have not received insurance coverage for Long Covid related testing and treatments [6]. This has led to significant debt for some patients [6]. In May 2023, JAMA estimated that average Long Covid related medical costs could be around \$9000 a year [6]. There is also the issue of lost wages due to Long Covid, which further complicates the medical debt. Making a case for those with Long Covid, by uncovering patient experiences, could be useful for public health officials and medical insurance companies who may need additional help in understanding how debilitating Long Covid can be.

Social media is a rich source of information regarding people's experiences and attitudes [7,8] due to the pervasiveness of social media apps and the freedom with which people engage in discourse on various topics. Such pervasiveness contributes to the increased size of health-related data [7]. This has encouraged researchers to use computational models to analyze social

media texts concerning COVID-19 [7,9–12]. One popular method of analysis is topic modeling. Topic modeling allows for discovery of thematic relationships and patterns within a body of text using natural language models [10]. Latent Dirichlet Analysis (LDA) is a probabilistic unsupervised classification method [13]. It has been widely used in studies employing topic modeling on a large set of documents [13].

For example, [10] used an LDA topic model to study COVID-19 related tweets in South Africa and found that conversations revolved around *alcohol consumption, staying at home, vaccine conspiracy theories, police brutality, statistics training* and *5G* [10]. In addition, authors explored public sentiment and discourse on COVID-19 vaccines on Reddit using an LDA model [9]. They found that posts covered the broader discussions of *vaccine, safety concerns, efficacy, and side effects*.

To date, there has been one other study that examined the experiences of Long Haulers on YouTube in hopes of understanding virtual health communication [7]. However, [7] did not employ any topic modeling methods. Instead, [7] manually coded the 100 most viewed Long Covid videos based on a predetermined list of themes.

In the following section, we provide a review of Long Haulers and health discussions on YouTube. Based on this, we seek to understand what types of videos are available on YouTube regarding Long Haulers and how users respond to Long Covid related content. Next, the procedure of data collection and analysis was provided. Results were then reported regarding salient themes in each type of content creators and positive and negative comments. Finally, we conclude with a discussion of the theoretical and practical implications of our findings.

## Related Work

### Long Haulers

Studies have focused on Long Haulers with hopes to understand their symptoms and concerns [14–16]. By analyzing Reddit posts, [15] found that discussions revolved around *symptoms, diagnostic concerns, broad health concerns, chronicity, support, identity, and anxiety*. In [17] news articles and videos were selected from a news media platform, Google News. They were then analyzed to identify common symptoms that appeared in Long Covid related content [17]. The authors found that 41% of news reports mentioned the *duration* of the symptoms which tended to range between 1 month to over a year. *Tiredness* and *fatigue* were the most mentioned symptoms, occurring in 74% of the news content. Though insightful, these studies do not focus solely on the YouTube platform wherein there can be interaction between the creators of long form content and those consuming their media.

### Health on YouTube

YouTube is a platform that motivates users to create, publish and comment on posts [18]. It has been developed to handle long form content. YouTube is unique in that creators of long form content can not only share their videos but engage with viewers within the comment section. As of April 2022, a report from



Statista estimates that YouTube has 247 million users in the US [19]. There have been studies in which researchers analyzed YouTube comments and transcripts to understand public sentiment on health-related matters. These studies have used either manual [20,21] or NLP-based [22,23] approaches.

Non-computational analyses of YouTube videos involved manually coding videos into various groupings. One study on anorexia-related YouTube videos employed the help of three physicians to classify 140 videos against a predetermined list of classification criteria [20]. Additionally, to understand discourse on YouTube videos that seeks to de-stigmatized mental health, [21] manually coded 100 randomly selected comments from 20 videos based on predetermined coding criteria.

In contrast, Aslam et al. used computational methods to understand the transcripts of 1000 COVID-19 related YouTube videos [16]. They used Gensim LDA topic model to understand the transcripts. They found that salient topics involved *symptoms*, *precautions*, and *home remedies* [22]. In their study, [23] fine-tuned the RoBERTA base to label comments from factual and misinformative COVID-19 related videos. Additionally, they extracted features from video titles and comments [23]. These features were used in a linear support vector machine (SVM) to detect misinformative videos [23].

We collected YouTube transcripts and comments between August 3, 2020 and October 29, 2021 to investigate Long Covid symptomatology and other related complications. We chose to use computational methods, more specifically topic modeling, because it can capture a wider distribution than manual studies [10]. To the best of our knowledge, this is the first study to examine YouTube video transcripts and comments related to Long Covid experiences. Our goals for this study were two-fold:

- **RQ1:** What types of videos are available on YouTube regarding Long Covid?
- **RQ2:** How do users respond to Long Covid content?

## Methods

### Data Collection

YouTube is a free to use social media platform, that has been adopted by individuals, organizations, and specialized professionals from various fields to share relevant and important information [18]. Because of this, we deemed YouTube to be a good source of data to capture videos uploaded by different types of content creators. This allows for diversity in our dataset.

We used Google's API, `googleapiclient.discovery`, to capture video comments and metadata (e.g. number of comment likes and responses). 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."

The search terms were found by first inspecting Covid Long Hauler-related news articles to find pertinent keywords. After this, Google Trends was inspected to see if there were any additional terms or versions of terms that had already been identified. We used these terms to find and inspect an initial list of videos

on YouTube. After completing this process, we were able to rule out the term “Longhauler” as many references were not related to COVID-19. The resulting videos were in the date range August 3, 2020 and October 29, 2021. The videos were collected on November 1, 2021. After removal of duplicates and irrelevant videos, we collected 152 videos.

We used the python package *YouTubeTranscriptAPI* to capture video transcripts from the videos. 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.

We then manually grouped videos based on the video source as previously done in a similar study [24]]. This is because the topic coverage of videos can vary widely depending on the source of the video. The resulting groups were News source, Medical source and Long Haulers. News source videos were those that were uploaded by news entities including local, national and 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. We collected 49 such videos. The last 25 videos belonged to the Long Hauler grouping which represented first-person accounts of those who considered themselves to have Long Covid. From these videos, we collected 2845 comments total: 1258 comments associated with the Medical source; 1078 comments were from News source videos; 509 comments were from Long Haulers.

We only analyzed publicly available documents in this study and did not analyze identifiable private information or involve any direct or indirect interactions with individuals. Per [Blank for review] policy (citation: 45 CFR 46 Definitions), the study is exempt from IRB review requirements because it does not meet the regulatory definitions of Human Subjects research. However, we removed any user identifiable information (e.g. usernames) and paraphrased or modified comments to preserve user pseudonymity while maintaining the content’s integrity in the manuscript.

