How did public sentiment towards COVID-19 vaccines on Twitter change throughout the key phases of the pandemic?

By: Ehren Dietrick, Huong Giang, Aminata Bangura

Agenda

Introduction to the dataset

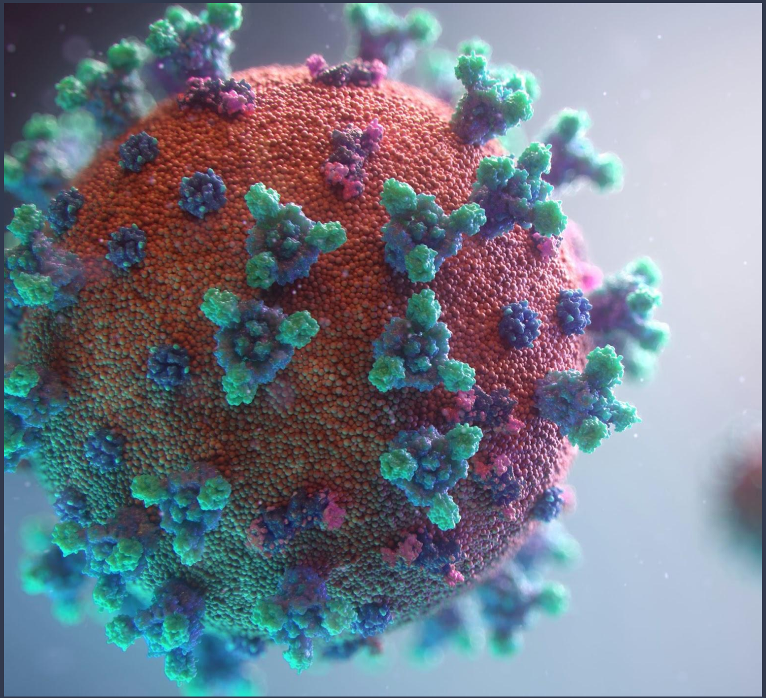
Data collection methods

LDA Results

Sentiment Analysis (bing + nrc)

Conclusion

Introduction



Social media gives a unique view into multiple perspectives

Pandemics create more than just physical and economic change - it shapes emotions

Social media captures real-time reactions from millions of people, so it’s a good resource to use

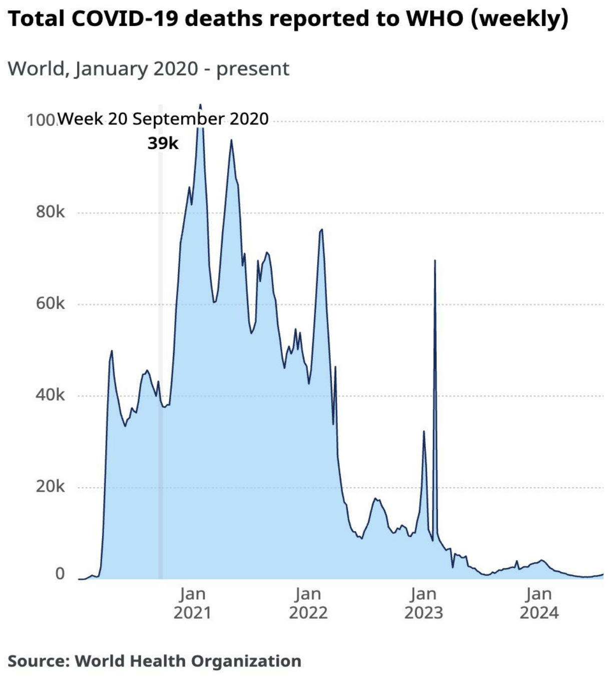
With a focus on death toll trends, we can look at reactions to trends found and look more into what each trend means

We aim to discover how uncertainty, information availability, and perceived progress shape public response over time

Data collection + preprocessing

- Source & Scope: Collected via Twitter's API using #covidvaccine, the dataset comprises 209,930 tweets from February to October 2020, featuring core metrics, content, and engagement variables. This timeframe capturing the initial global vaccine rollout, public debate, and the emergence of major variants → ideal for analyzing sentiment fluctuations
- Period Segmentation: The data was divided into six distinct phases aligned with global weekly death toll trends, using mortality as a concrete driver of public fear and discourse.
- Analytical Purpose: This segmentation created manageable, context-ric periods, from initial stability to sharp peaks and declines, enabling deeper analysis of how escalating and receding mortality influenced vaccine conversation dynamics.

