Words VS. Infections: What parameters determine public interest in infection?

Jennifer Roberts

Fall 2020

IDEMA

Abstract

A study was conducted to determine if the interest in SARS-CoV-2 only occurred after the number of confirmed positive cases in respective communities exceeded specific thresholds (significant spread occurred), or if outbreaks occurred. It was also hypothesized that public interest would not peak until after the first death had occurred in each state, as this severe symptom of disease would indicate a fairly significant risk to public health. Data for confirmed positive cases and confirmed deaths as a result of SARS-CoV-2 were collected from the CDC, John Hopkins University, and state government reports. Public interest in the virus was determined from data downloaded from Google Trends. All data was collected from January 1, 2020 until September 30, 2020. The results showed that public interest actually peaked before the majority of states achieved 100, 500, or 1,000 positive case thresholds. It was also found that public interest peaked before most states experienced the first death, indicating that significant spread and death tolls did not result in an increased public interest in SARS-CoV-2. It was also found that the peak interest time was the same for almost all states, regardless of outbreak timing, indicating that significant outbreaks played little to no role in public interest, which was unexpected. It was apparent that public interest in SARS-CoV-2 had peaked before the disease had a significant impact on each state locally, indicating that the public had interest in SARS-CoV-2, even before it affected them directly, unlike past infectious disease pandemics or epidemics.

Introduction

SARS-CoV-2 is a new respiratory infection that has led to millions of hospitalizations and deaths combined worldwide. It has been commonly spoken of in the news media and through the public as “coronavirus,” “COVID,” “COVID-19,” and “the virus.” In this study, it will be referred to as SARS-CoV-2, as this is a scientifically descriptive name used to refer to this virus. According to John Hopkins University Coronavirus Resource Center Map, as of October 23, 2020, there have been over 42,000,000 cases of SARS-CoV-2 globally, and over 1,100,000 deaths. In the United States, there has been over 8,400,000 cases and 223,000 deaths. Over 331,000 individuals have recovered in the United States alone (See Supplemental Table 1). There have been over 128,900,000 tests conducted and documented in the United States to this date.[1] Details of the virus composition has been well documented in the literature.[4]

According to the CDC’s website accessed on October 23, 2020, a wide range of symptoms have been reported from confirmed positive and symptomatic cases of SARS-CoV-2. The CDC list of SARS-CoV-2 symptoms includes fever, chills, cough, shortness of breath, difficulty breathing, fatigue, muscle or body aches, headaches, ageusia, hyposmia or anosmia, sore throat, congestion, rhinitis, nausea, vomiting, or diarrhea. In addition to this symptom list, the CDC lists symptoms that are considered an emergency and the individual suffering from symptoms in that list should seek immediate medical attention. Those symptoms include trouble breathing, persistent chest pain, a feeling of pressure in or on the chest, confusion, sleepiness, or blueish lips or face.[5] The literature tends to agree with this list of symptoms, with some adding pneumonia or death to the list.[6, 7, 8] This list is significant as it is possible that some individuals will search their symptoms online in an attempt to find out if they need to go to see their doctor, or even to get a diagnosis. Currently there is little to no data in the literature that compares how often the internet is used to search for symptoms of infectious diseases such as SARS-CoV-2, and more research is needed to analyze public behavior and reaction to disease pandemics.

While there are many cases that are symptomatic, the number of confirmed positive cases only accounted for 2% of each state population or less during the pandemic from January 1, 2020 until September 30, 2020.[5] In addition, many of these cases were asymptomatic or only mildly symptomatic (approximately 95%), and did not require hospitalization. According to the CDC, of all of the cases reported in the United States as of September 30, 2020, 408,649 people had been hospitalized, and 20,390 individuals had entered the ICU for treatment of the novel coronavirus. Only 2,319 people had required a ventilator.[24] This indicated that of all of the positive cases for SARS-CoV-2, only 5% were hospitalized, and less than 1% required ICU care or a ventilator. These percentages of hospitalizations, ICU care, and ventilators, while still unsatisfactory in our modern era, are low enough that public interest may not have been significant until after a significant number of cases threshold had been achieved or more deaths had occurred, as this study plans to investigate.

A study conducted by Mizumoto et al. found that in a sample of 3063 people, 634 tested positive. Of those 634 cases, 328 were counted as asymptomatic.[2] Another study by Rivett et al. showed similar results in a sample of 1032 healthcare workers. They concluded that over 57% of the healthcare workers were truly asymptomatic.[3] Therefore, considering the massive size of the population of the United States, the fraction of people negatively affected by the disease is small, and should not create mass panic through the public spectrum. However, this has not been the case for SARS-CoV-2, which has resulted in mask wearing mandates, as well as shutdowns of schools and businesses.