### **RQ1 Methods: What types of videos are available on YouTube regarding Long Covid?**

To understand themes within the video content, we used Biterm to generate topics of video transcripts as well. Biterm Topic Model learns topics by modeling the generation of word co-occurrence patterns in the whole documents to counter the sparse word co-occurrence pattern problem that occurs when evaluating at the document level [25]. Each group, Medical, News and Long Haulers were processed individually to preserve our groupings. Biterm was created with shorter social media texts in mind, given that they are usually much shorter than standard document sizes [25]. Because video transcripts were considerably longer thus it could contain multiple topics, chronological batches of 50 consecutive words were fed into each model as suggested by previous work on topic modeling [26]. It was important to divide the transcripts into shorter portions so that more specific topics would be generated. After pre-processing our data by lemmatizing words and removing stopwords, we fed our data into the topic models. To preserve our groupings, we created six separate models: one positive and one negative model for each News source, Medical source and Long Haulers. When fine tuning the number of topics, we tested four

numbers (3, 5, 10, 15). For each number, we assessed the coherence scores, strength for words within the same topic co-occurring in the same documents [25]. Biterm adopted a coherence score proposed by [27]. In [25], the average coherence score for a Biterm model with 5 topics was between -52.3 and -52.5. A limitation of coherence score is that it only accounts the most frequent topic words. To compensate for this limitation, we complement the evaluation with manual analysis in addition to considering coherence scores for selecting the most cohesive model. To elicit unknown, emerging themes grounded in the labelled topics, we further qualitatively analyze following an open coding procedure [28] similar to a previous study that analyzed social media content that includes YouTube on COVID-19 [29]. Following the collaborative identification of a list of topic labels, the research team independently labeled each topic using up to 50 most salient terms and up to 30 samples of the most representative content, followed by grouping into themes. At each iteration, the research team resolved any discrepancies through discussion.

## **RQ 2 Methods: How do users respond to Long Covid content?**

We conducted sentiment analysis to understand public sentiment with regards to the delivered content. We used Valence Aware Dictionary for sEntiment Reasoning (VADER) [30] to determine sentiment of video comments. VADER is a rule-based model for sentiment analysis. It was created specifically for social media contexts as it can recognize slang and emojis. It produces positive, negative, neutral and compound scores for each respective body of text by summing the valence scores of each word and normalizing them to be between -1 and 1. We chose VADER in lieu of other sentiment analysis tools, such as AFINN, BING or NRC because VADER was specifically developed for analyzing social media texts.

We used the compound score as the overall sentiment score for the comment. Positive comments included all comments with a compound score greater than zero. Negative comments included all comments less than or equal to zero, following methodological guidance of a previous study [31]. This process was completed independently for each group, Medical source, News source and Long Haulers.

After we created positive and negative subgroups of our comments, we created topic models to understand the thematic make up of positive and negative comments with regards to each respective group. We separated comments into positive and negative subgroupings prior to generating topics so that our resulting topic models would be more cohesive. Similar to RQ1 methods, we used Biterm to generate topics, and manual review to label topics and grouping them according to themes. Because comments are relatively short in length and typically have 1 topic, we used the entire comment as a document.

## **Results**

We organized the resulting topics into 28 themes across all sources. Medical Transcripts themes were *Explanations in Layman's Terms*, *Show Housekeeping* and *Biological Explanations*. News Transcripts themes were *Sharing Patient Experiences*, *Negative Experiences*, *Experts Weighing In*, and *Handling the Long*

*Haul*. Long Hauler Transcripts themes were *Taking Treatments into Own Hands*, *Changes to Daily Life*, and *Choosing Homeopathy over Pharmaceuticals*. Positive News Comments themes were *Extending Empathy*, *Expressing Distrust through Sarcasm*, and *Encouragement for Better Outcomes*. News source videos received the highest proportion of negative comments. Negative News Comment themes were *Reproduction of Debunked and Political Theories*, *Misinformation/Disinformation*, and *Issues with Healthcare System*. In contrast, Medical source comments received the highest proportion of positive comments. Positive Medical Comment themes were *Appreciation of Helpful Content*, *Hope and Encouragement*, and *Exchange of Helpful Information*. Negative Medical Comment themes were *Negative Impacts of Long Haul*, *Requiring Medical Alternatives*, and *Lack of Needs*. Positive Long Hauler Comment themes were *Appreciation*, *Exchange of Helpful Information*, and *Community Building*. Negative Long Hauler Comments were *Exchange of Additional Information*, *Disillusionment of Healthcare System*, and *Requiring More Visibility*.

### **RQ1 Results: What types of videos are available on YouTube regarding Long Covid?**

We collected the transcripts from 152 videos that were divided into three groups (News source: 78, Medical source: 49, and Long Haulers: 25). Transcripts were divided into subgroups of 50 consecutive words and fed into distinct Biterm topic models. The below chart shows the breakdown of videos by source type.

#### **Medical Source Transcripts**

The Medical source transcripts were captions from videos created by an individual or organization in the medical sector. This included doctors, medical insurance companies and medical schools.

Table 1. Medical Source Transcript Results.

<b>Theme</b>	<b>Topic Label</b>	<b>Keywords</b>	<b>Sample Transcripts</b>
<b>Explanations in Layman's Terms</b>	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.
	Symptom Etiology	<i>syndrome, severe, illness, chronic, different</i>	[...] 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>	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. [...]
<b>Show Housekeeping</b>	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 [...]
	Encouraging the audience to keep in touch	question, talk, help, data, group, better, information	[...] 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.
<b>Biological Explanations</b>	Immunophenotyping	<i>ccr5, data, number, antigen, cd16, cd14, interleukin, dotted, chord, monocytic, interstitial</i>	[...] 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>	[...] 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 [...] so long covid actually has an immune response with 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". As implied by the theme title, the transcript snippets comprising each topic displayed scientific speech that was relatively easy for the general public to understand. The first topic, Symptomatology, covers video transcripts in which the speaker is explaining the symptoms associated with COVID. Some Medical source content creators dedicated entire videos to just a few symptoms, or a particular health system, as was the case in a video from

UAB Medicine dedicated to Long Covid and hair loss; “*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.*” The second topic “Symptom Etiology” featured transcript snippets that offered explanations of how Long Covid symptoms might have originated. Lastly, “Symptom Management” featured transcript snippets wherein medical professionals offered potential treatments for symptoms.

*Show Housekeeping.* Show Housekeeping was another prevalent theme in Medical source video transcripts. Associated topics were: “Introducing the Show/Guest”, “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 offered links other social media platforms where they can continue the long hauler conversation with engagers.

*Biological Explanations.* In general, Biological Explanations represented transcript snippets that displayed more advanced scientific language than those shown in the Explanations in Layman’s Terms theme. 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 [32]. In one video, the speakers discussed using “*proprietary spark dyes*” which can be used for immunophenotyping [32]. 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.

### **News Source Transcripts**

The News source transcripts were the captions from News Media outlets. These outlets ranged from local to international audiences.

Table 2. News Source Caption Results

<b>Theme</b>	<b>Topic Label</b>	<b>Keywords</b>	<b>Sample Transcripts</b>
<b>Sharing Patient Experiences</b>	Symptoms	<i>patients, symptoms, life, understand, hair, feel, medical, sick, protein, heartbeat, health, doctors</i>	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.
	Treatments	<i>need, better, understand,</i>	[...] even though there’s not a magic pill yet, to cure a long COVID, at

		<i>doctors, months, trying, research, care, answers</i>	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.
<b>Negative Experiences</b>	Not being believed by others and doctors	<i>symptoms, covid, virus, physician, dr, feeling, need, progress, says</i>	[...] to those doctors that deny the existence of long covid that this thing of course it's really look at the science.
	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. The moment [...] 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.
<b>Experts weighing in</b>	Etiology of the disease	<i>effects, infection, different, virus, actually, research, seen, syndrome, persistent, fatigue</i>	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, group, body, covid19, physical, related</i>	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 [...]
<b>Handling the Long Haul</b>	Managing Symptoms	<i>test, hair, brain, doctor, fatigue, pain, disease, talk, common, exercise, need, home, levels</i>	[...] 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 [...]
	Handling Cardiac/Chest Problems Specifically	<i>oxygen, need, lung, blood, chest, pain, infection, shortness, ventilator, loss, pulmonary, complications, attack. Disease, breath</i>	[...] 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 [...]

