Analysis of Google Search Patterns during the COVID-19 Pandemic

(Milwaukee County, WI, USA)

Introduction

Why is this analysis interesting or important? (**color-coded)**

Does it solve a real problem or tackle an unresolved research question? (**color-coded**)

During any pandemic, individuals seek health information due to media curiosity, their own illness, or the illness of a family member. Health information seeking, also known as infodemiology, has been proposed as another disease surveillance method. During the epidemic, digital media played a massive role in spreading information. In addition to conventional surveillance, digital surveillance (internet) can provide intriguing trends concerning COVID-19's general concerns and aid in improving prediction models. The present digital surveillance analysis is driven by individuals' worries in the face of a pandemic. It is being used to determine the extent to which the Public Mask Mandate can safeguard people from getting COVID-19 symptoms. Google Trends (GT) is a valuable resource that may be used for digital monitoring. It may predict healthcare decisions in real time by following people's search behaviors and recording their worries (symptoms, vaccination availability, eligibility, and side effects).

Milwaukee is one of Wisconsin's largest counties with the highest number of detected COVID-19 cases. Illness-specific symptom search patterns on GT can alert the county's healthcare system to prepare and allocate resources needed ahead of time. The primary goal of this study is to leverage the GT to understand the trends in Milwaukee County citizens' search patterns for symptoms and correlate this information with the number of cases, fatalities, vaccinations, and impact of the public mask mandate in Milwaukee County. The findings can substantially aid in estimating the need for supplementary monitoring and policy methods and informing real-time public health choices.

Also, Vaccine hesitancy remains a severe challenge in ending the COVID-19 pandemic. Unfortunately, online platforms can also spread substantial misinformation about vaccines. The secondary goal of the analysis is to leverage the GT to understand the trends in vaccine misinformation and how public acceptance of vaccines changed over time with changes in mask mandate policies. Using this analysis, we can attribute these trends to public attitudes, interests during the varying vaccine availability, misinformation regarding the vaccines leading to a high unvaccinated population, and further target education among the citizens.

The proposed analysis can make the policy changes more human-centered as it indirectly incorporates the participatory design strategy. The epidemiological information about the virus spread in the county alone isn't sufficient to devise an effective plan to attenuate the infection rate and accelerate vaccine administration. Embedding the end users' thoughts and concerns about the pandemic and vaccinations can provide real-time, immediate, quick, and cheap feedback to the algorithms that predict/decide policy changes. For example, local governments can develop a practical vaccine distribution algorithm and make accurate infection predictions if they know the concerns of the citizens of the county. It eliminates the scope for wrong interpretations and provides more transparency to the end users of the algorithms developed specifically for the county. Uncovering these wide ranges of insights while considering the context of virus spread reveals insights about individual behaviors and the relationships between different entities involved in the policy-making process.

Background/Related Work

Hypothesis (**color-coded**)

There has previously been credible research published on mask effectiveness. A paper published in The Lancet by Dr. Chu et al. indicates that surgical and equivalent cotton masks are 67% effective in reducing COVID-19 transmission [1]. According to survey data from the New York Times [2], mask usage in Orange County is relatively high, with 60% of inhabitants wearing masks regularly or consistently. Initially, COVID-19 cases spread swiftly throughout the United States, particularly in big areas with huge populations, such as New York and Los Angeles. Though masks were heavily political in the United States, there is evidence that high mask adoption might save cases and associated fatalities, with one research in Nature suggesting that over 130,000 extra lives could be avoided.

For regression analysis, Jimenez AJ [3] used linear regression to analyze the association between the Google search trends and confirmed COVID-19 cases in Spain. The study assumed the data to be normal and independent as the sample size is small, considering only wave 1. To explore other flexible models, Abbas M [4] leveraged smoothed functional curves to explore the modes of variation in the data using functional principal component analysis (FPCA). The study also used functional clustering analysis and dynamic correlation to identify patterns of COVID-19-confirmed case and death trajectories across the US and their association with top Google search trends.

A common theme among all these regression studies about the association between symptoms is that they didn't analyze and account for the intervention effect – the effect caused by the change in mask policy. Also, they assume that the data follow normality and independence to model linear regression. But the data used for the current analysis spans 18 months, making it unique and violating all assumptions. Nguyen [5] conducted the survey analysis to examine how the Public Mask Mandate can protect individuals from developing the symptoms using a difference-in-differences (DID) framework. Drawing inspiration from this study, I used a similar methodology for my analysis with an additional dimension of specific symptom search trends. I plan to analyze how the mask mandate in the county impacted the search for all COVID-19 symptoms and how strongly the top symptom searches are impacted due to the mask mandate. I hypothesize that individuals are less likely to search for top and common COVID-19 symptoms like fever, cough, and pneumonia when the public mask mandate is effective in Milwaukee County.

For correlation analysis, Lippi et al. investigated the capacity of Google search volume of symptoms such as fever, cough, and dyspnea to predict the trajectory of the early 2020 COVID-19 outbreak in Italy using Spearman's correlation method. They concluded that GT's continuous monitoring is a valuable instrument in the early detection of COVID-19 outbreaks [6]. Jimenez A [7] also analyzed the association between the cases and symptom search trends using linear correlation. Most studies used conventional correlation methods to determine the relationship between symptom search and cases [8], [9]. Another approach is to use wave analysis to detect the synchrony between symptoms and cases [10], but it had the limitation of not seeing a correlation over time.

One common denominator in all these correlation studies that informed my analysis was the use of non-dynamic statistical procedures. Hence, I leveraged dynamic statistical procedures for correlation analysis to answer "What are the highly correlated symptom search terms with the daily confirmed COVID-19 cases and fatalities before and after the mask mandate policy?" I hypothesize that common COVID-19 health symptom searches like cough and shortness of breath will have a high time-lag correlation of 5 days (the time between contracting the virus and testing positive) with the daily number of confirmed cases, and the mask mandate will strongly impact these lag days.

Studies regarding vaccination intent used deterministic models for causal inference. The study on the effectiveness of face masks during vaccine rollout [11] aims to estimate the impact of community face mask use at varying levels of mask uptake and mask effectiveness during the rollout of vaccination in New York City. The process involved using an age-structured compartmentalized deterministic model. This study informed me about the causal relationship between mask usage and vaccine dynamics and helped me expand the similar thought for vaccination search terms in Milwaukee County. I plan to answer "How the vaccination intent and side effect search terms correlated with mask mandate policies?" by hypothesizing that County citizens showed more vaccination intention when the masking mandate was removed, and there is less negative news about the vaccine administration and availability.

