*Digital Media Analysis (MGMT 4084)*

*Guided By: Prof. Andy Ohemeng Asare*

**Sentiment Analysis on Public Opinions for Vaccines**

Dhwani Shah (101359323)

December 14, 2021

George Brown CollegE

500 Macpherson Ave, Toronto, ON M5R 1M3

# Part 1: Project Introduction

**Introduction**

As the COVID-19 pandemic has spread globally because of its highly contagious nature and several variations, public awareness of COVID-19 vaccine concerns grew. COVID-19 had about 235 million cases worldwide as of October 1, 2021. Around 211 million people had recovered from the sickness, whereas 4.8 million people had died. (Elflein, 2021) Since January 2020, scientists, and medical specialists all around the world have been producing and testing COVID-19 vaccinations; thus far, 16 vaccines have been licensed for use around the world. Vaccination progress has been hampered by skepticism, distrust, and disagreement. As a result, it's critical to know what public thinks about vaccination and, consequently, their willingness to get vaccinated. The World Health Organization has named vaccine hesitancy as one of the top 10 global health issues for 2019. In many countries, such fear, along with vaccination misinformation, has posed significant barriers to vaccinating a large enough proportion of the population to achieve herd immunity. (Yousefinaghani, Dara, Mubareka, Papadopoulos, & Sharif, 2021)

In this research we will explore public discourse about COVID-19 vaccines on social media to determine the subjects, overarching themes, and feelings around COVID-19 vaccines and immunization in public. By conducting this research, we aim to get a better picture of public opinion and concerns about COVID -19 vaccines, which will help us better inform public health education and campaigns aimed at increasing COVID- 19 vaccine acceptance. Furthermore, these findings may provide information that will support the promotion of other vaccines. The main objectives of this study are as follows:

1. What are the most common words associated with COVID-19 vaccines?
2. How do people feel about the COVID-19 vaccine?
3. What are the main points of contention regarding COVID-19 vaccines?

This study aims to provide answers to these questions by correctly classifying comments, reviews, or critics about COVID-19 vaccinations on social media platforms as positive, negative, or neutral.

The remainder of the paper is structured out as follows: The research approach, which includes data collection and cleaning, is presented in section 2. The study's empirical findings, including a word frequency, identified themes, sentiment analysis, cluster analysis, and the association between COVID-19 and public sentiments, are presented in Section 3. Section 4 discusses the findings and offers theoretical and practical advice and discusses the study's limitations as well as future research issues.

**Literature review**

### **A broader overview of the top 5 COVID-19 vaccines and its effectiveness.**

Efforts to prevent the spread of the infections caused by the covid virus and save lives began as early as March 11,2020. Vaccination is a way that proven to be successful in combating serious infections for the past 60 years and once scientists identified that SARS-CoV-2 is one of the main causes of COVID-19, many biotechnology firms began creating vaccines to combat the virus. Out of the many vaccines available in the world now, the following are the most widely used and effective:

#### Pfizer-BioNTech Covid 19 Vaccine

Pfizer BioNTech vaccine uses genetic material called messenger RNA encoding the virus’ spike protein to stimulate the immune system against the virus. Clinical trials of the vaccine in November 2020, showcased that it is safe and 95% effective. The US Food and Drug Administration (FDA) authorized the vaccine for emergency use in December 2020 and gave full approval in August 2021 for ages 16 and older. (Lewis, Moore, & McBratney, 2021)

The vaccine was found to be 90% effective in “real world conditions”. There were some doubts about the length of the vaccine’s effectiveness and researchers also proclaimed that other variants of the virus such as the Delta variant could blunt vaccine effectiveness further (Lewis, Moore, & McBratney, 2021). To combat the issues stated, FDA approved a booster shot for Pfizer, which aims to help control rising infection rates in populations across the world by creating more antibodies.

#### Moderna Covid-19 Vaccine

Moderna is another vaccine that utilizes messenger RNA in its vaccines. The vaccination was developed by a joint project between Moderna and the National Institute of Allergy and Infectious Diseases. Phase 1 trials of the vaccine were conducted in May 2020, and it quickly moved through Phase 2 and 3 trials. The vaccine highlighted effectiveness rate of 94.5% in its Phase 3 clinical trials conducted among 30,000 volunteers. The vaccine was found to be 90% effective in “real world conditions”. It received an Emergency Use Authorization (EUA) in December 2020 for people aged 18 and older. The FDA also had to authorize a booster shot for this vaccine since it was less effective against the Delta variant (Christ, 2021).

#### Johnson & Johnson Covid 19 Vaccine

J&J’s vaccine candidate Ad26.COV2.S uses S protein gene in an inactivated cold virus. The technology used in this vaccine is similar to the one used in making their Ebola vaccine and contrary to Pfizer and Moderna, J&J vaccine requires a single shot and can be stored in a standard refrigerator. Phase 1/2 trials for the vaccine began in mid-2020 and by September 2020, it concluded its phase 3 trials. The vaccine is found to be 85% effective in preventing COVID-19 complications. (Bai, 2021) The vaccine received authorization from FDA in February 2021 for people aged 18 and over. FDA has not yet approved a booster shot for the vaccine, but it is believed that it might be a requirement later in future.