*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 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. One Long Hauler 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 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 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 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.”*

### **Long Haulers Transcripts**

The Long Haulers transcripts were captions from individual content creators that talked directly about their own personal experiences with Long Covid.

Table 3. Long Hauler Caption Results

<b>Theme</b>	<b>Topic Label</b>	<b>Keywords</b>	<b>Selected Sample Transcripts</b>
<b>Taking</b>	Alternate	<i>new, health,</i>	Covid was my wake-up call to fix my



<b>Treatment into Own Hands</b>	Remedies	<i>try, better, care, fungus, changing, trusted, broken</i>	gut and ultimately fix my health I was declining I was already declining before covid I was getting weak [...]
	Dealing with Uncertainty	<i>changing, declining, shitty, new, life, work</i>	[...] 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 Covid 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 [...] I didn't think it would happen to me [...]
<b>Changes to Daily Life</b>	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.
	How Symptoms Interrupt Activities	<i>Day, symptoms, time, feel, bad, need, breath, overgrowth, taste, chronic, fever, life, nuts, sacrifice</i>	I had to stop eating eggs I recognize that eggs weren't agreeing with me anymore and [...] I was eating three eggs every day like that was you know that was a breakfast staple for me [...]
	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 [...]
<b>Choosing Homeopathy over Pharmaceuticals</b>	Use of CBD and THC	<i>gummies, high, work, try, started, need, help</i>	The cbd and thc gummies that I take to sleep at night [...] I just try to keep things as natural
	Turning Down Over the Counter Medicine	<i>deficiency, vitamin, blood, different, taking, bad</i>	[...] so naturally I assume that is still coronavirus so she encouraged me to take over the counter 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
--	--	--	--

*Taking Treatment into own Hands.* 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. In Uncertainty Management, 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 led 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 one woman 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 and person of color who felt she experienced 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. They mentioned that their 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 topics 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.*” Other Long Haulers shared their apprehension of using pharmaceutical drugs and mentioned turning to more natural options instead.

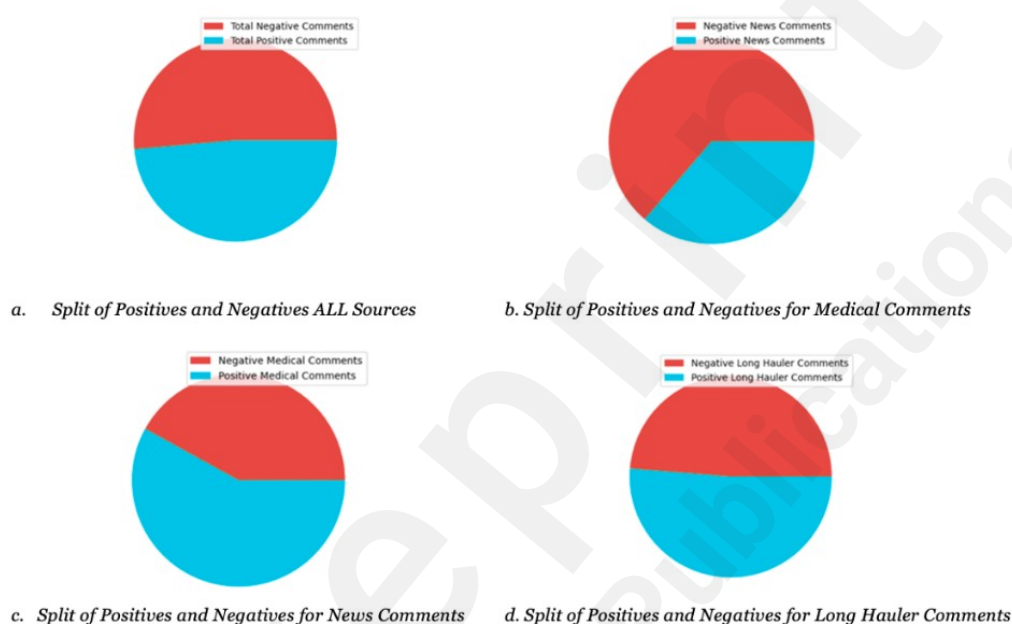
## **RQ2 Results: How do users respond to Long Covid content?**

To understand how users respond to Long Covid content, we separated comments for each respective category (News source, Medical source, and Long Haulers) into two subcategories based on sentiment (negative and positive). We then used Biterm to generate topics for these subcategories. The below figures (figures 2-5) show the breakdowns of positive and negative comments for our sources. When looking at all sources combined, there is not a large discrepancy between the number of positive and negative comments. Overall, there were 1463 positive comments and 1382 negative comments respectively.

However, when we begin to look at the split of positive and negative

comments by source, we can see that News sources received a greater share of negative comments. There were 687 negative comments and 391 positive comments. In contrast, Medical sources saw more positive comments than negative. There were 528 negative comments compared to 730 positive comments. Lastly, Long Hauler comments only showed a 13-point difference between number of positive and negative comments. There were 261 positive comments 248 negative comments.

Figure 1. Positive and Negative Splits of Comments.



In addition to capturing the comments themselves, we also captured metadata associated with the comments. This included comment replies, comment likes and video description. Comment likes and replies indicate the level of engagement that other YouTube users had with the comment posted. Medical source commenters saw an average of 16.02 likes per comment. The most liked comment received 602 likes. The most replied to comment received 474 replies. Conversely, News source commenters saw an average of 36.46 likes per comment. The most liked comment received 2520 likes. The most replied to comment received 184 replies. Lastly, Long Hauler video commenters saw an average of 54.55 likes per video. The most liked comment received 4127 likes. The most replied to comment received 223 replies.

Table 4. Comment Metadata

Source	Average No. Comment Likes	Most Likes on Comment	Average No. Comment Replies	Most Replies on Comment
Medical Source	16.02	602	3.15	474

<b>News Source</b>	36.46	2520	4.23	184
<b>Long Hauler</b>	54.55	4127	3.06	223

### News Source Comments

The following table represents the resulting topics and themes from positive comments found under News source videos.

Table 5. Results of Positive Comments in News Source

<b>Themes</b>	<b>Topic Label</b>	<b>Keywords</b>	<b>Sample Comments</b>
<b>Extending Empathy</b>	Relating to Others	people, get, think, symptoms, many, felt	omg, i can totally relate
	Well wishes	everyone, really, hope, take, care, able, want	It would be terrible to lose your ability to taste or smell. Here's to hoping they improve soon
	Gratitude	Better, still, hope, people, feel, heart, help, something, good	Your story was gut-wrenching, but still worth the share. Thank you. People need to hear this.
<b>Expressing Distrust Through Sarcasm</b>	Sarcasm	See, think, often, say, real, know	Okay so they survived a cold like most do. With a 99.8% survival rate, I'm sooo surprised.
<b>Encouragement For Better Outcomes</b>	Prayers and Scriptures	unto, shall, ye, people, peace, hath, forgive, love, reward, presence, pray, temple, holy	Phillippians 4:7 - And the peace of God, which surpasses all comprehension, will guard your hearts and your minds in Christ Jesus.
	Potential Solutions and Sharing Symptoms	Ask, receive, keep, know	Leronlimab is in clinical trials you guys. Don't worry, help is on the way.