Methodology

Why I chose the methods (**color-coded**), human-centered considerations (**color-coded**)

In the following sections, all the proposed analysis uses a general response metric - the infection rate. The infection rate is a good proxy for the daily confirmed cases and accurately represents the virus's progression and change over time. Below is the derived calculation for the infection rate where k = 100.

$$\frac{\text{\# of Infections}}{\text{Population at Risk}} \textbf{ X constant (k) = Rate of Infection}$$

The population at Risk is the susceptible population, i.e., the population susceptible to infection. Considering a **recovery period of 14 days**, I calculated the recovered population and subtracted it from the cumulative cases to get the infected population. These values are then used to obtain the susceptible population, i.e., the population at Risk.

Population at Risk = Total Population - Infected - Recovered - Deaths

The infection rate time series is used in the derivative and regression analysis mentioned below.

Derivative Analysis - Change point detection [12]

A time series signal will have a few components like trend, seasonality, and cyclicity embedded in it. The signal can forego changes for many direct and indirect reasons. It is crucial to detect whether or not a change has occurred, or whether several changes might have occurred and identify the times of such changes. Change point detection (CPD) [12] is one such technique used in time series analysis to detect the changes in the signal. CPD detects abrupt changes in time series when the property of the signal changes. Using CPD, we can identify the pivot points in virus progression and draw inferences for the detected changes. I leveraged ruptures [13] that capture the change in the infection rate derivative. It leverages the Pelt Search method to identify and plot the change points (brown dotted lines) in the infection rate data, indicating slope changes over the infection rate. Figure 1 shows the sample change points detected using ruptures on a time series signal.

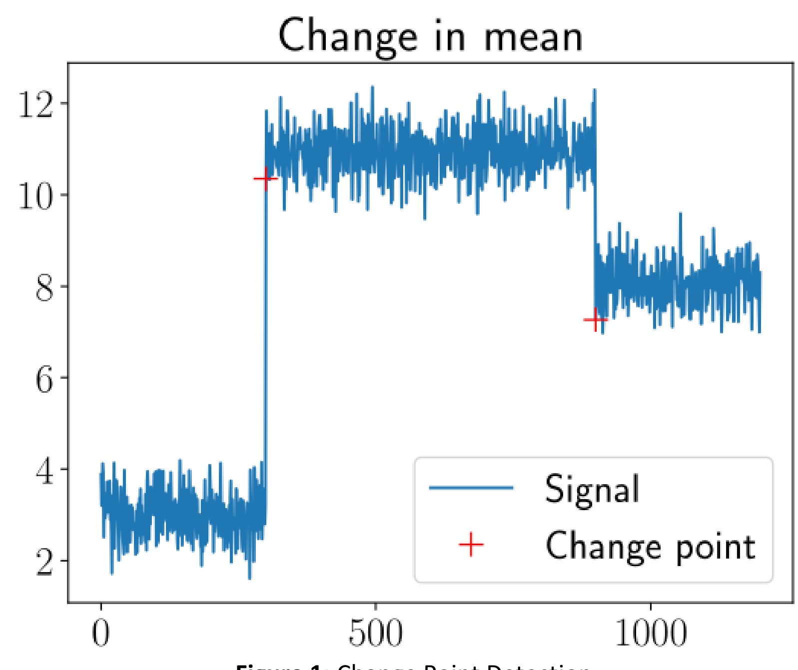

Figure 1: Change Point Detection

The red change points are the detected changes in the signal sequence. I leveraged the algorithm to detect the change points in infection rate, symptom search terms, and vaccination search terms and drew inferences about these change points with respect to mask mandate and vaccine-related news in the county.

Regression Inference – Difference in Difference [14]

To understand the impact of the mask mandate, we need to design an experimental technique and observe how the experiment's environment changes with the intervention. Difference-in-Difference (DID) [14] is an experimental design technique using longitudinal time series data to estimate a causal effect after an intervention. DID accurately estimates the effect of mask mandate on Symptom searches (a proxy for symptoms) by comparing similar counties. DID uses regression theory to model the differential timing of the mask mandate implementation in Milwaukee County. In this framework, we compare the health outcomes for individuals under the Public Mask Mandate period with those in the same county and similar counties reported when the mandate had not been enforced.

The method involves fitting a regression line between the control group (another county – Bergen, NJ) and the treatment group (Milwaukee County) and obtaining the coefficients. The covariates involve the relative search values (RSV) of the top COVID-19 symptom searched during the time frame. Figure 2 is the visual representation of the observable effect using the algorithm.

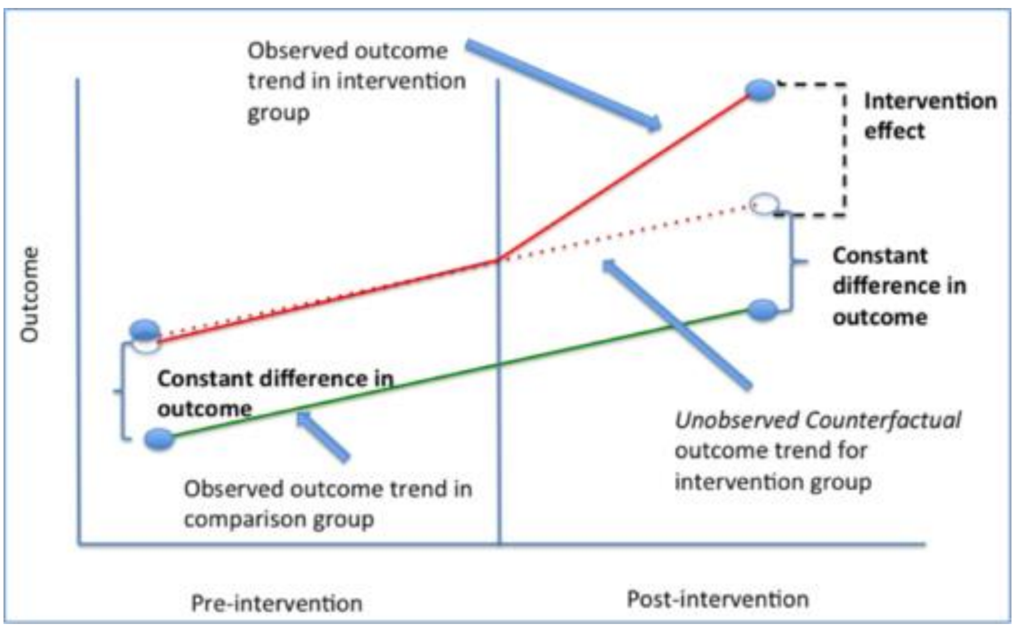

Figure 2: Difference-In-Difference

The coefficients of these covariates β1 summarize the level of impact the Public Mask Mandate had on individuals' search for COVID-19 symptoms. The symptom searches with the highest coefficient values quantify the highest impacted symptoms due to the masking policy.