The FDA paused the vaccine on April 13, 2021, after some people developed a serious blood clot combined with low platelet count within three weeks of receiving their dosage. After thorough review, the pause has been lifted and the vaccine became available again however coming with a warning. (Grady & Zimmer, 2021)

#### AstraZeneca and Oxford Vaccine

The vaccine developed by Oxford also used the S protein gene found in an inactivated common cold. The candidate named AZD1222, was found to be 76% effective in the phase 3 clinical trials conducted in US (Lewis, Moore, & McBratney, 2021). The US has still not authorized the vaccine for emergency use even though many countries in the world such as the UK are administering the vaccine for emergency use. European medicines agency reported similar blood clot cases as reported after the J&J vaccine, within two weeks of vaccination. The cost of the vaccine is less when compared to others and it requires only normal refrigeration for storage. (Eurepeon Medicines Agency, 2021)

#### Novavax Covid-19 vaccine

Novavax’s vaccine candidate named NVX-CoV2373 requires two doses of vaccine 21 days apart and is also a protein-based vaccine. In its clinical trials conducted in US and Mexico, the vaccine was found to be 90% effective against Covid-19. Site pain, muscle pain, fatigue and headache are common vaccine side effects. On top of its two required doses, Novavax also conducted trials in the US and Australia to test the effectiveness of a booster dose (Say & Crawford, 2021). This booster shot is to be taken six months after the two doses and it is said to be effective against all variants of the virus. An EUA from FDA was said to be approved for the vaccine in fall 2021. After receiving the EUA, it would deliver 100 million doses to the US government (Lewis, Moore, & McBratney, 2021).

### **Describe some of the issues surrounded with the top 5 COVID-19 vaccines and how the public reacted to them.**

There are different reasons for vaccine hesitancy. The most common of all are directly related to the vaccines themselves – whether the vaccines are safe, if there are any side effects, and how effective they are (Lin, Tu, & Beitsch, 2020). Further to common side effects including fever, fatigue, aches, chills and nauseas, there was also a known risks with Johnson & Johnson’s and AstraZeneca’s vaccine of causing blood-clotting disorders; In the meantime, while Pfizer-BioNTech and Moderna are the first ever mRNA vaccines in the market, some have expressed worries that they would alter human’s genetic information and pose long term health threats, albeit ungrounded (Jeong, 2021).

Such concerns are further amplified given the fast-tracked development of current COVID-19 vaccines, taking significantly less time compared to the average of 10-15 years for existing vaccines for other diseases (Davies, 2020). Although the development of COVID vaccines was legitimately accelerated with simultaneous trials and suspension of non-essential research, the public still shows inconfidence in the vaccine as a poll showed 75% of Americans were worried about the safety of such fast-tracked development, and 60% would rather wait than to take the first generation vaccine (Lin, Tu, & Beitsch, 2020). There are also rising concerns over the vaccines’ efficacy and duration of immunity as variants of the virus emerge (Jeong, 2021).

To tackle such public concerns, governments and public health agencies have a crucial role in explaining and communicating the benefits of the immunization program, but at times they might have delivered inconsistent or even contradicting information. According to Thunstrom, Ashworth, Finnoff and Newbold (2020), White House have been continuously communicating lower risk messages than the Center of Disease Control (CDC) did. This undermined public trust in the government and led to growing vaccine hesitancy.

Nonetheless, rather than learning about and understanding the development process and the proven efficacy of vaccines, many have instead received misinformation or even conspiracy theories, for example how the virus and the 5G network were related (Megget, 2020), and the conspiracy that the virus being a biological weapon created in China (Pennycook, McPhetres, Zhang, Lu, & Rand, 2020). Such misinformation, initiated from fear, could eventually lead to a viscous cycle of growing public fear and skepticism towards vaccines (Loomba, de Figueiredo, Patiek, de Graaf, & Larson, 2021).

There is also a general belief among the public that vaccines are not necessary (Lin, Tu, & Beitsch, 2020). While the virus shows a low infection-fatality rate, many do not see their loved ones getting critically ill, and they believe that they would survive without any significant health issues even if they unfortunately catch the disease (Davies, 2020). Other reported reasons for vaccine hesitancy include insufficient information, costs of self-financing vaccination programs and vaccine’s country of origin (Lin, Tu, & Beitsch, 2020).

### **Describe the COVID-19 immunization program**

COVID-19 vaccination programs have been off and running around the world. As of 9th October 2021, 47.6% of the global population have received at least one dose of vaccines (Holder, 2021). The vaccine co-developed by Oxford and AstraZeneca has the widest coverage of countries (182) followed by Pfizer-BioNTech (136), Moderna (76), Sinopharm (75), and Johnson & Johnson (55) (Holder, 2021).

