Data Mining on COVID-19 Vaccines: Side Effects

Jing You  
Department of Biomedical Engineering, University of North Texas, USA  
JingYou@my.unt.edu

Nagma Shaik  
Department of Information Science, University of North Texas, USA  
nagmashaik@my.unt.edu

Haihua Chen  
Department of Information Science, University of North Texas, USA  
Haihua.Chen@unt.edu

ABSTRACT
COVID-19 is a pandemic disease affecting billions of people worldwide. Taking vaccines is a most effective approach to gain fully control. Thanks to the coordinated efforts from all over the world, several brands of vaccines targeting COVID-19 have passed through clinical trials and been brought to the public. Growing numbers of people are taking vaccines and share their feedback on social media, mostly on Twitter. In this study, we used Twitter data to analyze the side effects on each individual and quantify these side effects in a brand-wise and country-wise manner. Based on Twitter data, we found that the United States has the largest number of people getting vaccinated, Pfizer is the most widely used vaccine brand around the world and the most frequent side effect is cold. From our analysis, the side effects of vaccines are under controllable and are acceptable, and everyone can join the vaccinated camping without hesitation.

KEYWORDS
COVID-19 vaccines; side effects; Twitter; keyword extraction; N-gram.

INTRODUCTION
COVID-19, the most challenging pandemic in the 21st century, has infected over 174 million people and lead to the death of 3.7 million individuals around the world (World Health Organization, 2021). Thanks to the development of COVID-19 vaccines, we can gain control of the pandemic. However, there are restrictions and side effects followed by the vaccination. According to the statement of Centers of Disease Control and Prevention (CDC), people should not get the COVID-19 vaccine if they had a severe allergic reaction after a previous dose of this vaccine or had a severe allergic reaction to any ingredient of the vaccine (Centers for Disease Control and Prevention, 2021). A list of side effects has been reported: headache, muscle pain, redness and swelling injection site, cold, fever, tired and sleepy. There is also a remote chance to get severe allergic reaction, and will occur within a few minutes to one hour after getting injection (Centers for Disease Control and Prevention, 2021). Although more and more people are getting vaccinated, a big number of people are still hesitating mainly due to the side effects of vaccines. Our goal for drafting this paper is to dispel concerns of vaccines side effect by making proof with Twitter data.

Many people are sharing their feedbacks in social media after taking COVID-19 vaccines. Twitter data is the most accessible one among them. In this article, we collected tweets from Mar.15 to May 15 of 2021 and using brand names and side effects as unigrams and bigrams to extract the keywords of side effects from tweets and performed brand-wise and country-wise analysis to answer these research questions below.

Q1: What are the side effects for dose 1 and dose 2 according to Twitter data?
Q2: What are the side effects for different vaccine brand based on Twitter data?
Q3: What’s the percentage of people taking different brand of vaccines reported in Twitter?
Q4: What’s the country-wise vaccination progress based on Twitter data?

METHODOLOGY
Tweets collection & data cleaning. Tweepy library (Tweepy, 2020) is used to collect tweets together with 3 hashtags including #COVID19vaccine, #CoronaVaccination, #vaccinated, brand names are not used as hashtags for fear that it may bias the result. Data cleaning and preprocessing were performed on collected tweets. Retweeted tweets and irrelevant information were removed together with Emoji and special characters.

Data Modeling. After data is cleaned, keyword extraction is performed to extract tweets with both side effects and brand (Sarker, 2020). Extracted data is manually reviewed to delete false positive tweets. We then used filtered data to generate the unigrams and bigrams. Unigrams were generated for performing analysis on Vaccine brands. Since vaccine brands being single words. Bigrams were generated for vaccine reactions observation analysis. The dictionaries were created to represent the datasets generated out of n-gram models which later proceeded with data visualization using Matplotlib & Pyplot.
RESULT & DISCUSSION
In this study, we collected more than 440,000 tweets. After data cleaning and removed retweets and irrelevant info, 178,137 tweets were kept for further analysis. Based on user location, country-wise information of vaccinated people are calculated (Figure 1B). United States has the largest percentage of people vaccinated reported on Twitter. Based on the description in tweets, we classified the side effects into 12 categories, counts of each category is shown in figure 1A. Cold is the most reported side effects from Twitter data.

![Figure 1. Overall statistics of side effects (A) and country-wise analysis of vaccinated people (B)](image)

We cross-linked the side effects with vaccines brand name to find out which is the most frequent side effect in each brand (Figure 2A). Taking Pfizer as an example, the most frequently reported side effect is sore arm. We also quantified the vaccine usage of different brand around the world (Figure 2B). Pfizer has the most widely usage around the globe.

![Figure 2. Brand-wise side effects analysis (A) and vaccine brand usage around the globe (B)](image)

These results reported the vaccination trend around the world, and concluded the major side effects reported by people using Twitter. The side effects in each vaccine brand is organized and clear, although not distinguish between dose 1 and dose 2. There is some inevitable bias in data selection, as Twitter is not a mainstream social media in some countries, such as China, and we also deleted the non-English written tweets, so it may not represent the real trend in the world. But the study is meaningful, it reported that side effects of all the available vaccines are acceptable and these results can encourage more and more people to join the vaccination camp and finally save the world from coronavirus.

CONCLUSION
Based on the Twitter data, the side effects of current COVID-19 vaccines are acceptable and under control. There is no need to hesitate about side effects after taking vaccines and everyone should join the group of getting vaccinated to protect themselves and people they contacted with. Vaccines is a key step in fighting against the pandemic COVID-19 disease. No matter getting vaccinated or not, we should still remember to wear masks when attending gathering activities to stay safe and health. It’s a long run in the fight against the COVID-19 but we will succeed at last.

REFERENCES
Centers for Disease Control and Prevention. (2021, March 4) What to Do if You Have an Allergic Reaction After Getting A COVID-19 Vaccine. Retrieved May 24, 2021 from https://www.cdc.gov/coronavirus/2019-ncov/vaccines/safety/allergic-reaction.html
Centers for Disease Control and Prevention. (2021, May 5) Possible Side Effects After Getting a COVID-19 Vaccine. Retrieved May 24, 2021 from https://www.cdc.gov/coronavirus/2019-ncov/vaccines/expect/after.html

Sarker, A., Lakamana, S., Hogg-Bremer, W., Xie, A., Al-Garadi, M. A., & Yang, Y. C. (2020). Self-reported COVID-19 symptoms on Twitter: an analysis and a research resource. *Journal of the American Medical Informatics Association, 27*(8), 1310-1315.

Tweepy. (2020, December 25) Tweepy: Twitter for Python! Retrieved March 10, 2021 from https://www.tweepy.org/

World Health Organization. (2021, June 9) *WHO Coronavirus (COVID-19) Dashboard*. Retrieved June 9, 2021 from https://covid19.who.int/.