Correlation Inference –Rolling Windowed Time Lagged Correlation [15]

To detect the temporal relationship between the symptom-related search terms and infection rate, we need to assess the association's direction and strength. For this purpose, we need to calculate the time-lagged correlation values (Pearson correlation coefficients) and *p*-values for each symptom-related search term. Windowed TLCC [15] is a synchrony measurement technique that obtains the correlation values with a lag between signals. WTLCC helps identify the directionality, synchrony, and leader-follower relationship between symptom searches and daily COVID-19 confirmed cases. Figure 3 shows the sample sinusoidal signal and the respective time-lagged cross-correlation for the time series.

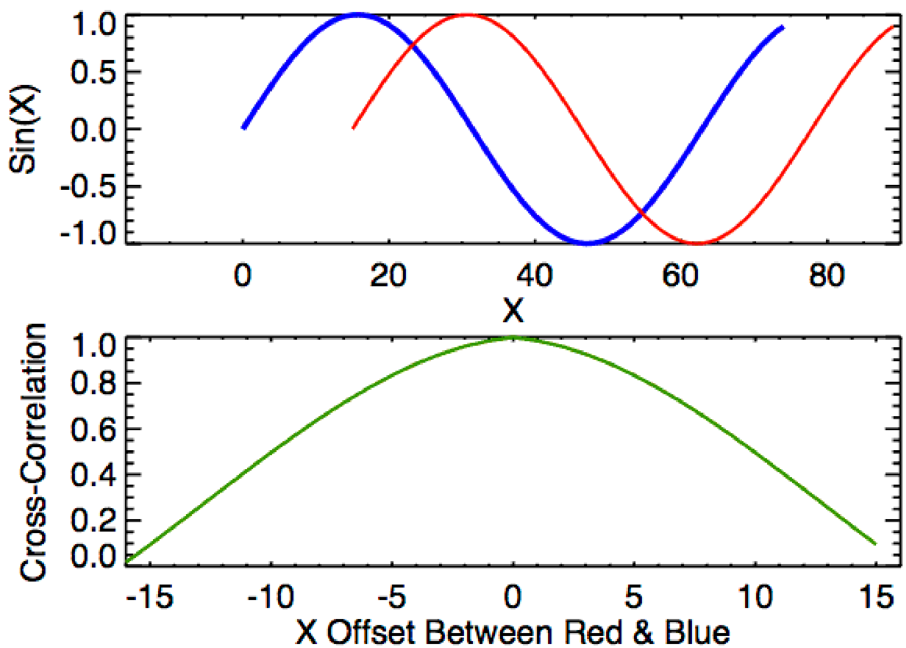

Figure 3: Rolling Windowed Time Lagged Cross-Correlation

It uses sliding window offset and calculates Pearson correlation values for each lag value between the cases and COVID-19 top searched symptoms. For each symptom search, the high correlation lag value over time with statistical significance will help us quantify the impact of mask mandate on search terms.

Causal Inference – Granger Test [16]

National, state, and county news about vaccines will have a bi-directional impact on the citizens' vaccination interest and side-effects concerns. We must leverage a technique that understands this causal relationship of relevance in the news with varying vaccination search patterns in Milwaukee County. For this purpose, we need to try techniques that don't reflect "mere" correlations, like regression. The Granger causality test [16] is a statistical hypothesis test for determining whether one-time series is helpful in forecasting another. It effectively measures the significance of the association between vaccination news and the vaccination search time series. Below is the sample graph, which describes the causal association between the two-time series that the granger causality test can detect.

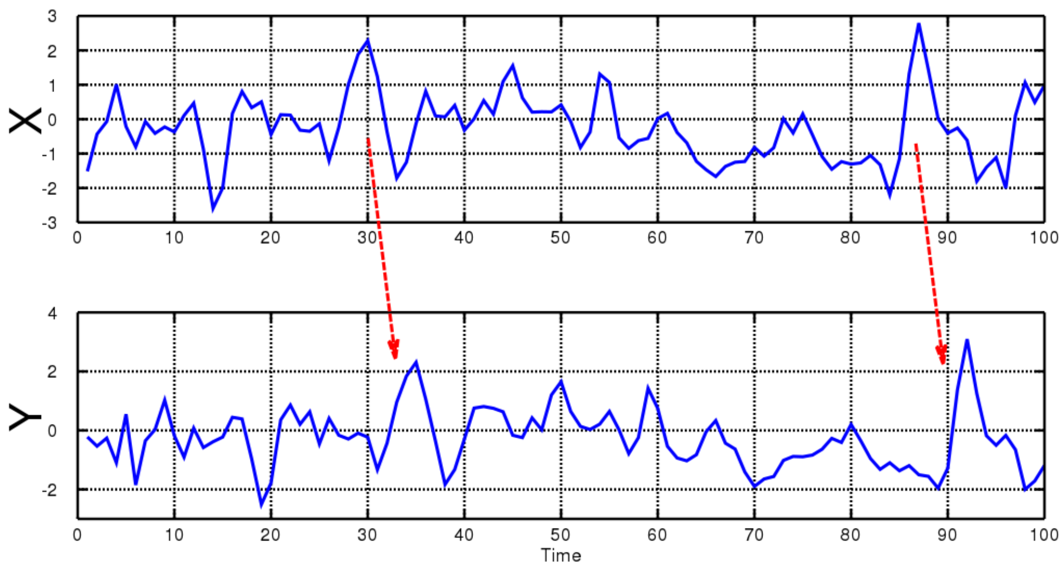

Figure 4: Granger Test

All the above-proposed analysis methods are designed with human-centeredness in mind. Firstly, the proposed design for the study is people-centered. It embeds the end users' thoughts and concerns about the symptoms and vaccines, leveraging them to develop policy health care changes for citizens. It leverages the data produced by the people for the people. Secondly, I adopted a participatory design where the data from users' searches is incorporated into the developed solution in real time. This immediate, quick, and cheap feedback will help proactively update the policy suggestions based on the actions performed by the people. Especially for this task, given the anonymity in the virus dynamics and limited knowledge about the nuances in the study, it is critical to give and take inputs in a collaborative cohort setting.

Thirdly, the design process involves ethical considerations. It uses masked data produced by the people of Milwaukee County and is free of other demographic biases. For the data leveraged by the above three methods proposed, differential privacy has been used by adding artificial noise, enabling high-quality results without identifying anyone. To further protect people's privacy, it is ensured that the study leverages no personal information or individual search queries. Finally, the proposed solution of the study is highly reproducible. The methodology proposed can be leveraged to fit any data consisting of people's searches by validating and incorporating the assumptions made in regression and correlation analysis.

Findings

Change point detection – Infection Rate

The change point detection analysis is performed on the infection rate and the derivative of the infection rate. The infection rate change point visualization Figure 5 shows the change points that capture the difference in the signal. Figure 6 shows the change points in the derivative function in daily infection rate along with the timelines of masking mandate policies in Milwaukee County, Wisconsin State, USA. The vertical lines indicate the detected change points.