While there is a vast wealth of information regarding SARS-CoV-2 clearly documented in the scientific literature from reputable sources, the public continues to use non-peer reviewed sources to access information. According to Jardine et al, use of internet to find out information regarding disease had increased from 25% during SARS to 56% during the H1N1 flu pandemic.[9]

This study will compare the public's interest in SARS-CoV-2 to the spread of the actual disease. It is hypothesized that the spread of interest in information regarding an infectious disease (such as SARS-CoV-2) will reach its peak after the disease has either a) experienced a significant outbreak, b) significantly spread through the local population, or c) a first death or many deaths have occurred. It is proposed that the public will not show significant interest in infectious diseases that can result in extreme symptoms such as death until after one or more of these proposed criteria have been met.

Methods

This study was conducted in two parts. The first part of the study collected data of the cumulative number of cases per day per state to determine when outbreaks had occurred. This data was also used to determine when states achieved case milestones such as 100, 500, 1000, or 10,000 positive confirmed cases cumulatively from the CDC to determine the extent to which the virus was spreading through different populations.[5] Some states did not achieve greater than 25,000 cases in the study, which was attributed to state population size as reported by the United States Census Bereau.[15] As a result, measure of intensity was calculated by dividing the number of cases in a particular state[5] by the number of individuals in the population of that state in total[15] and multiplying by 100,000 so that states of both large and small population sizes could be compared. The dates of first death and cumulative number of deaths were also collected by state by day to determine if death played a role in public interest in the virus. The second part of the study used Google Trends[13] to determine when public interest in SARS-CoV-2 peaked. After all of the data had been collected, it was compared to determine if public interest peaked after a significant outbreak occurred, if public interest peaked after significant spread occurred through a particular population, or if public interest peaked after any deaths had occurred.

First, the study needed to determine how many confirmed positive cases of SARS-CoV-2 were recorded by state over the course of the SARS-CoV-2 pandemic. The cumulative data was collected from the CDC[5] and graphed using Microsoft Excel (Supplemental Figure 1). The states were divided into upper, middle, and lower quartiles based on their total final cumulative case numbers from January 1, 2020 to September 30, 2020 for ease of comparison. The top 10 states with the most cumulative cases were compared together forming the upper quartile, the middle 30 states were compared together forming the middle quartile, and the bottom 10 states were compared together, forming the lower quartile.

Next, the study tracked the dates in which each state achieved a specified number of cases (Table 1) in order to determine if significant spread resulted in an increase in public interest in SARS-CoV-2. The study looked at the dates in which all states achieved levels of cumulative positive confirmed cases at intervals of 100 cases, 500 cases, 1,000 cases, and 10,000 cases. The date in which each state achieved each cumulative case count milestone was compared to the date of peak interest on Google Trends which was determined to be 100% interest. The cumulative number of deaths were also tracked to determine if deaths played a role in stimulating public search interest (Table 1).

After the cumulative cases over time were established, a measure of intensity was needed so that states with high populations of susceptible individuals could be compared to states with low populations of susceptible individuals. Since SARS-CoV-2 is new, it was assumed that all members of the population were susceptible at the start of the pandemic. The number of new cases per 100,000 individuals per state per day were calculated by dividing the number of cases cumulatively for that state by the population of that state and multiplying by 100,000. The number of new cases per day was compared to the state population to observe severity of outbreaks, regardless of population size (Figure 1). States were graphed in groups of upper, middle, and lower quartiles based on their final cumulative number of cases for ease of comparison between states with large numbers of cases and states with low numbers of cases. This data was then compared to interest data to determine if outbreak peaks correlated with interest peaks.

Throughout the study, data was collected from Google Trends to evaluate how the terms relating to SARS-CoV-2 were searched and how those terms compared to other popular searches on the internet. Interest in all search terms were performed with a state (or the United States overall) selected, and over the same time period (January 1, 2020 to September 30, 2020). All searches were performed with “all categories” and “web search” options selected. Search terms were compared in groups of 5 terms together, or as single terms alone. Data was downloaded, graphed, and analyzed using Microsoft Excel.

Results

Table 1. **Dates of Cumulative Case Milestones by State.** The peak interest in SARS-CoV-2 was recorded from Google Trends data from January 1, 2020 until September 30, 2020. Peak interest was counted as 100% interest. Columns recording 100, 500, 1,000, 10,000, and 25,000 cases were based on cumulative number of cases. Dates indicate when each state achieved that milestone of case count. Blank boxes indicate that the state did not reach the threshold number of cases by September 30, 2020. The CDC reported the first case of SARS-CoV-2 confirmed by laboratory testing in the United States on January 22, 2020.[14, 15] In this report and in the CDC database, some US territories were excluded due to lack of data and reporting. John Hopkins University Coronavirus Dashboard[1] and Google Trends[13] started reporting data relating to SARS-CoV-2 for US territories and other countries on and after March 6, 2020.