*Extending Empathy.* Extending Empathy comprises of the topics "Relating to Others", "Well Wishes", and "Gratitude". Comments in which people Related to Others involved people explicitly sharing that they related to the content shown or explaining how their symptoms were similar to those interviewed in the news segments. For example, one commenter wrote *"You are not alone. I had COVID in April 2020 [...] I am currently in pulmonary rehab [...] I want others to know you are not alone. I'm praying for everyone. God Bless."* Well Wishes was the second topic in this theme. In this topic commenters sought to verbally empathize with those experiencing negative COVID-related symptoms; *"Too bad for that young man, hopes he gets better!"* Lastly, in Gratitude, commenters

were also grateful that Long Covid content was being shared at all; *“So glad she is sharing her struggles.”*

*Expressing Distrust through Sarcasm.* Although the comments observed in this analysis were rated neutral or positive by Vader, some comments seemed to take on a sarcastic tone. For example, one commenter wrote “The greatest nation in the world is your imagoNATION.” These sarcastic comments often appeared to exhibit political or skeptical undertones.

*Encouragement for Better Outcomes.* The topics comprising Encouragement for Better Outcomes were “Prayers and Scriptures” and “Potential Solutions.” Many commenters left prayers and extensive Bible verses underneath videos as a form of encouragement for those battling Long Covid; *“God heal these people from this virus. Give them strength.”* Lastly, Potential Solutions and Sharing Symptoms was a topic that covered suggestions that commenters made to improve the symptoms of those dealing with Long Covid as well as sharing symptomatology in general; *“Leronlimab is in clinical trials you guys. Don’t worry, help is on the way.”*

The following table represents the resulting topics and themes from negative comments found under News source videos.

Table 6. Results of Negative Comments in News Source

<b>Themes</b>	<b>Topic Label</b>	<b>Keywords</b>	<b>Sample Comments</b>
<b>Reproduction Of Debunked and Political Theories</b>	Conspiracy Theories	vaccine, face, different, affected, system, situation, avoid, dreadful, resist, corona	This is all because of 5G poisoning
	Political Influences	capability, dreadful, never, responsible, fight, answer	They got their butts kicked by Kung flu.
<b>Misinformation/Disinformation</b>	Fear of Impending Doom	never, stress, know, affected, death, worry, stop	Something’s coming, and we won’t be able to stop it.
	Skepticism/Rationalization	already, must, know, affected, response, another,	Elderly people are susceptible to viruses. This is well

		nothing, personal	known.
<b>Issues With Healthcare System</b>	Not Believed	medical, sick, normal, pain, feeling, never, anxiety, help, see, hope, think	My primary care physician doesn't believe me either [...]
	Other Illnesses	sick, heart, pain, long, time, blood, fatigue, brain, chronic, help	I had this for decades with me /cfs. Imagine dealing with it for that long [...]

**Reproduction of Debunked and Political Theories.** This theme consisted of two topics "Conspiracy Theories" and "Political Influences." In an example of Conspiracy theories topics, one commenter offered alternate causes of Long Haul symptoms which was reflected public disdain for mask-wearing "*'Long-haulers' may actually be suffering from effects of prolonged mask-wearing [...]*" instead of veritable information. In contrast, Political Influences covered suspected country and/or political involvement that contributed to the pandemic. When referring to individual damages lost due to Long Covid, one commenter wrote "*take the cost off the debt to china.*"

**Distrust of Information Shared.** This theme consisted of two topics "Fear of Impending Doom" and "Skepticism/Rationalization." Fear of Impending Doom consisted of comments that pointed to a grim future for Long Haulers and/or the general public, "*they're just trying to kill all the long haulers when all you need is some ivermectin [...]*." Skepticism and Rationalization consisted of commenters who were not convinced that the information presented on the Long Haul was veritable "*they had flu colds bacterial lung infections pneumonia, many caused by face mask, no sunlight, fear and confinement [...]*."

**Issues with Healthcare System.** This theme consisted of two topics. "Not Believed" covered comments condemning health care workers for dismissing the symptoms of their patients, "*typical doctor behavior: when in doubt, blame anxiety.*" Other Illnesses covered comments in which people drew similarities between Long Covid and other chronic illnesses. "*This is so real... the Lyme community feels all your pain. And being denied by Dr's that this is real. Its criminal to ignore this.*"

### **Medical Source Comments**

The following table represents the resulting topics and themes from positive comments found under Medical source videos.

Table 7. Results of Positive Comments in Medical Source

<b>Themes</b>	<b>Topic Label</b>	<b>Keywords</b>	<b>Sample Comments</b>
<b>Appreciation Of Helpful Content</b>	Gratitude	help, medical, doctors, hope, thank, much, understand, positive	This is the first thing that I have seen that explains anything besides the news trying to sensationalize and leave out important details.
	Health Literacy	need, information, understand, research, know, specific, narrative	Your lectures are always easy to understand. Thank you Dr.
<b>Hope And Encouragement</b>	Prayers	hope, believe, feeling, days	Jesus loves you [...]
	Voice of Reason	help, know, say, think, test, research	You've always cared, been of a sound mind, and shared such insightful information. Thank you.
	Bravery	research, doctor, video, help, know, people, feel good, positive, believe	Even though this subject is controversial, you're still brave enough to comment on it. Thank you
<b>Exchange Of Helpful Information</b>	Seeking Additional Information	symptoms, help, information, video, wonder, please	Has the Dr. released the additional information?
	Seeking Translated Information	please, videos	Can you please translate to Arabic?
	Sharing Helpful Information	think, information, help, vitamin, research, medical, may	Yesterday, I saw an article that said we needed to be aware of [...]

*Appreciation of Helpful Content.* This theme covers two topics "Gratitude" and "Health Literacy." Gratitude covered general professions of thanks for the content shown. One commenter wrote "Dr. Hansen, this is exactly the information I was hoping for! Thank you." Health Literacy in this case was covered in a positive light. Commenters thanked content makers for presenting information in a clear manner, "as a lay person with zero medical background, I learn a lot."

*Hope and Encouragement.* This included three topics "Prayers", "Voice of Reason" and "Bravery." Prayers included well wishes for those dealing with the Long Haul or reading the comment section. This included requesting prayers as well "Please pray for my mom... she is positive for covid 19." The Voice of Reason topic alluded to the idea that commenters deemed it important to find useful and truthful information, "Thank you for your commitment to keeping the world informed." Lastly, Bravery featured comments that alluded to the negativity that those sharing about Long Haul and, more generally, Covid-19

face. One commenter noted, *“this subject is controversial and you’re still brave enough to comment on it.”*

*Exchange of Helpful Information.* This theme covered three topics “Seeking Additional Information”, “Seeking Translated Information” and “Sharing Helpful Information.” Seeking Additional Information featured those primarily asking questions such as, *“What about cutaneous hyperesthesia?”* In Seeking Translated Information, many sought to understand content by having it translated into their native language. In Sharing Helpful Information, commenters tried to share what they deemed to be helpful to others *“Find a hyperbaric oxygen therapy chamber and a doctor checkup for compassionate use.”*

The following table represents the resulting topics and themes from negative comments found under Medical source videos.