The ideal way of reading the plot is first to understand the virus progression using change points at various periods and how it has brought the necessity of masking policy. Next, by looking at the post-facto lens for impact analysis, the reader must observe the masking policy dates and infer how the policy changes impact the infection rate - Change points in the data after a few time intervals post-mandate.

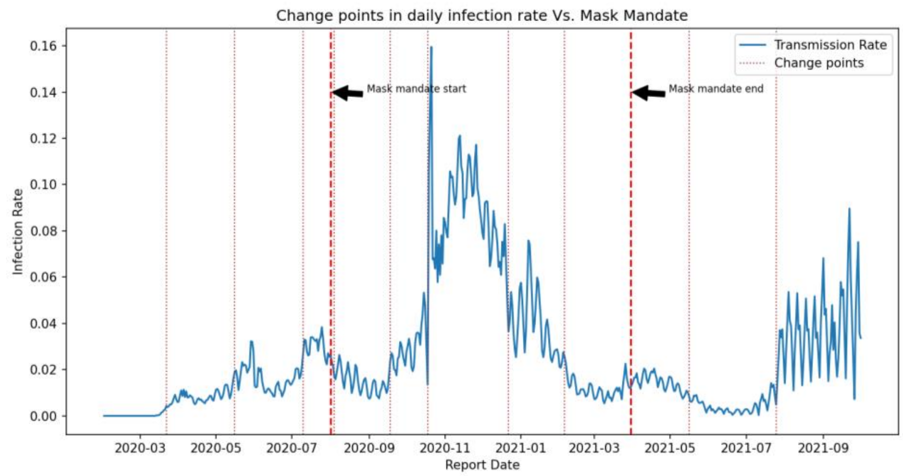

Figure 5: Change Points in Infection rate

From the 4th change point, we can observe how the masking mandate lowered the infection rate. This change shows evidence of the impact of the masking policy. From the 8th change point to the 9th, there is no considerable difference in infection rates. Hence, removing the masking policy can be attributed to stability. However, at the final change point (10th), the COVID-19 infection rate started to peak.

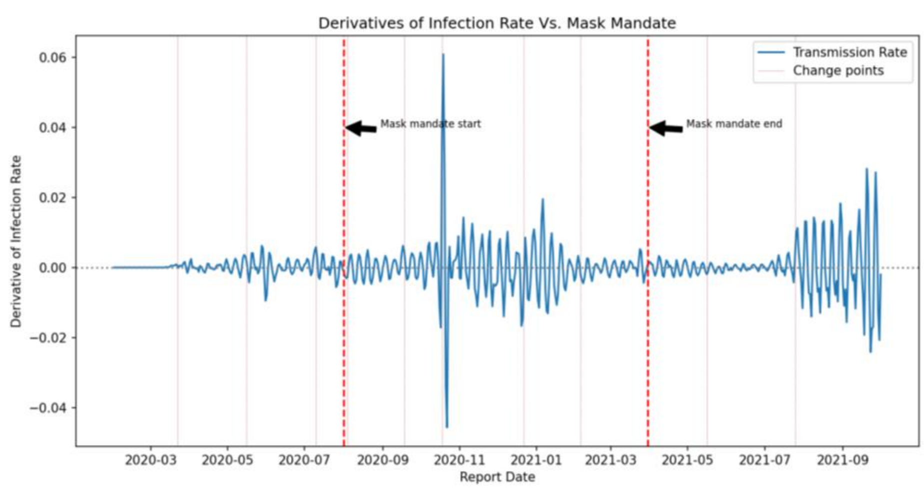

Figure 6: Change Points in derivative of Infection rate

We can observe that at the 4th change point after the masking mandate started, the derivate is closer to 0, indicating the change in infection rate over time is not frequent. The same pattern can be observed after removing the mandate but oscillates more at the end.

Key Takeaways:

Overall, it is evident from the above analysis the reasons for enforcing and lifting the masking policy took place in Milwaukee County. It indicates the impact of masking on the infection rate by showing how it decelerated after the mandate was enforced and accelerated after the mandate was removed. However, a few exceptions are observed where we see peak infection rates even after the masking policy is in place. This can be attributed to the high impact of other aspects like vaccinations, recovery rates, and hospitalizations.

Change point detection – Google Symptom Search Trends

Similar to the above analysis, the change point detection is implemented on top of COVID-19-related symptom searches. First, using the relative search values (RSV) and the correlation with infection rate, we observe that FEVER, COUGH, and PNEUMONIA are the top symptom search terms that are associated with the infection rate. Figures 7, 8, and 9 represent the change points in these top symptom searches.