- Using the tm package, the tweet text was converted into a corpus, systematically cleaned and normalized, and then transformed into a Term-Document Matrix for all subsequent analysis.



LDA topic

During the initial low-volume phase (Feb-Mar), discourse centered on personal emotions, vaccine logistics, and political commentary. A sharp surge in conversation (Mar-Jun) saw a dramatic expansion into global conspiracy theories, intensified personal skepticism, and sustained critiques of media and leadership.

Feb 22 – Mar 14 (Fluctuated under 1,000)
Topic 1: Personal Acknowledgement & Vaccination Debate (informal, emotional language).
Topic 2: Vaccine Development & Accessibility (logistics, science, equity).
Topic 3: Political & Social Commentary (government response, mandates, economic impact).

Mar 15 – Apr 25 (Sharp increase: 2.7k to 50k)
Topic 1: Globalized Conspiracy Theories (Bill Gates narratives, 5G/microchip myths).
Topic 2: Personal Skepticism & Testing Debates.
Topic 3: Vaccine Development & Equity Concerns (“vaccine nationalism”).
Topic 4: Political Response & Leadership Critique.
Topic 5: Media Critique & Minimization of severity.

Apr 26 – Jun 6 (Gradual decline: 50k to 33k)
Topic 1: Personal Choice & Skeptical Hesitancy.
Topic 2: Geopolitics of Vaccine Development (global race, key players).
Topic 3: Tracking Progress (impatience with trial stages).
Topic 4: Evaluating Vaccine Safety & Efficacy.
Topic 5: Scientific Conspiracies (genetic modification, global control theories).

Jun 7 – Jul 11 (Fluctuated around 35k)
Topic 1: Anticipation & Cautious Optimism (trial timelines).
Topic 2: Personal Participation in Clinical Trials.
Topic 3: Skepticism of Rushed Safety Protocols.
Topic 4: Geopolitical Demand & Nationalism (focus on India’s access).

J***ul 12 – Sep 19 (Fluctuated around 40k)***
Topic 1: Global Vaccine Politics & Distrust (“big pharma”).
Topic 2: Personal Hesitancy & Media Distrust.
Topic 3: The Russian Vaccine & Crisis of Trust (Sputnik V announcement).
Topic 4: Russia-West Geopolitical Drama.
Topic 5: U.S. Political Pressure & Regulatory Scrutiny.

Sep 20 – Oct 22 (Sharp decline)
Topic 1: Corporate Rollout & Public Messaging (company/agency communications).
Topic 2: Political Pressure on Regulatory Approval (Trump, FDA, election timing).
Topic 3: Election-Linked Skepticism (fears of rushed approval).
Topic 4: Logistical Realities & Side Effect Concerns.
Topic 5: Crisis of Trust in Institutions (media, leadership, science).

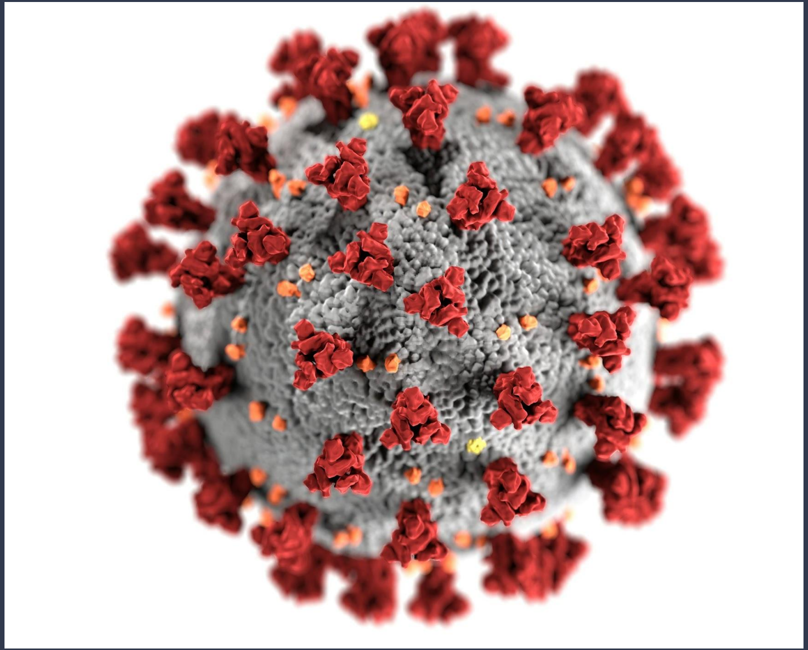
From summer through the fall (Jun-Oct), the dialogue shifted to a protracted phase dominated by geopolitical competition, institutional distrust, and mounting public anxiety over safety, logistics, and the politicization of the rollout.