Table 8. Results of Negative Comments in Medical Source

<b>Themes</b>	<b>Topic Label</b>	<b>Keywords</b>	<b>Sample Comment</b>
<b>Negative Impacts Of The Long Haul</b>	Comorbidity	fatigue, disease, symptoms, pain, chronic, heart, brain, chest, feel, body, diagnosed	How does this effect those with diabetes. I’m experiencing a range of symptoms.
	Loss	family, end, life, months, never	COVID-19 took my mom last year. I don’t know how I’ll move on.
	Symptoms	symptoms, fatigue, disease, chest, heart, brain, taste, body, severe, hearing, memory	I had a headache so bad that I had to seek treatment [...]
<b>Requiring Medical Alternatives</b>	Criticism of Doctors	doctor, bad, know, experience, need, study, poor, data, must	These doctors have no idea what they’re doing. His advice makes no sense. I think we’ll be sick forever.
	Debunked Recommendations	study, poor, suffering, need, research, must, know	How could you share so much but not talk about Ivermectin? You’re doing everyone an injustice.
	Misinformation	vaccine, last, illness, death, bad	Misinformation has gotten so bad that my own family won’t even believe me [...]
<b>Lack Of Needs</b>	Lack of Improvement	symptoms, feel, since, weeks, pain, back, effects, suffering, last, vaccination	The vaccine didn’t improve my symptoms
	Lack of Information	medical, would,	He mentions



		think, cause, whether, help	promising treatments but he never tells us what they are.
--	--	-----------------------------	---

*Negative Impacts of the Long Haul.* This theme consisted of three topics “Comorbidity”, “Loss” and “Symptoms.” Comorbidity featured comments and questions that sought to relate Long Covid to other diseases, *“Childhood obesity might be a factor [...]”* In Loss, some commenters spoke explicitly about those they lost to Long Covid or COVID-19. Lastly, in Symptoms commenters spoke candidly about the symptoms they face, *“I had a headache so bad that I had to seek treatment”*.

*Requiring Medical Alternatives.* In this theme there are three categories “Criticism of Doctors”, “Debunked Recommendations” and “Misinformation.” In Criticism of Doctors, commenters spoke about how they often felt dismissed by doctors when presenting their symptoms, *“if he went to visit my gp he would tell him he was stressed and it was in his head told me the same [...] it turned out to be lung scarring and a tumor.”* In Debunked Recommendations, commenters pushed for the use of medications that had already been proven not helpful and even toxic for human consumption. Ivermectin was notably one of these medications, *“should we be taking Ivermectin since our DNA now expresses spike protein forever?”* Lastly, Misinformation comments reverberated common anti-mask and anti-vaccine comments *“John do you have the list of ingredients of the vaccines? My daughter makes cupcakes and she has to list every ingredient by law...”*

*Lack of Needs.* This theme covers “Lack of Improvement” and “Lack of Information”. Lack of Improvement largely relates to symptoms not improving despite medical and home remedy attempts. Lack of Information included criticism of content sources for not providing enough information regarding content (e.g. treatment, research), *“he mentions promising treatments, but he never tells us what they are”*.

### Long Haulers Comments

The following table represents the resulting topics and themes from positive comments found under Long Hauler videos.

Table 9. Results of Positive Comments in Long Hauler Source

Themes	Topic Label	Keywords	Sample Comment
<b>Appreciation</b>	Bravery	sharing, thank, glad, feeling, believe, recovery, care, story, bless, post	Your bravery hasn't gone unnoticed. Thank you for all that you do.
	Compliments	bless, share, thank, good, feel, positive, love, believe	what a beautiful person inside and out.
<b>Exchange</b>	<b>Of</b> Seeking Additional	back, still, help,	eliminating carbs

<b>Helpful Information</b>	Information	know, could, think, test, say, check, treatment, better, work, natural, support, right	could potentially make things better. That's what worked for me [...]
	Sharing Additional Information	support, scheduling, doctor, right, treatment, better, work, great, different, try, may, take	If you check my channel you'll see why you should check your CRP. It could really help your lungs [...]
<b>Community Building</b>	Reaching Out	need, people, help, sharing, video, time, want, research, appreciate, experience	is there any way that I can talk to you please or message you?"

*Appreciation.* The Appreciation theme consisted of “Bravery” and general “Compliments”. Commenters lauded the content creator for being brave enough to share their experiences. This may allude to the idea that some who speak on their Long Covid experiences may face backlash. Additionally, commenters also gave content creators various accolades regarding their personalities and their decisions to share information, *“what a beautiful person, inside and out”*.

*Exchange of Helpful Information.* “Seeking Additional Information” and “Sharing Additional Information” were two topics in this theme. Commenters often initiated or tried to engage in dialogue about topics such as potential treatments and tests for Long Covid symptoms, *“if you check my channel, you'll see why you need to check your CRP”*.

*Community Building.* “Reaching Out” was the topic in this theme. Commenters sought to connect with Long Haulers to continue conversations elsewhere.

The following table represents the resulting topics and themes from negative comments found under Long Hauler videos.

Table 10. Results of Negative Comments in Long Hauler Source

<b>Themes</b>	<b>Topic Label</b>	<b>Keywords</b>	<b>Sample Comment</b>
<b>Exchange Of Additional Information</b>	Asking Follow Up Questions	get, think, covid, would, take, symptoms, different, know, since, less, help, maybe	Are you still sick?
	Sharing Info via Experience	symptoms, many, swollen, frustrating, lymph, body, covid, chest, get, pain, take, help	You should check into your thyroid levels. I had issues with mine [...]

	Seeking Answers for Symptoms	help, feel, usually, maybe, less, symptoms	did anyone else experience long COVID anxiety?
<b>Disillusionment With Healthcare System</b>	Disappointment with Doctors	medical, pain, enough, nurse, support, right, hard, felt, know, fear, frustrating, people, deal	Overachievers will never admit they don't know something.
	Unfair Treatment	people, frustrating, deal, problem, felt, hard, know	women and women of color are often treated this way. i'm not really surprised.
<b>Requiring More Visibility</b>	Gratitude	symptoms, sick, since, without, feeling, believe, almost	My girlfriend has long COVID and she has so many of these issues.
	Wanting more Awareness	weeks, always, sick, bad, body, infection, low, feel, know, symptoms	I barely see any information like this in the media. Why is that?!

*Exchange of Additional Information.* “Asking Follow Up Questions” and “Sharing Information via Experience” and “Seeking Answers for Symptoms” were the three topics in this theme. This theme was very similar to the theme that appeared in the Positive Long Hauler Comments analysis. There were slight differences between the examples in the two themes. The theme in this instance focused more on symptomatology in the case of the content creators and/or the commenters, “*did anyone else experience long COVID anxiety?*”.