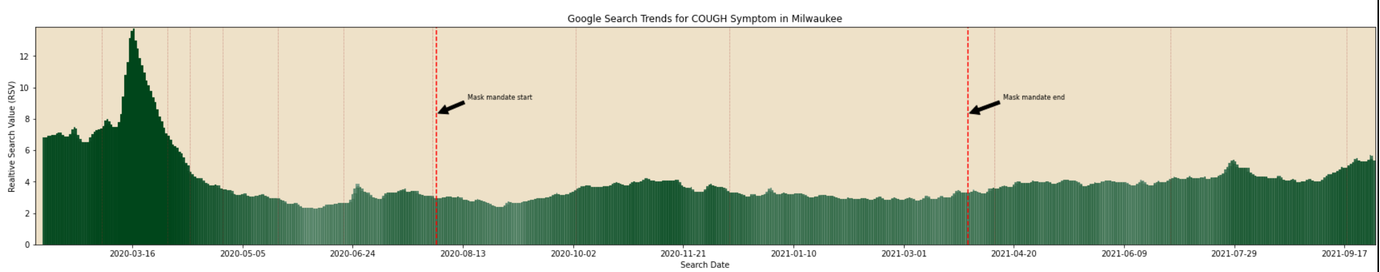

Figure 7: Change Points in COUGH symptom search

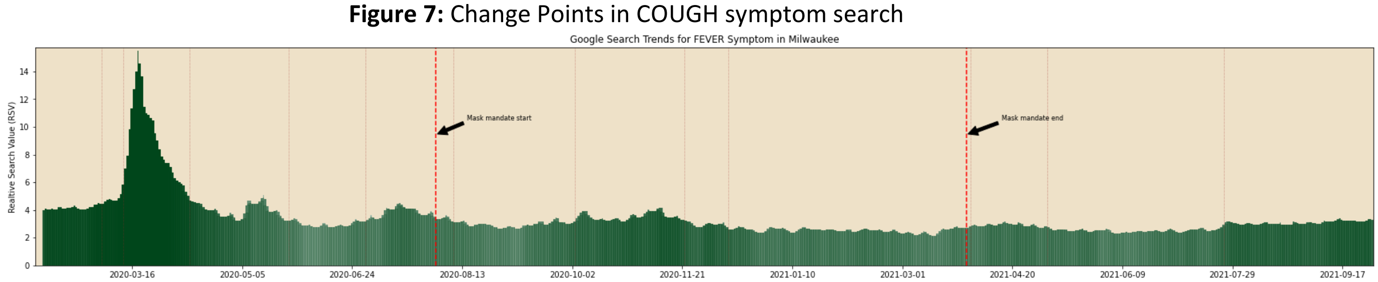

Figure 8: Change Points in FEVER symptom search

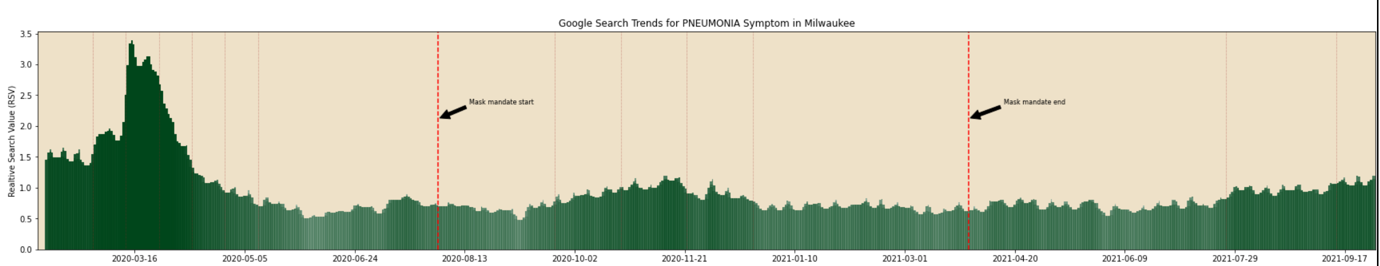

Figure 9: Change Points in PNEUMONIA symptom search

Key Takeaways:

FEVER symptom search term is strongly associated with daily COVID-19 confirmed cases. The change points detected for the FEVER search term closely align with those of the infection rate, concluding that the mask mandates evidently changed the symptom search trends. COUGH is the highest searched symptom with high RSV values but is moderately correlated with confirmed COVID-19 cases. This can be because it is a common symptom; the searches can be associated with other seasonal illnesses.

The PNEUMONIA symptom search term is strongly associated with daily COVID-19 fatalities proving that it is a highly critical health issue that requires immediate hospitalization. For PNEUMONIA, a new set of change points are identified that are inconsistent with those of infection rate and mask mandate changes.

Regression Inference – Impact of Mask Mandate on Symptom Searches

Difference-in-Difference model output coefficients are leveraged to visualize the predicted percentage change in symptom searches in Milwaukee County, where there is a change in mask mandate. For Bergen County, the percentage change values are calculated by subtracting the mean RSV values before and during the public mask mandate period in Milwaukee County.

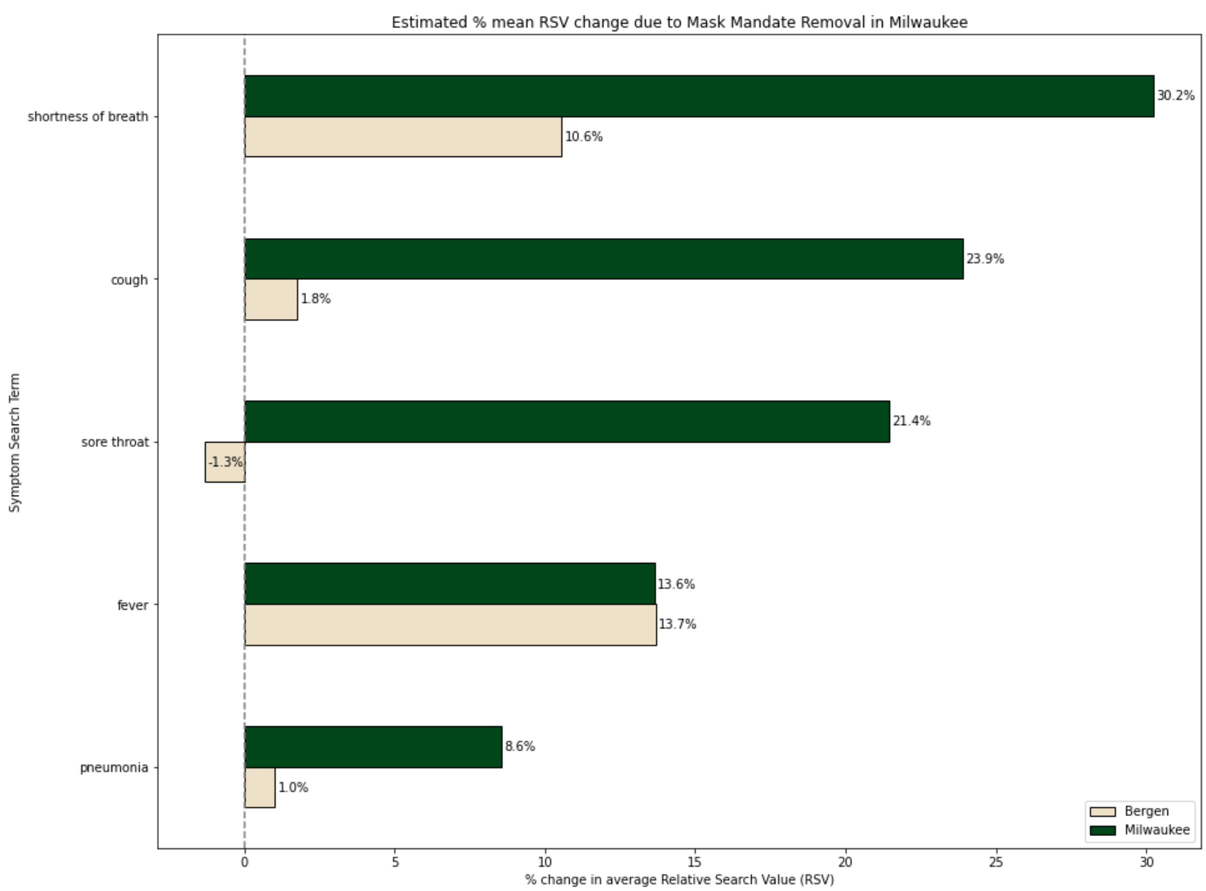

Figure 10: Change Points in PNEUMONIA symptom search

Key Takeaways:

SHORTNESS OF BREATH, COUGH and SORE THROAT are the top 3 symptom searches impacted by the mask mandate. SHORTNESS OF BREATH symptom search is highly impacted by the removal of the mask mandate in Milwaukee County (20% greater than another county with a continuous mask mandate). FEVER symptom search growth rate is approximately the same with and without mask mandate. PNEUMONIA symptom search term is the least impacted symptom search with the mask mandate in Milwaukee County.