Bing Sentiment Analysis

Sentiment analysis per topics

Dataset	Period	Negative Words	Positive Words	Net Sentiment	Total Words	Pos/Neg Ratio	Sentiment Character
Covid 6	Feb 22 - Mar 14 (<1k)	316	371	+55	687	1.067	Slightly Positive
Covid 5	Mar 15 - Apr 25 (2.7k → 50k)	5880	6372	+492	12252	1.0837	Slightly Positive
Covid 4	Apr 26 - Jun 6 (50k → 33k)	12536	12934	+398	25470	1.0317	Slightly Positive
Covid 3	Jun 7 - Jul 11 (~35k)	12223	12199	-24	24422	0.998	Most Neutral
Covid 2	Jul 12 - Sep 19 (~40k)	71107	74008	+2901	145115	1.0408	Slightly Positive
Covid 1	Sep 20 - Oct 22 (Sharp Decline)	31610	29973	-1637	61583	0.9482	Slightly Negative

Sentiment Analysis conclusions (Bing)



While ultimately relatively neutral with only slightly positive or negative tone for the entirety of each period, the shift between each are important to note.

Language during March-April periods are important: “Want” “Hope” “Need” - not showing positivity, but more intense emotional engagement

Moving forward, LDA trends seemed to stay consistent

Largest dataset (most tweets) July-September, with a high death toll, and a positive ratio, aligning with solution-focused language

Sentiment fell soon after, right as language started to trend more towards social conflicts, despite lower death rates and better conditions

Throughout the pandemic, the public's sentiment evolved from initial observational concern and logistical hope to confrontational backlash and fatalism, ultimately solidifying into a central conflict over compliance, safety, and trust, which, by the end, revealed a scientific solution overshadowed by the sobering reality of loss.

Bing Sentiment Analysis

Positive Wordcloud (feb 12 - march 14)

Positive Wordcloud (march 15 - april 25)

Positive Wordcloud (april 26 - june 6)

Positive Wordcloud (june 7 - july 11)

Positive Wordcloud (july 12 - sep 19)

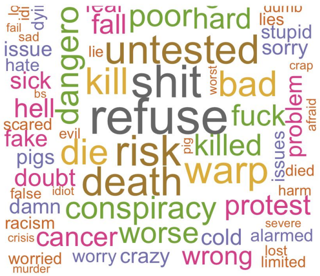
Positive Wordcloud (sep 20 - oct 22)

Negative Wordcloud (feb 12 - mar

Negative Wordcloud (march 15 - april 25)

Negative Wordcloud (april 26 - june 6)

Negative Wordcloud (june 7 - july 11)



Negative Wordcloud (july 12 - sep 19

Negative Wordcloud (sep 20 - oct 22)

NRC Sentiment Analysis

The NRC lexicon was used to identify emotional patterns in Covid-19 discussions across six time periods. Ten emotions were detected, fear, trust, surprise, sadness, anger, surprise, joy and disgust–along with positive and negative sentiment.

Covid 6: February 22-March 14

•Emotions such as fear and anticipation were the most prominent as the virus emerged. Limited information and uncertainty about the virus influenced public reactions.

Covid 5: March 15-April 25

•Huge rise in fear, sadness and negative sentiment during initial lockdowns

Covid 4: April 26-June 6

•Continued emotional intensity with growing frustration and anger

Covid 3: June 7-July 11

•Joy, trust, and positive emotion increase due to reopenings and vaccine trial announcements

Covid 2: July 12-September 19

•Highest emotional volume, reflecting debates over restrictions, public concerns, and misinformation

Covid 1: September 20-October 22

•Rise in trust and anticipation influenced by vaccine announcements, but also political anger