*Disillusionment with Healthcare System.* The topics in this theme were “Disappointment with Doctors” and “Unfair Treatment.” In “Disappointment with Doctors” commenters mainly criticized the behavior of doctors in the context of long COVID diagnosis and lack thereof. Additionally, in “Unfair Treatment” commenters mentioned how specific groups may experience worse health care treatment than others, “*female patients and patients of color [...] there is so much research on patients reporting doctors not believing them or not treating them with the same level of compassion.*”

*Requiring more Visibility.* This theme consisted of “Gratitude” and “Wanting more Awareness.” Interestingly, although these comments were marked as negative, there was still a number of comments that expressed gratitude for the content creator sharing their message. This was often accompanied by sharing of their experiences as well. Relatedly, “Wanting More Awareness” reflected the desire of commenters to see additional Long Covid content in the media, insinuating that there was not yet enough coverage.

## Discussion

## Principle Findings

Symptomatology was a prevalent theme across all sources. Video creators and commenters shared and empathized with each other regarding symptoms that occurred because of Long Covid. These symptoms included *prolonged fatigue, cognitive dysfunction, shortness of breath, cardiac issues and lingering pulmonary symptoms*. This was consistent with the findings of several studies [4,14–16,33]. In Medical source videos, medical professionals explained symptomatology and symptom etiology in both laymen's and more scientific terms. In both News source and Long Hauler videos, personal experiences were shared as well as how Long Covid symptoms had impacted their daily lives. Upon inspection of the comments, we found that symptoms were shared for a range of purposes. At times it was purely to exchange knowledge and offer informational support. Additionally, it was used as a means to connect with others to exchange emotional support [34].

## Emotional and Informational Support

The positive themes identified in our findings can be operationalized as emotional and informational support. The emotional support category of themes were those in which commenters or video creators sought to empathize with others. This was through words of encouragement, prayers, sharing of similar experiences and community building. Informational support themes covered themes in which users sought or shared information.

In both transcripts and comments, people discussed experiences of not being believed by doctors and having a perilous relationship with the healthcare system. This sentiment appeared to be common across the board, however three groups stood out in particular: those with other chronic illnesses, such as Chronic Fatigue Syndrome (CFS) and Myalgic Encephalomyelitis (ME), women and people of color (POC). Those who had been battling chronic disease for years prior to the emergence of Long Covid empathized with Long Haulers who felt that they were not being heard, as can be seen in table 6. Complaints centered around being told that they were overexaggerating their symptoms or insinuations that patients were hypochondriacs (table 6).

Women and people of color discussed how they felt dismissed by health care workers. There was a general sentiment of distrust. This notion has been backed by an NBC article, wherein one woman of color explained she had been brushed off by physicians and labeled as aggressive [35]. This was despite the fact that she had lost 30 pounds and sight in her right eye as a result of Long Covid [35]. People of color have been disproportionately affected by Long Covid [35–37]. Two studies conducted by the NIH [36,37] found that Hispanic and African Americans had greater health problems and symptoms related to Long Covid but were less likely to be diagnosed. This corroborates anecdotal evidence from video comments (table 10).

Though these topics occurred in comments with negative sentiment, there were positive repercussions: emotional and informational support. This general distrust of the healthcare symptoms appears to have led to the adoption of homeopathic medicine, alternative medicine and home remedies. In attempts to take their health into their own hands, users resorted to alternative treatments even if it put their freedom at risk. These comments were shared freely between video creators and commenters, exemplifying informational support. For

example, one commenter noted that they smuggled marijuana into their state and felt that their insomnia had improved as a result of consuming it (table 9). Others suggested changes in dietary habits (table 9).

Another aspect of informational support dealt with health literacy. Health Literacy was a theme that appeared most often in medical source related videos. Health Literacy has been defined by the Centers for Disease Control and Prevention as the degree to which individuals have the ability to understand and use information to form health related decisions [38]. The content from medical sources exhibited two distinct tones. In the first, information was delivered in layman's terms which would likely be easier for the average person to understand. In the second, scientists presented biological explanations of Long Covid in more jargon-filled language. There were mixed reactions. Commenters noted that, at times, they had issues understanding the content (table 8). Issues with health literacy can impede one's ability to properly advocate for themselves and to understand what their options are. In other instances, commenters thanked the medical professionals for explaining Long Covid in a digestible manner as can be seen in table 7.

Symptom management was another topic that came up often in Medical sources and Long Haulers transcripts and comments. In videos, medical professionals outlined steps that those with Long Covid could take to mitigate their symptoms (table 1). In the comments section of medical source videos, commenters shared helpful information as well, (table 7).

As implied, emotional support operationalized as comments that extended empathy and compassion. This could often be found when there were accounts of personal experiences. Bible verses were shared as a means of offering hope. Commenters also thanked creators for sharing their story and offered prayers (table 9). There was support from those living with other long-term illnesses, notably those with CFS or ME. Such discourse often led to community building in the comments section. This was particularly prevalent in the comments section of Long Hauler's videos. To continue discussion, commenters asked follow-up questions regarding the progression of symptoms (table 10). Additionally, they sought other avenues to connect with and support each other (table 9).

### ***Skepticism, Misinformation, and negatively charged comments for News and Medical Videos***

We also observed a high frequency of negatively charged content, particularly in the comments for news and medical videos. Skepticism regularly appeared in news related content. Theories suggested by prominent politicians, such as Ivermectin as a cure for Covid were abounded. Many also criticized the credibility of the news source and its supposed neutrality. News stations and reporters were at times labeled as people pushing liberal agendas and fear mongering propaganda.

Misinformation and Disinformation were major themes in both Medical source and News source videos and comments. Some commenters felt that doctors were not sharing correct information or misinterpreting the information that they had received (table 8). 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 (table 7).

Other negatively charged comments dealt with the lack of needs: lack of information, lack of visibility and lack of improvement. In general, commenters sought for more information from health care professionals (table 8). On a related note, commenters expressed wishes for more visibility regarding Long Covid. Commenters noted that their symptoms did not improve even once given the vaccine.

Based on our sentiment analysis, News source received by far the greatest proportion of negative comments (figure 1). When assessing the topics and themes that came up in comments under News videos, criticism and sharing of misinformation was dominant. 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 was shifted onto China and liberal politicians. Vaccine hesitancy and opposition expressed by commenters was reverberated by some politicians as well. Some commenters appeared to experience extreme fear with reference to the vaccine. They mentioned that those administering the vaccine and treating Long Haulers had motives to kill (table 6). This seems to shed light on the idea that, although many previously debunked sentiments of politicians are being repeated, there is a genuine fear of vaccines, the healthcare system and some members of the government. The sentiment analysis of videos by Medical sources revealed that only a smaller portion (42%) of the comments were negative. There were 528.

## Implications

Results of this study could help public health agencies, policymakers, organizations and health researchers to understand symptomatology and experiences related to Long Covid. The information includes a description of the diverse range of symptoms and informational and emotional needs of Long Covid patients. This information can help public health professionals to develop and implement effective interventions to manage Long Covid. Voices of Long Covid [39] is one campaign, promoted by the U.S. Department of Health and Human Services, that has emerged in November 2021 as a community for those with the syndrome. In addition to providing a forum for Long Covid patients to share their experiences, the campaign offers resources for vaccinations and updates on developing research. The findings of this study demonstrate the potential of computational analysis of social media to provide insights and communication strategies into the public's responses to future health crises. This information can be used to add additional perspective and information to such campaigns.