Correlation Inference – Change in the association of Symptom Searches over time

Extending the above analysis, I picked the top and bottom-most public mask mandate-impacted symptom searches and analyzed the correlation values with infection rate. Figure 11 represents these correlation values in a heat map, with X-axis representing the lag values and Y axis representing the sliding window epoch used for calculation.

Time lagged Correlation plots before, during and after mask mandate in Milwaukee County

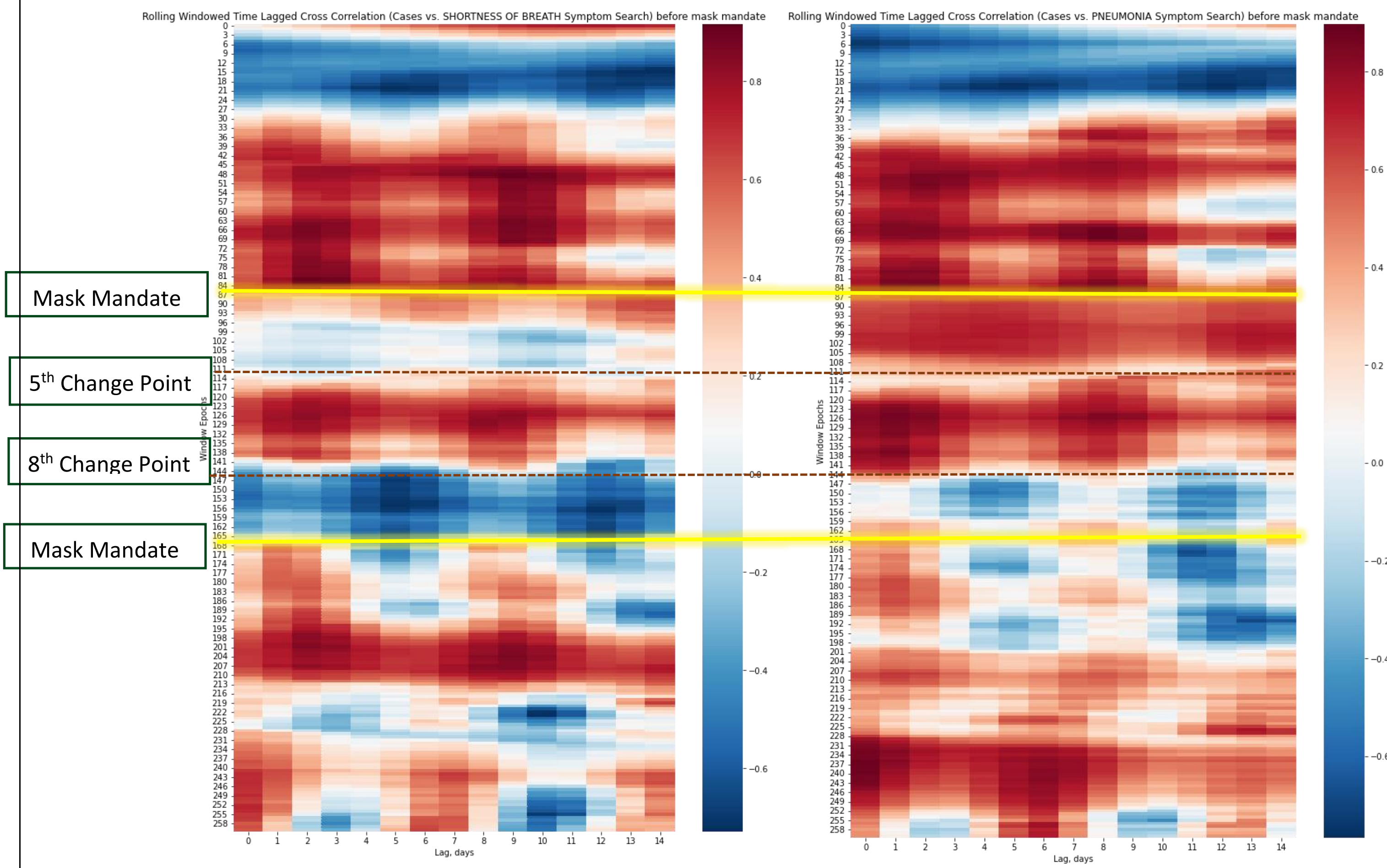

Figure 11: Correlation windows of SHORTNESS OF BREATH (**11a**) and PNEUMONIA (**11b**) symptom searches

Key Takeaways:

The SHORTNESS OF BREATH search term is strongly correlated with a lag of 4 days when there is no mask mandate. However, this correlation is not seen much during the mask mandate except between the 5th and 8th change point period caused by the 2nd wave of COVID-19 (Red region in the middle of Figure 11a). The PNEUMONIA search term is highly correlated at day 11 with daily fatalities and is not impacted much due to the mask mandate (red regions in Figure 11b). These two figures collectively show the variation in the association between the symptom searches with confirmed cases over time.

Causal Inference – Vaccination Intent and side-effects

The visualization Figure 12 represents the change points in vaccination search terms, vaccine intent, and side effects search. The vaccination-related national, state, and local county news is collected, and the events with high significance observed in the causal test are represented along with the change points below.