Deals have been signed by government with vaccine producers to reserve vaccines for their people. It was reported that over 130 deals involving a at least 9.6 billion doses were signed as of March 2021, which would allow us to fully immunize over half of all the people in the world (Randall, Sam, Tartar, & Cannon, 2021). Wealthier countries, however, have enjoyed an advantage over less well-off countries when it came to securing vaccines from developers. A strong correlation is observed between the country’s income level and its vaccination rate (defined by proportion of population having received at least one dose), as per Holder (2021) as of 9th October 2021:

* High-income countries: vaccination rates range from UAE (96%), Portugal (87%) and Chile (84%), to Romania (31%), Trinidad & Tobago (42%) and Croatia (45%);
* Higher-mid income countries: range from Cuba (84%), China (79%) and Malaysia (76%), to Turkmenistan (0.5%), Gabon (4.9%) and Armenia (8.9%);
* Lower-mid income countries: range from Cambodia (82%), Mongolia (70%) and Sri Lanka (67%) to Tanzania (1%), Cameroon (1.5%) and Papua New Guinea (1.5%);
* Low-income countries: highest rates observed in Tajikistan (26%), Rwanda (17%) and Guinea (9.8%).

Geographically, Africa only has an overall vaccination rate of 7.1% as of 9th October 2021 (Holder, 2021), which has also drawn WHO’s attention as the third wave of the disease emerged on the continent in June (Felter, 2021).

There are more vaccines are under development, which would hopefully increase vaccine supplies. One of such would be Novavax, which has passed phase III trials and applied to the WHO for approval of emergency use on 23rd September 2021 (McGregor, 2021). It however still has to overcome regulatory concerns and manufacturing delays before it could be rolled out to the market (Zimmer, Corum, & Wee, 2021).

### **Describe at least 5 COVID-19 vaccine immunization challenges faced by public health agencies.**

#### Distribution Challenges

As more and more efficacious COVID-19 vaccinations are being introduced to the market, a set of practical and ethical challenges have emerged. Specifically, supply chain and distribution issues — from manufacturing to shipping to storage – are restricting the public's access to these life-saving vaccines. It is practically impossible to have people around the world vaccinated at the same time due to supply chain constraints (Yale School of Medicine, 2021). As a result, prioritizing who to vaccinate first must be deliberate and fair. Further to distribution, storing the vaccines in refrigerators after they arrive at the final distribution or administration site is a challenge, as many of them require extreme storage conditions (Petkov, 2021).

#### Hesitancy by health workers and the public

Detecting the hesitation of health workers is a big challenge for the public sector to overcome. One obvious source of this is vaccine misinformation, which has been inflamed by anti-vaccination groups and reinforced by the internet. Many workers are more cautious in their information consumption, yet they still have valid concerns. In the meantime, public segments such as pregnant women are afraid of the vaccine because they concluded there were insufficient evidence to support its safety for mothers and their unborn children, particularly for a vaccine that was hurried to market (Smith, 2021).

#### Ensuring Transparency and Integrity of Advisory Bodies

Another issue confronting public health is the lack of transparency and honesty in special advisory committees. It is critical to gain the public's trust. Many governments have set up such organizations to help enlighten public decision-making in the wake of the outbreak (OECD, 2021).

Governments and public health organizations have a responsibility to guarantee that information about research into treatments and vaccinations are being communicated in an open and thorough manner. Several companies published their clinical trial protocols during the COVID-19 vaccine development process, but the results of key trials were initially communicated in headlines and press releases with little detail, leading to speculation about the underlying data prior to publication and peer review (OECD, 2021).

#### Proactively Releasing Timely Information Regarding Vaccination

The COVID-19 pandemic has demonstrated how a lack of clear and timely information can lead to ambiguity in decision-making, as well as a general lack of faith in public health and governance. Having timely and granular open-source data on key issues like the number of people vaccinated, the number of doses administered, geographic coverage, and the number of people who experienced adverse reactions will make data analysis and dissemination easier in online trackers, news sites, and other places (OECD, 2021).

It is also critical for individuals outside government to have trust in the effectiveness of government vaccination programs and policies by releasing up-to-date information about vaccine procurement and funding in accordance with access to information laws (OECD, 2021).

### **Describe the possibilities for future COVID-19 vaccines immunization programs such as the booster shots.**

It seems like just yesterday that individuals were calculating when their immunization would totally protect them; now there is speculation that our safety may require another shot in the arm. The current vaccines are still effective against the new variations, especially in terms of preventing serious sickness that might necessitate hospitalization or result in death. However, if the virus evolves further and a more dangerous strain emerges, alterations might be made to the vaccine as **Booster Shots**, which are additional doses to achieve longer-lasting protective immunity (MacMillan, 2021). Furthermore, experts believed that **nasal antibody spray** could be the future of immunizations. Using a nasal spray to provide a new COVID-19 medication derived from small llama antibodies could be the future of giving vaccines and defending against infectious diseases. Also, the nasal nanobodies spray has several advantages over ordinary human antibodies injected with a needle, including ease of use and the possibility for a person to self-administer. The nasal spray is still in the experimental phase. However, the possibilities appear to be promising. (Lao, 2021)

# Part 2: Methodology

### **Overview**

To begin the analysis, we developed codes in Python language and collected the data from Twitter using different keywords. Once the data was collected, we used Python codes to prepare our data for analysis, which includes data cleaning, removing any unnecessary columns and words. The entire data collection, cleaning, and storing procedure is briefly explained below.