As referenced in the NBC News article [35], there are Long Covid patients have been met with resistance by some medical professionals. For example, one patient felt that she was dismissed after explaining her Long Covid symptoms. This dilemma led to the creation of Long Hauler support groups on various social media platforms [35]. By mining YouTube, a rich source of our daily experiences, we began to uncover multifaceted challenges faced by Long Haulers. Our findings align with experiences of patients who have lost work due to Long Covid and are unable to receive insurance coverage.

## Limitations and Future Work

There are some limitations within this work. Our study was performed on YouTube transcripts. In many cases, transcripts on YouTube videos are automatically generated. This means that the captioning process is imperfect and at times, incorrect words were recorded instead of words that the speaker said.

Additionally, we only reviewed top level comments related to our videos and our analyses on comments does not reflect the full scope of the discourse in the comment section. Thus, we may be missing important insights from responses to the comments. Future studies should extend this study to include reactions to comments as well. Another limitation is that we cannot assume that the comments present underneath the videos in our study are representative of all viewers. Many viewers do not comment on videos [40], thus their opinions are not captured.

It is difficult to detect sarcasm and linguistic nuances using LDA and sentiment analysis. Despite this, sarcasm is often used commonly in everyday speech. Because of this, the computational models may have interpreted some texts that was different than originally intended.

Future research could be focused on longitudinal experience of Long Haulers to examine how they are being perceived and their overall experience over time. Long Hauler sentiments towards the healthcare system and doctors could potentially have changed over time. Additionally, as more information has surfaced and more COVID-19 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 experience. Longitudinal studies would be able to explore this shift in their experience. Future research can explore the effectiveness of various public health strategies in mitigating the impact of Long Covid, taking into account potential changes in public awareness and understanding fostered by increased media coverage, including YouTube.

Regarding recent Long Covid treatments, we started our research before drugs like Paxlovid received full FDA approval on November, 2023 [41]. We collected the videos on November 1, 2021, which included videos made in August, after the spread of COVID-19 and until October 2021. A future study should investigate how the availability of Long Covid treatments changed the perceptions, management, and psychological impact of long covid.

It is important to acknowledge that the commenters and video creators in our YouTube study may be subjective to selection bias and excluded a certain geographic and demographic perspectives. These perspectives hold some weight in how public sentiment should be perceived [42–45]. However, over 95% of the internet population spanning across 88 countries regularly interacts with YouTube [46]. This highlights the potential for an opportunity for broader exploration.

## Conclusions

In this study we used topic modeling to investigate the videos concerning Long Covid were present on YouTube. Additionally, we assessed public responses to these videos by analyzing the comments section using sentiment analysis and

topic modeling. We found that videos mostly focused on symptomatology, potential treatments and sharing experiences. There were a range of response types, with News sources receiving the highest proportion of negative comments and Medical sources receiving the least. Some were negative and often referenced conspiracy theories and distrust of the shared content. They also included negative experiences regarding Long Covid symptoms and treatment. Positive comments were those that exhibited community building, sharing of information and offering of support. This information, which is based on social media analyses, can assist public health professionals in comprehending the responses to Long Covid. The information includes a description of the diverse range of symptoms and informational and emotional needs of Long Covid patients. This information can help public health professionals to develop and implement effective interventions to manage Long Covid. The findings of this study demonstrate the potential of computational analysis of social media to provide insights and communication strategies into the public's responses to future health crises.

## References

1. 60 Minutes. 'Post-acute covid-19 syndrome': COVID 'long-haulers' suffering symptoms months after initial diagnosis. YouTube. United States: 60 Minutes; 2020. Available from: <https://www.youtube.com/watch?v=0gLmMPOHDwM> [accessed Apr 21, 2021]
2. Cutler DM. The Costs of Long COVID. JAMA Health Forum 2022 May 12;3(5):e221809. doi: 10.1001/jamahealthforum.2022.1809
3. Rubin R. As Their Numbers Grow, COVID-19 "Long Haulers" Stump Experts. JAMA 2020 Oct 13;324(14):1381. doi: 10.1001/jama.2020.17709
4. Wisk LE, Nichol G, Elmore JG. Toward Unbiased Evaluation of Postacute Sequelae of SARS-CoV-2 Infection: Challenges and Solutions for the Long Haul Ahead. Ann Intern Med 2022 May;175(5):740-743. doi: 10.7326/M21-4664
5. Pires L, Reis C, Mesquita Facão AR, Moniri A, Marreiros A, Drummond M, Berger-Estilita J. Fatigue and Mental Illness Symptoms in Long COVID: Protocol for a Prospective Cohort Multicenter Observational Study. JMIR Res Protoc 2024 Jan 19;13:e51820. doi: 10.2196/51820
6. Berkeley Lovelace Jr. Long Covid patients face medical debt after insurance denies claims. NBC News. 2023.
7. Jacques ET, Basch CH, Park E, Kollia B, Barry E. Long Haul COVID-19 Videos on YouTube: Implications for Health Communication. J Community Health 2022 Aug 12;47(4):610-615. doi: 10.1007/s10900-022-01086-4
8. Oyeboode O, Ndulue C, Adib A, Mulchandani D, Suruliraj B, Orji FA, Chambers CT, Meier S, Orji R. Health, Psychosocial, and Social Issues Emanating From the COVID-19 Pandemic Based on Social Media Comments: Text Mining and Thematic Analysis Approach. JMIR Med Inform 2021 Apr 6;9(4):e22734. doi: 10.2196/22734
9. Melton CA, Olusanya OA, Ammar N, Shaban-Nejad A. Public sentiment analysis and topic modeling regarding COVID-19 vaccines on the Reddit social media platform: A call to action for strengthening vaccine confidence. J Infect Public