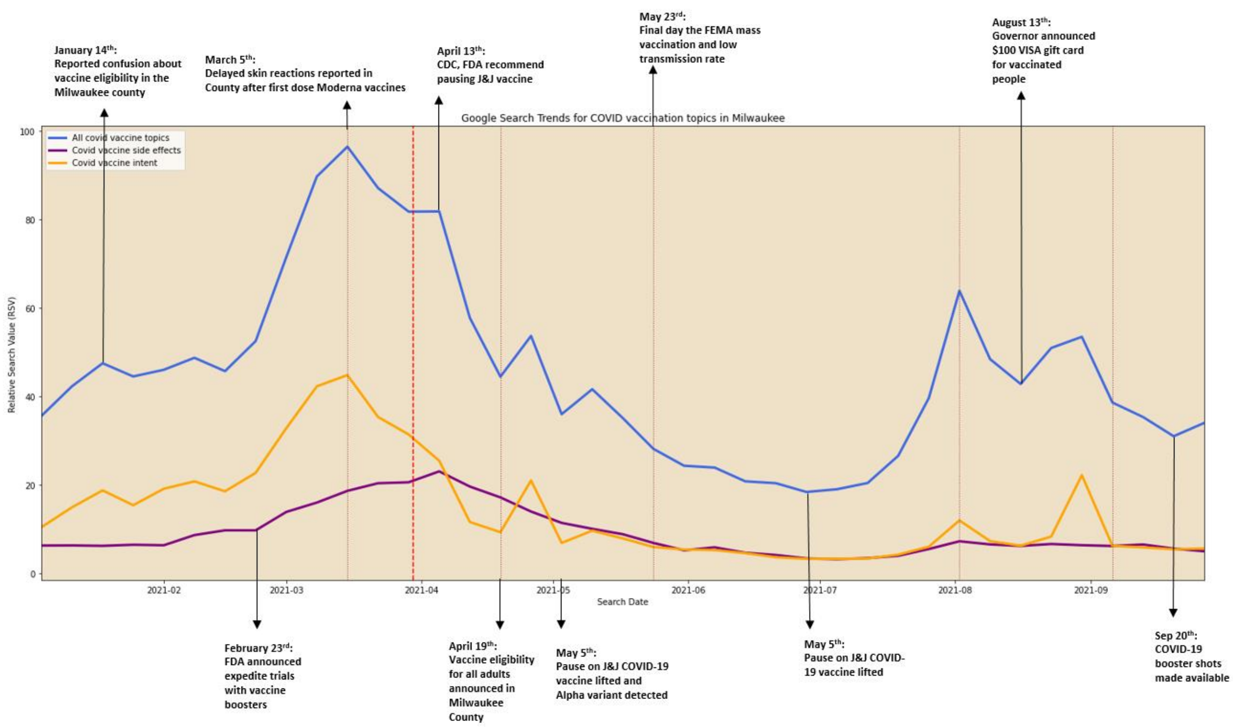

Figure 12: Vaccination causal inference graph (with change points)

Key Takeaways:

Most of the change points with positive slopes are aligned with the positive news regarding the availability, eligibility, and promotions like coupons announced by the government. Change points at negative sloped are associated with the negative news regarding the side effects and CDC pausing the vaccine administration. News about the COVID-19 variants also strongly impacted the vaccination search trends in Milwaukee County.

Discussion/Implications

How could future research build on this study? (**color-coded)**

Ways human-centered data science principles informed decision-making (**color-coded**)

Examining research that uses nontraditional data sources like Google trends has several implications. The results of the study demonstrated that Google could potentially be used as a complementary tool to aid in understanding online search behavior, which could help mitigate the adverse effects of the pandemic and quantify the impact of policy changes by the government. The study also proved that internet search patterns reveal a robust temporal pattern of disease progression for COVID-19.

The search data can be used to track and predict the local spread of COVID-19 before widespread laboratory testing becomes available and help guide current public health responses. When used correctly, internet search patterns have several powerful advantages over testing, including the following: (1) surveillance data are available immediately when a new pandemic emerges, (2) data are available at a population scale in countries with sufficient internet access, (3) delays are minimal, as search data are available the same day, and (4) there is no need for individuals to travel to a testing location; people can stay at home, thereby avoiding travel costs.

Future research can be focused on checking the progression of symptom-related search terms over time to characterize the clinical course of COVID-19 by examining a range of possible search term–based definitions for initial symptom onset. This should be based on various combinations of the earliest peaking search terms and a detailed understanding of the stage of illness and the manifestations of COVID-19 in the local environment and over time. Additionally, there should be a mechanism to separate symptom searches for general/seasonal illnesses from that for COVID-19. The model predictions would be more accurate if the data were segregated without overlaps. This makes the data more independent of other biases as well.

Studies have indicated that the spread and severity of COVID-19 can be affected by local conditions, and search volume data can be a valuable complementary tool for studying potential local variations in disease presentation. Also, further research is necessary to determine if the lag detected in our study is related to the results of clinical studies that postulate 97.5% of symptomatic COVID-19 cases develop within 11.5 days after exposure [17].

For counties where the inflection curve has not yet occurred, the proposed methodology can act as a systematic approach for the governments to monitor the spread of the virus. They can analyze the Google queries in their country to foresee the best use of their hospital systems. When faced with a COVID-19-like pandemic in the future, policymakers can utilize the symptom search data to plan effective test-track-treat mechanisms in their respective counties. Using the results, citizens can also better understand the importance of vaccination in their county and not spread the misinformation about vaccines any further. It also avoids panic actions by people regarding the side effects of vaccines during upcoming pandemics.

Limitations

There are several limitations to Google symptom and vaccine search trend analysis. The first major limitation is that Google Trends does not give absolute numbers of searches, which makes understanding search patterns in the context of broader information-seeking trends more challenging. Also, the selection of keywords and their related spelling is a significant constraint of this research since they play a critical role in assuring the validity of the results. Because there is no standard reporting technique, various phrases have the same meaning, multiple meanings of the same term, and different acronyms. As a result, good keyword selection will make the study more accessible to a broader audience.

Second, a significant restriction of search data is the socioeconomic, regional, or other biases inherent in the local digital divide. While Google Trends accounts for the vast majority (74%) of all Internet searches, it excludes smaller search engines, which may add bias [18]. As a result, data from different search engine platforms must be obtained to capture a more diversified community of consumers. For example, an automated software [19] can increase the accuracy of data gathered and processed in nations with a high frequency of viral infection. Because the data evaluated is solely from users with internet access who are actively looking for one of the defined phrases in our research, sampling bias may exist.

Third, while Google Trends gives useful information about a person's interests and behavioral intent, it may not predict actual infection and vaccination behavior. The current study looks at how the Public Mask Mandate impacts the chance of searching for symptoms and vaccination rather than the Risk of getting infected or immunized against COVID-19. Even though symptom searches are strong predictors of infection, the link between the Public Mask Mandate and individual Risk of having COVID-19 is still indirect, offering only suggestive evidence for the relationship of interest. Furthermore, further research should be conducted to offer clear evidence of this association.

Finally, due to established symptom search tendencies, using basic regression theory to analyze the data without trend and seasonal correction may be insufficient to determine the underlying link between cases and search patterns in Milwaukee County. There are more intricate ways of seasonal adjustment, but each has its own set of difficulties and limitations. This work employed the Difference-in-Difference approach, which has both the advantage and disadvantage of simplicity.

Conclusion

The study reveals the advantages of digital surveillance and infodemiology using Google Trends to monitor an infectious disease like COVID-19 in Milwaukee County, WI. The correlation between Google Trends of symptom searches data and confirmed COVID-19 cases in the county represents the search behavior induced by general symptom-related concerns of the pandemic or health-seeking for COVID-19 illness. Symptoms like COUGH and FEVER are common for most seasonal illnesses, but the study proved that they are highly associated with the COVID-19 pandemic. The public mask mandate was enforced in many areas across the United States. Knowing the mask mandate's effectiveness on people's health is key to avoiding fatalities. The current study results in a positive impact of mask mandate on symptom searches like SHORTNESS OF BREATH and COUGH, but the mask mandate does not much impact a few severe health issues like PNEUMONIA.