Datasets

```
> covid6_nrc
       sentiment   n   dataset
1       positive 941 Dataset 6
2       negative 425 Dataset 6
3           fear 306 Dataset 6
4          trust 296 Dataset 6
5   anticipation 256 Dataset 6
6        sadness 168 Dataset 6
7          anger 160 Dataset 6
8       surprise 154 Dataset 6
9            joy 140 Dataset 6
10       disgust 107 Dataset 6
```

```
> covid5_nrc
      sentiment     n   dataset
1      positive 19637 Dataset 5
2      negative  7703 Dataset 5
3         trust  6593 Dataset 5
4  anticipation  5098 Dataset 5
5          fear  4956 Dataset 5
6           joy  3142 Dataset 5
7       sadness  2959 Dataset 5
8         anger  2661 Dataset 5
9      surprise  2479 Dataset 5
10      disgust  1974 Dataset 5
```

Datasets

```
> covid4_nrc
      sentiment     n   dataset
1      positive 41311 Dataset 4
2      negative 15391 Dataset 4
3         trust 13528 Dataset 4
4  anticipation 10833 Dataset 4
5          fear 10546 Dataset 4
6           joy  6521 Dataset 4
7       sadness  6514 Dataset 4
8         anger  5639 Dataset 4
9      surprise  5221 Dataset 4
10      disgust  3672 Dataset 4
```

```
> covid3_nrc
      sentiment     n   dataset
1      positive 41013 Dataset 3
2      negative 14355 Dataset 3
3         trust 13098 Dataset 3
4  anticipation 11072 Dataset 3
5          fear  9390 Dataset 3
6           joy  6230 Dataset 3
7       sadness  5815 Dataset 3
8         anger  5383 Dataset 3
9      surprise  4785 Dataset 3
10      disgust  3543 Dataset 3
```

Datasets

```
> covid2_nrc
      sentiment      n   dataset
1      positive 239537 Dataset 2
2      negative  83747 Dataset 2
3         trust  79148 Dataset 2
4  anticipation  59622 Dataset 2
5          fear  57283 Dataset 2
6           joy  35970 Dataset 2
7       sadness  35079 Dataset 2
8      surprise  33812 Dataset 2
9         anger  31267 Dataset 2
10      disgust  20834 Dataset 2
```

```
> covid1_nrc
      sentiment     n   dataset
1      positive 99886 Dataset 1
2      negative 38135 Dataset 1
3         trust 33347 Dataset 1
4  anticipation 26943 Dataset 1
5          fear 26329 Dataset 1
6       sadness 17246 Dataset 1
7         anger 14754 Dataset 1
8      surprise 14629 Dataset 1
9           joy 14138 Dataset 1
10      disgust  8785 Dataset 1
```

Summary of key emotions across each dataset

Time period	Dominant Emotions
February	Fear, Anticipation
March-April	Fear, Sadness
April-June	Anger, Fear
June-July	Joy, Trust, Anticipation
July-September	High emotional volume
September-October	Trust, Anticipation, Anger

Conclusion

When deaths were low, conversations were causal and focused more on reactions.

More deaths generally equalled more mistrust and misinformation, except when people felt real progress was happening

Patterns suggest that public reactions are influenced less by the objective severity of a crisis and more by uncertainty, trust, and perceived momentum toward solutions

Future research can help focus on demographics and political affiliation (More local trends) or with other disasters (hurricanes, wildfires, etc.)

References

John Hopkins University. (2023, March 10). Mortality Analyses. Johns Hopkins Coronavirus Resource Center. https://coronavirus.jhu.edu/data/mortality

Keller, A., Litzelman, K., Wisk, L. E., Maddox, T., Cheng, E. R., Creswell, P. D., & Witt, W. P. (2012). Does the perception that stress affects health matter? The association with health and mortality. *Health psychology : official journal of the Division of Health Psychology, American Psychological Association*, *31*(5), 677–684. https://doi.org/10.1037/a0026743

Levin, A. (2008). Death, Injury Tallies Mask Toll on Families' Mental Health. *Psychiatric News*, *43*(21), 17–17. https://doi.org/10.1176/pn.43.21.0017a

Mohammad, S. (2011). *Sentiment and Emotion Lexicons*. Saifmohammad.com. http://saifmohammad.com/WebPages/lexicons.html

World Health Organization. (2024). *COVID-19 deaths | WHO COVID-19 dashboard*. World Health Organization Data. https://data.who.int/dashboards/covid19/deaths?n=o

Yadav, R. (2020). *COVID-19 🦠 Vaccine Tweets*. Kaggle.com. https://www.kaggle.com/datasets/ritesh2000/covid19-vaccine-tweets/data

Thank You!