- Health 2021 Oct;14(10):1505–1512. doi: 10.1016/j.jiph.2021.08.010
10. Mutanga MB, Abayomi A. Tweeting on COVID-19 pandemic in South Africa: LDA-based topic modelling approach. *African Journal of Science, Technology, Innovation and Development* 2022 Jan 2;14(1):163–172. doi: 10.1080/20421338.2020.1817262
  11. Li C, Jordan A, Song J, Ge Y, Park A. A Novel Approach to Characterize State-level Food Environment and Predict Obesity Rate Using Social Media Data: Correlational Study. *J Med Internet Res* 2022 Dec 13;24(12):e39340. doi: 10.2196/39340
  12. Oduru T, Jordan A, Park A. Healthy vs. Unhealthy Food Images: Image Classification of Twitter Images. *Int J Environ Res Public Health* 2022 Jan 14;19(2):923. doi: 10.3390/ijerph19020923
  13. Jelodar H, Wang Y, Rabbani M, Ahmadi SBB, Boukela L, Zhao R, Larik RSA. A NLP framework based on meaningful latent-topic detection and sentiment analysis via fuzzy lattice reasoning on youtube comments. *Multimed Tools Appl* 2021 Jan 28;80(3):4155–4181. doi: 10.1007/s11042-020-09755-z
  14. Leviner S. Recognizing the Clinical Sequelae of COVID-19 in Adults: COVID-19 Long-Haulers. *The Journal for Nurse Practitioners* 2021 Sep;17(8):946–949. doi: 10.1016/j.nurpra.2021.05.003
  15. Thompson CM, Rhidenour KB, Blackburn KG, Barrett AK, Babu S. Using crowdsourced medicine to manage uncertainty on Reddit: The case of COVID-19 long-haulers. *Patient Educ Couns* 2022 Feb;105(2):322–330. doi: 10.1016/j.pec.2021.07.011
  16. Siegelman JN. Reflections of a COVID-19 Long Hauler. *JAMA* 2020 Nov 24;324(20):2031. doi: 10.1001/jama.2020.22130
  17. Basch CH, Park E, Kolia B, Quinones N. Online News Coverage of COVID-19 Long Haul Symptoms. *J Community Health* 2022 Apr 3;47(2):306–310. doi: 10.1007/s10900-021-01053-5
  18. Uddin SMJ, Albert A, Tamanna M, Alsharef A. YouTube as a source of information: early coverage of the COVID-19 pandemic in the context of the construction industry. *Construction Management and Economics* 2023 May 4;41(5):402–427. doi: 10.1080/01446193.2022.2162096
  19. L. Ceci. YouTube users by country 2022. Available from: <https://www.statista.com/statistics/280685/number-of-monthly-unique-youtube-users/> [accessed Jul 18, 2022]
  20. Syed-Abdul S, Fernandez-Luque L, Jian W-S, Li Y-C, Crain S, Hsu M-H, Wang Y-C, Khandregzen D, Chuluunbaatar E, Nguyen PA, Liou D-M. Misleading Health-Related Information Promoted Through Video Based Social Media: Anorexia on YouTube. *J Med Internet Res* 2013;15(2).
  21. McLellan A, Schmidt-Waselenchuk K, Duerksen K, Woodin E. Talking back to mental health stigma: An exploration of YouTube comments on anti-stigma videos. *Comput Human Behav* 2022 Jun;131:107214. doi: 10.1016/j.chb.2022.107214
  22. Aslam A Bin, Syed ZS, Khan MF, Baloch A, Syed MSS. Leveraging Natural Language Processing For Public Health Screening On YouTube: A COVID-19 Case Study. 2023 Jun 1;
  23. Serrano JCM, Papakyriakopoulos O, Hegelich S. NLP-based Feature Extraction for the Detection of COVID-19 Misinformation Videos on YouTube. *Proceedings of the 1st Workshop on NLP for COVID-19 Association for Computational Linguistics*;

- 2020.
24. Choi B, Kim H, Huh-Yoo J. Seeking Mental Health Support Among College Students in Video-Based Social Media: Content and Statistical Analysis of YouTube Videos. *JMIR Form Res* 2021 Nov 11;5(11):e31944. doi: 10.2196/31944
  25. Yan X, Guo J, Lan Y, Cheng X. A biterm topic model for short texts. *Proceedings of the 22nd international conference on World Wide Web - WWW '13* New York, New York, USA: ACM Press; 2013. p. 1445–1456. doi: 10.1145/2488388.2488514
  26. Schwarz C. Idagibbs: A command for topic modeling in Stata using latent Dirichlet allocation. *Stata Journal* 2018 Mar;18(1):101–117.
  27. Mimno D, Wallach HM, Talley E, Leenders M, McCallum A. Optimizing semantic coherence in topic models. *EMNLP '11: Proceedings of the Conference on Empirical Methods in Natural Language Processing Association for Computational Linguistics*; 2011. p. 262–272.
  28. Strauss A, Corbin J. *Basics of qualitative research*. Sage publications; 1990.
  29. Kwon S, Park A. Examining thematic and emotional differences across Twitter, Reddit, and YouTube: The case of COVID-19 vaccine side effects. *Comput Human Behav* 2023 Jul;144:107734. doi: 10.1016/j.chb.2023.107734
  30. Hutto C, Gilbert E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media* 2014 May 16;8(1):216–225. doi: 10.1609/icwsm.v8i1.14550
  31. Ahmed A, Aziz S, Khalifa M, Shah U, Hassan A, Abd-Alrazaq A, Househ M. Thematic Analysis on User Reviews for Depression and Anxiety Chatbot Apps: Machine Learning Approach. *JMIR Form Res* 2022 Mar 11;6(3):e27654. doi: 10.2196/27654
  32. Immunophenotyping. stanfordhealthcare.org.
  33. Indang Trihandini, Manendra Muhtar, Dea Allan Karunia Sakti, Chintya Putri Erlianti. Effects of long-haul COVID on the health-related quality of life among recovered hospitalized patients. 2022 Mar.
  34. Park A, Conway M. Harnessing Reddit to Understand the Written-Communication Challenges Experienced by Individuals With Mental Health Disorders: Analysis of Texts From Mental Health Communities. *J Med Internet Res* 2018;20(4):e121. PMID:29636316
  35. Bellamy C, Adams C. Black Covid long-haulers felt invisible to the health care system, so they formed their own support groups. *NBC News* 2022;
  36. Pfaff ER, Madlock-Brown C, Baratta JM, Bhatia A, Davis H, Girvin A, Hill E, Kelly E, Kostka K, Loomba J, McMurphy JA, Wong R, Bennett TD, Moffitt R, Chute CG, Haendel M. Coding long COVID: characterizing a new disease through an ICD-10 lens. *BMC Med* 2023 Feb 16;21(1):58. doi: 10.1186/s12916-023-02737-6
  37. Khullar D, Zhang Y, Zang C, Xu Z, Wang F, Weiner MG, Carton TW, Rothman RL, Block JP, Kaushal R. Racial/Ethnic Disparities in Post-acute Sequelae of SARS-CoV-2 Infection in New York: an EHR-Based Cohort Study from the RECOVER Program. *J Gen Intern Med* 2023 Apr 16;38(5):1127–1136. doi: 10.1007/s11606-022-07997-1
  38. Health literacy: report of the Council on Scientific Affairs. Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. *JAMA* 1999 Feb 10;281(6):552–7. PMID:10022112
  39. Voices of Long Covid. Resolve to Save Lives. Available from: <https://voicesoflongcovid.org/> [accessed Mar 8, 2024]

40. Soukup PA. Looking at, through, and with YouTube. *Communication Research Trends* 2014;3-34.
41. Pfizer's PAXLOVID Receives FDA Approval for Adult Patients at High Risk of Progression to Severe COVID-19. Pfizer. 2023. Available from: <https://www.pfizer.com/news/press-release/press-release-detail/pfizers-paxlovidtm-receives-fda-approval-adult-patients> [accessed Mar 8, 2024]
42. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. *PLoS One* 2018 Jun 14;13(6):e0198857. doi: 10.1371/journal.pone.0198857
43. Mitchell L, Frank MR, Harris KD, Dodds PS, Danforth CM. The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place. *PLoS One* 2013 May 29;8(5):e64417. doi: 10.1371/journal.pone.0064417
44. Hussein E, Juneja P, Mitra T. Measuring Misinformation in Video Search Platforms: An Audit Study on YouTube. *Proc ACM Hum Comput Interact* 2020 May 28;4(CSCW1):1-27. doi: 10.1145/3392854
45. Gore RJ, Diallo S, Padilla J. You Are What You Tweet: Connecting the Geographic Variation in America's Obesity Rate to Twitter Content. *PLoS One* 2015;10(9):e0133505. PMID:26332588
46. Osman W, Mohamed F, Elhassan M, Shoufan A. Is YouTube a reliable source of health-related information? A systematic review. *BMC Med Educ* 2022 Dec 19;22(1):382. doi: 10.1186/s12909-022-03446-z

## Supplementary Files

## Figures

Positive and negative splits for comments.