### **Data collection process and techniques**

#### Data gathering

Twitter is one of the most popular social media platforms in the world. Twitter users primarily use the service to share their views, opinions, and facts. Many researchers have gathered data from this site to investigate attitudes on a variety of social topics due to the large number of expressed views on this platform. In this research we collected tweets from the Twitter platform using the Tweepy package with Twitter API in Python language using Jupyter Notebook.

COVID-19 vaccines were considered to be officially introduced to the market on December 31, 2020, when the World Health Organization (WHO) issued its first-ever COVID-19 vaccine emergency validation for the Pfizer-BioNTech vaccine (World Health Organization, 2020). However, due to certain challenges we were able to collect data with Tweepy on two consecutive weeks from 5th to 19th November 2021.

#### Keywords

Below are the keywords which are used to extract data from Twitter.

* Vaccine
* Booster shots
* Covid-19
* Moderna
* Johnson & Johnson
* Janssen
* Pfizer
* BioNTech
* Comirnaty
* AstraZeneca
* Novavax

### **Data cleaning**

After fetching our raw data, steps have been taken to remove unnecessary content from the dataset and to prepare it for analysis. The process included removing unwanted contents, tokenization, and deduplication, with the use of Python language on Jupyter Notebook.

#### Removing unwanted contents

Many tweets include more information than needed for the analysis. For example, Twitter tags such as “RT” for retweets, “@” for mentioning other accounts, and “#” for hashtags, although useful for Twitter users to share and read content on the platform, does not carry any meaning for our analysis and topic modelling. With the re library imported, these Twitter tags have been removed using regular expressions. Other unwanted content being removed include hyperlinks, emojis (using the emoji library), pictures, and unwanted symbols like “amp” as translated from “&”.

#### Tokenization

After removing unwanted content, the results were changed to lowercase and were tokenized such that sentences were being split into words according to white spaces between words. This step was necessary for natural language processing so that the sentiments and topic modelling could be processed based on analyzing tokens. After tokenization, clean text was created by joining the tokens back using white space as separators.

#### Deduplication

Lastly, the cleaned results were being checked for duplication to see if there were any repeating or identical results. Only one occurrence should be kept for any repeating results because repeating content being posted would not contribute to our analysis or understanding of the sentiment or topics that were being discussed around the subject matter.

### **Data storage**

Results before and after data cleaning were saved as interim local data files in the format of comma-separated values (CSV). After doublechecking the data files using Excel, we imported the cleaned data into Jupyter Notebook and begin analyzing with Python packages.

# Part 3: Data Analysis

### **Overview**

Python was used to collect relevant Tweets over the course of two weeks due to its robust and customizable features. The code was used to generate the following analyses:

#### Word Frequency

It is important for us to understand what the most common words are being used when Twitter users post content about COVID-19 vaccines. As shown later in this report, a word cloud was generated using the word frequency data to visualize the most frequently used words.

#### Polarity and Subjectivity

Sentiment analysis is important for us to understand public perception towards vaccines. In the study, the TextBlob library was used to calculate subjectivity and polarity scores of each tweet. A subjectivity score ranges between 0 to 1, where 0 indicates that the content of the tweet is purely factual, while 1 means that it is entirely based on opinions. As for polarity, scores would be calculated on a scale of -1 to +1, with 0 meaning that the content is purely neutral, while a positive score denotes positive sentiment and vice versa.

#### Topic Modelling

Eight topics were identified by the code based on frequency of keywords used. Using manual interpretation, these topics were further classified based on what the actual meaning of the topic was.

### **Word Frequency**

As shown in *Figure 1*, the words “dose”, “children”, “adult”, “FDA”, “Authorization”, “month” and “age” had the highest word frequency.

The Food and Drug Administration (FDA) is in charge of authorizing many vaccines, hence the phrase is widely used. Some Phase 2 populations were vaccinated based on age and risk factors, therefore terms like “health”, “adult”, “children” and “eligible” refer to vaccine eligibility. As there were two shots from approved vaccines, dose and shots are commonly utilized, and many are also concerned about booster shots. This is owing to the fact that the majority of the tweets were about vaccines and public perceptions of vaccination as a result of COVID-19.

“Vaccinated”, “vaccines”, “Covid19”, and a variety of other terms are also used. However, since these words were used as search keywords, we did not consider them in the analysis process to avoid research bias.

A picture containing text

Description automatically generated

Figure 1 – Word Cloud

**Sentiment Analysis**

Of the 1,916 unique tweets analyzed, 615 (32%) were based on pure facts, and 1,462 (76%) had a subjectivity score of less than 0.5, indicating that Twitter users have refrained from expressing strong opinions towards the vaccines (see figure 2). This was also reflected in the distribution of polarity scores, where over half (1,048; 55%) fell within the range of -0.1 to +0.1, and 728 (38%) of tweets had a polarity score of 0 (see figure 3).