Additionally, the symptoms searches like FEVER and COUGH showed a lag correlation of 5 days with COVID-19 confirmed cases which acts as a good indicator for the outbreak of the virus before imposing quarantines and mask mandates. PNEUMONIA has a high fatality lag correlation of 11 days, waving a red flag to the health agencies to be prepared to track and treat the citizens suffering from such symptoms in a specific county. Regarding the vaccination search patterns, misinformation RSVs were highest after FDA authorization and had multiple repeated spikes after subsequent vaccine announcements. General information-seeking terms peaked concurrently with increased vaccination uptake in Milwaukee County. Search interest has decreased with wider vaccine availability, despite many individuals in Milwaukee County remaining unvaccinated. GT can be used to monitor trends in public attitudes and misinformation regarding COVID-19 vaccines and further target education.

Given its effectiveness in decreasing infection, public mask-wearing might still be an appropriate policy response to future outbreaks. Overall, the study provides suggestive evidence about the benefits of wearing masks in public during the various stages of the COVID-19 pandemic. The study also highlights the relevance of public mask-wearing for the ongoing pandemic, where the vaccination rate is precarious, and access to vaccines is still limited in many countries.

References

[1] Chu, Derek K., et al. "Physical distancing, face masks, and eye protection to prevent person-to-person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis." *The Lancet* 395.10242 (2020): 1973-1987.

[2] Sarah Almukhtar, Aliza Aufrichtig, Anne Barnard, Matthew Bloch, Weiyi Cai, Julia Calderone, Keith Collins "Coronavirus (Covid-19) Data in the United States. "- https://github.com/nytimes/covid-19-data

[3] Jimenez AJ, Estevez-Reboredo RM, Santed MA, Ramos V. COVID-19 Symptom-Related Google Searches and Local COVID-19 Incidence in Spain: Correlational Study. J Med Internet Res. 2020 Dec 18;22(12):e23518. doi: 10.2196/23518. PMID: 33156803; PMCID: PMC7757783.

[4] Abbas M, Morland TB, Hall ES, El-Manzalawy Y. Associations between Google Search Trends for Symptoms and COVID-19 Confirmed and Death Cases in the United States. Int J Environ Res Public Health. 2021 Apr 25;18(9):4560. doi: 10.3390/ijerph18094560. PMID: 33923094; PMCID: PMC8123439.

[5] Nguyen, My. "Mask mandates and COVID-19 related symptoms in the US." *ClinicoEconomics and Outcomes Research: CEOR* 13 (2021): 757.

[6] Sulyok, Mihály, Tamás Ferenci, and Mark Walker. "Google Trends Data and COVID‐19 in Europe: Correlations and model enhancement are European wide." *Transboundary and Emerging Diseases* 68.4 (2021): 2610-2615.

[7] Jimenez A, Estevez-Reboredo R, Santed M, Ramos V COVID-19 Symptom-Related Google Searches and Local COVID-19 Incidence in Spain: Correlational Study J Med Internet Res 2020;22(12):e23518 URL: https://www.jmir.org/2020/12/e23518 DOI: 10.2196/23518

[8] Lippi, Giuseppe, Camilla Mattiuzzi, and Gianfranco Cervellin. "Google search volume predicts the emergence of COVID-19 outbreaks." *Acta Bio Medica: Atenei Parmensis* 91.3 (2020): e2020006.

[9] Kardeş, Sinan, et al. "Public interest in rheumatic diseases and rheumatologist in the United States during the COVID-19 pandemic: evidence from Google Trends." *Rheumatology international* 41.2 (2021): 329-334.

[10] James, Nick, Max Menzies, and Peter Radchenko. "COVID-19 second wave mortality in Europe and the United States." *Chaos: An Interdisciplinary Journal of Nonlinear Science* 31.3 (2021): 031105.

[11] Raina MacIntyre C, Costantino V, Chanmugam A. The use of face masks during vaccine rollout in New YorkCity and impact on epidemic control. Vaccine. 2021 Oct 8;39(42):6296-6301. doi: 10.1016/j.vaccine.2021.08.102. Epub 2021 Sep 16. PMID: 34538699; PMCID: PMC8443976.

[12] Aminikhanghahi S, Cook DJ. A Survey of Methods for Time Series Change Point Detection. Knowl Inf Syst. 2017 May;51(2):339-367. doi: 10.1007/s10115-016-0987-z. Epub 2016 Sep 8. PMID: 28603327; PMCID: PMC5464762.

[13] https://centre-borelli.github.io/ruptures-docs/

[14] https://en.wikipedia.org/wiki/Difference_in_differences

[15] https://towardsdatascience.com/four-ways-to-quantify-synchrony-between-time-series-data-b99136c4a9c9

[16] https://en.wikipedia.org/wiki/Granger_causality

[17] Información científica-técnica, enfermedad por coronavirus, COVID-19. *Ministerio de Sanidad Centro de Coordinación de Alertasy Emergencias Sanitarias.* [2020-06-10]. https://www.mscbs.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov-China/documentos/ITCoronavirus.pdf.

[18] Search engine market share. Accessed April 6, 2021. https://www.netmarketshare.com/search-engine-market-share.aspx.

[19] GeneralMills/pytrends: Pseudo API for Google Trends. *GitHub.* [2020-09-12]. https://github.com/GeneralMills/pytrends.

Data Sources

Data Source 1: COVID-19 Data from John Hopkins University

The cumulative confirmed case counts were gathered from the Kaggle repository of the John Hopkins University COVID-19 raw United States confirmed cases dataset.

Link - https://www.kaggle.com/antgoldbloom/covid19-data-from-john-hopkins-university?select=RAW_us_confirmed_cases.csv

Data Source 2: U.S. State and Territorial Public Mask Mandates From April 10, 2020, through August 15, 2021, by County by Day

The data for masking mandates was sourced from the CDC dataset of masking mandates by county.
Link -
https://github.com/aaliyahfiala42/data-512-a7#us-state-and-territorial-public-mask-mandates-from-april-10-2020-through-august-15-2021-by-county-by-day-mask-use-by-countycsv

https://data.cdc.gov/Policy-Surveillance/U-S-State-and-Territorial-Public-Mask-Mandates-Fro/62d6-pm5i.

Data Source 3: COVID-19 Search Trends symptoms dataset
The dataset consists of aggregated, anonymized trends in Google searches for more than 400 health symptoms, signs, and conditions, such as cough, fever, and difficulty breathing. The dataset provides a time series for each region, showing each symptom's relative volume of searches.
Link - https://github.com/GoogleCloudPlatform/covid-19-open-data/blob/main/docs/table-search-trends.md

Data Source 4: COVID-19 Vaccination Search Insights
Description - This aggregated, anonymized data shows trends in search patterns related to COVID-19 vaccination. These trends in search patterns are made available with the intention of helping design, target, and evaluate public education campaigns. These trends reflect the *relative interest* in Google searches related to COVID-19 vaccination.
Link - https://github.com/GoogleCloudPlatform/covid-19-open-data/blob/main/docs/table-vaccination-search-insights.md