Figure 2 - Distribution of subjectivity scores

Figure 3 - Distribution of polarity scores

Despite the reservedness from expressing strong opinions, tweets have mostly been positive in terms of polarity. More specifically, around half (965; 50%) of the tweets analyzed were positive or partially positive, compared to only 233 (12%) tweets that showed negative or partially negative sentiment.

Figure 4 - Proportion of tweets by polarity

By comparing subjectivity against polarity of tweets, it has been observed that tweets, particularly those of positive polarity, have centered around the lower-middle part of the subjectivity scale, which indicated that while Twitter users have used fair amount of facts to support their views in favor of or against the vaccines.

Figure 5 - scatter plot: subjectivity against polarity

### **Topic Modelling and Sentiment Labelling**

The pyLDAvis interactive topic modelling visualization for the covid-19 vaccination sentiment for public opinion can be found by hovering over the first subject in the *Figure*-*6.* The red bar depicts the most common terms in each subject, and the chart reveals that there are 8 cluster themes associated with the sector. The blue bar shows the overall phrase frequency of that particular word across all topics. Also, any overlapping between clusters indicates similar terminology being shared across topics.

Chart

Description automatically generated

Figure 6 – Topic Modelling

|  |  |
| --- | --- |
| **Difficulties and Risks** | **Top 10 words for topic #0:**  ['adult', 'saturday', 'asks', 'disease', 'requested', 'regulators', 'church', 'doses', 'biontech', 'immune', 'prevent', 'eligibility', 'americans', 'expand', 'pfe', 'recent', 'help', 'died', 'requests', 'caused', 'icu', 'problems', 'germany', 'thailand', 'data', 'dies', 'heart', 'days', 'year', 'old'] |
| **Health and Safety** | **Top 10 words for topic #1:**  ['high', 'immunity', 'getting', 'months', 'like', 'biontech', 'breaking', 'infection', 'vaccinated', 'friday', 'mrna', 'protection', 'pending', 'dose', 'effective', 'doses', 'approval', 'study', 'boosters', 'people', 'just', 'risk', 'vaccines', 'authorized', 'shots', 'adults', 'cdc', 'fda', 'got', 'moderna'] |
| **Vaccine Distribution** | **Top 10 words for topic #2:**  ['adverse', 'weeks', 'countries', 'coronavirus', 'forecast', 'day', 'vaccines', 'age', 'group', 'approval', 'years', 'sales', 'clinics', 'pediatric', 'million', 'according', 'week', 'approved', 'received', 'shots', 'doses', 'cdc', 'said', 'billion', 'parents', 'biontech', 'ages', 'astrazeneca', 'children', 'kids'] |
| **Vaccine Shots** | **Top 10 words for topic #3:**  ['united', 'just', 'age', 'shots', 'states', 'expand', 'breaking', 'regulators', 'moderna', 'people', 'announced', 'americans', 'week', 'ages', 'expected', 'request', 'asked', 'tuesday', 'asks', 'food', 'drug', 'older', 'administration', 'use', 'emergency', 'biontech', 'authorize', 'authorization', 'adults', 'fda'] |
| **Vaccine Research** | **Top 10 words for topic #4:**  ['big', 'term', 'federal', 'clinic', 'people', 'chest', 'considered', 'care', 'moderna', 'years', 'adult', 'novavax', 'live', 'says', 'older', 'read', 'close', 'valuable', 'keeping', 'formula', 'rich', 'favorite', 'drugmaker', 'news', 'world', 'fully', 'adults', 'soon', 'vaccinated', 'eligible'] |
| **Eligibility for Vaccination** | **Top 10 words for topic #5:**  ['weeks', 'provide', 'today', 'vaccines', 'time', 'myocarditis', 'past', 'workers', 'got', 'adults', 'years', 'vaccination', 'janssen', 'dropped', 'receiving', 'doses', 'million', 'single', 'receive', 'older', 'ago', 'eligible', 'people', 'second', 'moderna', 'dose', 'received', 'months', 'amp', 'johnson’] |
| **Free Vaccination** | **Top 10 words for topic #6:**  ['young', 'day', 'results', 'centers', 'country', 'disease', 'tomorrow', 'getting', 'control', 'clinic', 'given', 'prevention', 'said', 'wednesday', 'got', 'city', 'make', 'vaccination', 'year', 'don', 'biontech', 'nov', 'says', 'kids', 'free', 'health', 'aged', 'receive', 'children', 'flu'] |
| **Vaccine types** | **Top 10 words for topic #7:**  ['news', 'visit', 'nov', 'clinic', 'drug', 'inflammation', 'spinal', 'regulator', 'type', 'europe', 'effect', 'rare', 'thursday', 'schedule', 'today', 'comirnaty', 'remended', 'authorizes', 'public', 'vaccines', 'shots', 'johnson', 'biontech', 'approved', 'use', 'moderna', 'amp', 'available', 'canada', 'health'] |

Further to grouping tweets into different topics, sentiment analysis was conducted again for each of the topics. As seen in figure 7, majority of tweets across the eight themes identified showed positive to neutral sentiment. The figure also illustrates that tweets related to vaccine eligibility and distribution have shown the highest proportion of negative sentiment, as there have been times when vaccines were unavailable for a specific number of people in numerous nations. In the meantime, tweets related to vaccine research and free vaccination generated the most positive sentiments.

Chart, bar chart

Description automatically generated

Figure 7 – Sentiment Analysis

# Part 4: Research Discussion

### **Overview**

The purpose of this study was to use the Twitter platform to investigate the thoughts and perceptions of individuals all around the world on vaccines for COVID-19. Because of its widespread use and ease of collection, Twitter was chosen for this study. To find helpful insights, a variety of analysis approaches were used, including world clouds, sentiment analysis, cluster analysis, correlations analysis, and trends analysis. ‘FDA, ‘doses’, ‘Covid-19’, ‘Adult, ‘Health’, ‘People, ‘Eligible, ‘Children, and ‘authorize’ were among the most frequently referenced words in the study. Furthermore, roughly half of the public's feelings were positive (50%), with the rest being neutral (38%) and negative (12%). Furthermore, the majority of the data in the tweets (68%) was based on opinions rather than facts (32%). This research supports the notion that people expressed positively about different vaccines.

### **Practical Recommendations**

Although our results showed that around half of the tweets showed positive sentiment towards vaccines, it is also reasonable to believe that users expressing positive opinions are more likely to have already been vaccinated. Herd immunity, however, requires the buy-in of the vast majority, as an estimate of 60-70% ofthe world population is required to have immunity against the virus in order to break the disease transmission (Swaminathan & Gupta-Smith, 2020), not to mention that vaccination rates vary across different countries. To improve the effectiveness of immunization programs, it is important to address concerns and improve public perception towards COVID-19 vaccines.

From the results in topic analysis, it was observed that the public was the most concerned about the eligibility issues related to vaccines, whereas research has received the most positive sentiment. Concerns with eligibility could be addressed by conducting more studies and research to restore trust and ensure safety of the vaccines for different groups of populations, such as minors, senior members, and patients with chronical illnesses. This also has to go hand in hand with consistent messaging from the governments and medical or public health professionals on the risk of vaccines, in contrary to what we saw in the White House and the CDC (Thunstrom, Ashworth, Finnoff, & Newbold, 2020).

Distribution of vaccines was another topic that received relatively higher proportions of negative sentiment, while polarized comments seen regarding free vaccines indicated that the public are also concerned with the accessibility of vaccines. Amid distribution challenges, governments, pharmacies, and international bodies like WHO should put continuous effort into distributing vaccines fairly to people around the world, and subsidizing vaccinations so that they are accessible to everyone.

With vaccination research being the topic that returned the most positive sentiment, governments and public health agencies should communicate information related to the benefits and risks of vaccines with supporting of research results to tackle vaccine hesitancy and improve public acceptance.

### **Study Limitations and Recommendations for Future Research**

#### Limitation: data collection

COVID-19 vaccines were considered to be officially introduced to the market on December 31, 2020, when the World Health Organization (WHO) issued its first-ever COVID-19 vaccine emergency validation for the Pfizer-BioNTech vaccine (World Health Organization, 2020). When designing the research project, the initial plan was to analyze tweets starting from January 2021, which unfortunately did not go as planned due to technical limitations.

With the desired analysis period in plan, the team started with using the Twint library for scraping tweets. There was however a bug with Twint where it could only return results from the past one day of the search. The team later switched to Tweepy for scraping tweets, which was only marginally better as it could return data from the past week. As such, the team was only able to run the codes again one week after the first run to collect two weeks of data in total for this research.

#### Recommendation: search query and keywords

Before scraping tweets around the research subject, a list of search keywords was prepared as the team saw relevant to the topic. Although the list was kept simple, keywords such as “COVID-19” as well as the most popular vaccines like “Moderna”, “AstraZeneca” and “Pfizer” were included, with which the team was confident such that these keywords would be sufficient to provide a fairly representative overview of COVID-19 vaccines as a subject. However, the team would recommend future studies to use a more extensive set of keywords to cover the topic more exhaustively, such as by including more variations of words and names (e.g. “COVID19”, “coronavirus”, “SARS-CoV-2”, etc.), including more vaccine producers rather than only including the mainstream ones (e.g. Sinovax, Sinopharm, Gamaleya, CanSino, etc.) (Zimmer, Corum, & Wee, 2021), as well as adding any newly developed vaccines at the time of research.

#### Recommendation: data collection and analysis tool

Given the challenges in using Python for data collection, the team would recommend, if resources permit, future studies to use paid social listening platforms that have official partnerships with Twitter, such as Talkwalker, Meltwater, or even Brandwatch which covers all the data up to the first ever tweet in 2006 (Twitter, n.d.). These platforms also often provide a guided set up for the search and support users with online resources and representatives, which would assist researchers who are less familiar with the tool.

#### Recommendation: additional analysis angles

Shall the limitation with the period of data collection be overcome, the team would strongly recommend covering data dated back to January 2021, as compared to just two weeks of tweets in this research, for a few reasons. Firstly, with data covering a more extensive period, the representativeness of the analysis and findings would be greatly improved, as compared to findings based on data from a relatively limited period where opinions are easily skewed due to updates and policies during that time. Secondly, while the vaccines were first rolled out to the market in late 2020 / early 2021, it is also reasonable to assume that the public would be more expressive with their opinions when vaccines are newly introduced to them. Another benefit of analyzing data over an extensive period is that it would allow us to understand any trends and changes in public sentiment on the subject over time, which could provide useful insights for experts, policy makers and healthcare professionals on the important factors influencing public perception on the ongoing immunization programs.

In addition, we would recommend future studies to investigate patterns by different subgroups. For example, as different countries have different policies and hence immunization programs, it would be interesting to look at the difference between them such as by sentiments, topics of discussion, and share of topics. It would also be meaningful to look into the difference in sentiment and perception towards the vaccines produced by different pharmacies and find patterns such as public preference of technologies behind vaccines (mRNA, deactivated virus, vector, etc.), and of pharmaceutical companies.

### **Conclusion**

The goal of this study was to look at public perceptions of vaccines. Around 3000 tweets were extracted using Python and pre-processed to uncover useful findings utilizing the Twitter platform. During these uncertain times, people have displayed positive attitudes regarding immunizations, according to the report. The findings of the study are consistent with prior research on the public's mixed attitudes to various vaccines as well as some of the risks linked with them. As a result, this research will assist health organizations, government agencies, individual groups, and other stakeholders in collaborating closely to make better-informed judgments on various vaccines while considering the views of all individuals.

### **Abstract**

The world had come to a standstill since the first cases of the COVID-19 virus were reported. Medical researchers around the globe undertook the task to develop vaccinations, in a bid to bring back the world to normalcy. The following research project aims to judge whether people trust and are willing to get their doses, since such factors can be major contributors to the success of vaccination drives. Data was gathered from popular social media platform Twitter for analyses such as word cloud, polarity and subjectivity, and topic modelling. “FDA”, “authorization”, “children”, “dose” were some of the most common words in discussions around COVID-19 vaccines, and the discussion was generally positive and supported by fair number of facts. Topic modelling analysis identified eight themes of discussion, and further sentiment analysis showcased that negative tweet were centered around vaccine distribution and eligibility whereas vaccine research and types garnered positive response. With growing use of social media, we would like to conclude by mentioning that this powerful tool should be utilized while taking decisions that affect public and, to clarify perceptions surrounding sensitive topics, to bring about transparency and harmony in the world.

# References

Lin, C., Tu, P., & Beitsch, L. M. (2020, December 30). Confidence and Receptivity for COVID-19 Vaccines: A Rapid Systematic Review. *Vaccines, 9*(1), https://doi.org/10.3390/vaccines9010016.

Davies, C. (2020, November 20). *How do we tackle vaccine hesitancy and effectively communicate vaccine safety to the general public in the COVID-19 era?* (The Haldine Group) Retrieved October 4, 2021, from Imperial Bioscience Review: https://imperialbiosciencereview.com/2020/11/20/how-do-we-tackle-vaccine-hesitancy-and-effectively-communicate-vaccine-safety-to-the-general-public-in-the-covid-19-era/

Jeong, M. (2021, September 6). *Global COVID-19 vaccine summary: Side effects* . Retrieved from Medical News Today: https://www.medicalnewstoday.com/articles/global-covid-19-vaccine-summary-side-effects

Lewis, S., Moore, C., & McBratney, S. (2021, August 31). *The Most Promising COVID-19 Vaccines* . Retrieved from Health Grades Web site: https://www.healthgrades.com/right-care/vaccines/the-most-promising-covid-19-vaccines

Christ, C. (2021, August 26). *Moderna Requests Full FDA Approval for COVID-19 Vaccine*. Retrieved from WebMD News Web site: https://www.webmd.com/vaccines/covid-19-vaccine/news/20210826/moderna-requests-full-fda-approval-covid-vaccine

Say, D., & Crawford, N. (2021, June). *COVID-19 Vaccine: Novavax*. Retrieved from Melbourne Vaccine Education Centre Web site: https://mvec.mcri.edu.au/references/covid-19-vaccine-candidate-novavax/

Yale School of Medicine. (2021, February 08). *Understanding COVID-19: Vaccine Distribution Challenges*. Retrieved from Yale School of Medicine: https://medicine.yale.edu/media-player/understanding-covid-19-vaccine-distribution-challenges/

Petkov, H. (2021, January 29). *The challenges of the global distribution of COVID-19 vaccines*. Retrieved from Maersk: https://www.maersk.com/news/articles/2021/01/29/the-challenges-of-the-covid-19-vaccine-global-distribution

Smith, T. M. (2021, March 04). *Dealing with COVID-19 vaccine hesitancy among health care workers* . Retrieved from ama-assn.org: https://www.ama-assn.org/delivering-care/public-health/dealing-covid-19-vaccine-hesitancy-among-health-care-workers

OECD. (2021, May 10). *Enhancing public trust in COVID-19 vaccination: The role of governments*. Retrieved from OECD: https://www.oecd.org/coronavirus/policy-responses/enhancing-public-trust-in-covid-19-vaccination-the-role-of-governments-eae0ec5a/

Lao, D. (2021, September 23). *COVID-19: Nasal antibody spray could be ‘future’ of vaccines, experts says* . Retrieved from Global News: https://globalnews.ca/news/8216134/nasal-covid-spray-vaccine-llama/

Elflein, J. (2021, October 1). *Covid-19 cases, recoveries, and deaths*. Retrieved from Statista: https://www.statista.com/statistics/1087466/covid19-cases-recoveries-deaths-worldwide/

Holder, J. (2021, October 4). *Tracking Coronavirus Vaccinations Around the World*. Retrieved October 9, 2021, from The New York Times: https://www.nytimes.com/interactive/2021/world/covid-vaccinations-tracker.html

Randall, T., Sam, C., Tartar, A., & Cannon, C. (2021, March 09). *Covid-19 Deals Tracker: 9.6 Billion Doses Under Contract*. Retrieved October 09, 2021, from Bloomberg: https://www.bloomberg.com/graphics/covid-vaccine-tracker-global-distribution/contracts-purchasing-agreements.html

Felter, C. (2021, June 30). *A Guide to Global COVID-19 Vaccine Efforts*. Retrieved October 09, 2021, from CFR: https://www.cfr.org/backgrounder/guide-global-covid-19-vaccine-efforts

Yousefinaghani, S., Dara, R., Mubareka, S., Papadopoulos, A., & Sharif, S. (2021, June). An Analysis of COVID-19 Vaccine Sentiments and Opinions on Twitter. *Journal of Medical Internet Research, 23*(6), https://doi.org/10.2196/24435.

MacMillan, C. (2021, October 6). *Will You Need a COVID-19 Booster? What We Know So Far*. Retrieved from Yale Medicine: https://www.yalemedicine.org/news/covid-19-booster

McGregor, G. (2021, September 24). *A WHO-approved Novavax vaccine could upstage mRNA jabs—if it can solve its manufacturing delays*. Retrieved October 09, 2021, from Fortune: https://fortune.com/2021/09/24/novavax-who-approval-covid-vaccine-mrna-covax/

Zimmer, C., Corum, J., & Wee, S.-L. (2021, October 07). *Coronavirus Vaccine Tracker*. Retrieved October 09, 2021, from The New York Times: https://www.nytimes.com/interactive/2020/science/coronavirus-vaccine-tracker.html

Megget, K. (2020, June 04). Even covid-19 can’t kill the anti-vaccination movement. *BMJ*(8250), https://doi.org/10.1136/bmj.m2184. Retrieved from https://www.bmj.com/content/369/bmj.m2184

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychological Science, 31*(7), doi:10.1177/0956797620939054.

Loomba, S., de Figueiredo, A., Patiek, S. J., de Graaf, K., & Larson, H. J. (2021, February 05). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nature Human Behaviour, 5*, https://doi.org/10.1038/s41562-021-01056-1.

Thunstrom, L., Ashworth, M., Finnoff, D., & Newbold, S. (2020, June 30). *Hesitancy Towards a COVID-19 Vaccine and Prospects for Herd Immunity*. Retrieved October 26, 2021, from SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract\_id=3593098

Bai, N. (2021, March 16). *How Effective Is the Johnson & Johnson COVID-19 Vaccine? Here’s What You Should Know*. Retrieved from University of California San Francisco Web site: https://www.ucsf.edu/news/2021/03/420071/how-effective-johnson-johnson-covid-19-vaccine-heres-what-you-should-know

Grady, D., & Zimmer, C. (2021, April 13). *J&J Vaccine and Blood Clots: The Risks, if Any, Are Very Low*. Retrieved from The New York Times Web site: https://www.nytimes.com/2021/04/13/health/blood-clots-johnson-vaccine.html

Eurepeon Medicines Agency. (2021, April 07). *AstraZeneca’s COVID-19 vaccine: EMA finds possible link to very rare cases of unusual blood clots with low blood platelets*. Retrieved from European Medicines Agency Web site: https://www.ema.europa.eu/en/news/astrazenecas-covid-19-vaccine-ema-finds-possible-link-very-rare-cases-unusual-blood-clots-low-blood

World Health Organization. (2020, December 31). *WHO issues its first emergency use validation for a COVID-19 vaccine and emphasizes need for equitable global access*. Retrieved November 02, 2021, from World Health Organization: https://www.who.int/news/item/31-12-2020-who-issues-its-first-emergency-use-validation-for-a-covid-19-vaccine-and-emphasizes-need-for-equitable-global-access

Twitter. (n.d.). *Brandwatch*. Retrieved from partners.twitter.com: https://partners.twitter.com/en/partners/brandwatch

Swaminathan, S., & Gupta-Smith, V. (2020, August 28). *Episode #1 - Herd immunity* . Retrieved from World Health Organization: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/media-resources/science-in-5/episode-1