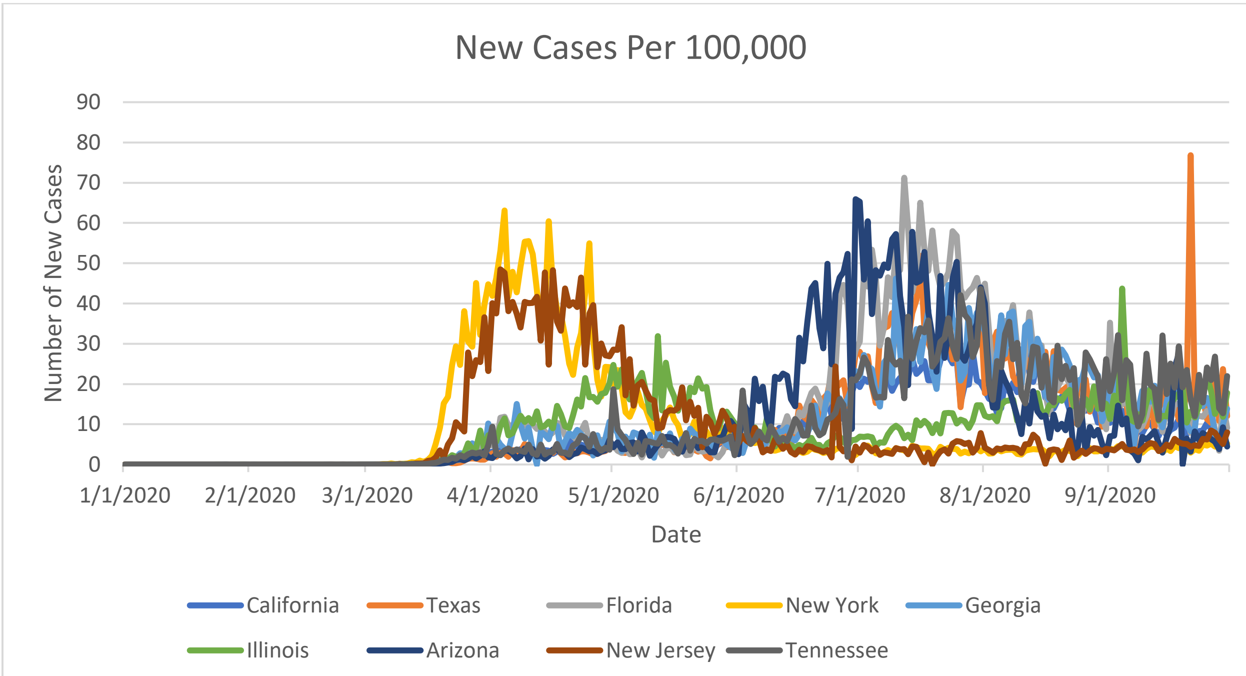
State Name	Peak Interest (100%)	First Death	100 cases	500 cases	1000 cases	10000 Cases	25000 Cases
Alabama	3/13/2020	3/25/2020	3/20/2020	3/26/2020	4/1/2020	5/11/2020	6/14/2020
Alaska	3/13/2020	3/27/2020	3/28/2020	6/3/2020	6/26/2020		
Arizona	3/15/2020	3/20/2020	3/21/2020	3/26/2020	3/30/2020	5/8/2020	6/6/2020
Arkansas	3/12/2020	3/24/2020	3/20/2020	3/30/2020	4/8/2020	6/9/2020	7/8/2020
California	3/15/2020	3/4/2020	3/7/2020	3/16/2020	3/19/2020	4/2/2020	4/14/2020
Colorado	3/15/2020	3/12/2020	3/14/2020	3/22/2020	3/25/2020	4/20/2020	5/28/2020
Connecticut	3/12/2020	3/18/2020	3/19/2020	3/24/2020	3/26/2020	4/10/2020	4/26/2020
Delaware	3/12/2020	3/26/2020	3/24/2020	4/4/2020	4/8/2020	6/9/2020	
Florida	3/15/2020	3/6/2020	3/15/2020	3/20/2020	3/22/2020	4/3/2020	4/18/2020
Georgia	3/12/2020	3/12/2020	3/16/2020	3/21/2020	3/24/2020	4/8/2020	4/30/2020
Hawaii	3/15/2020	3/31/2020	3/26/2020	4/13/2020	7/5/2020	9/8/2020	

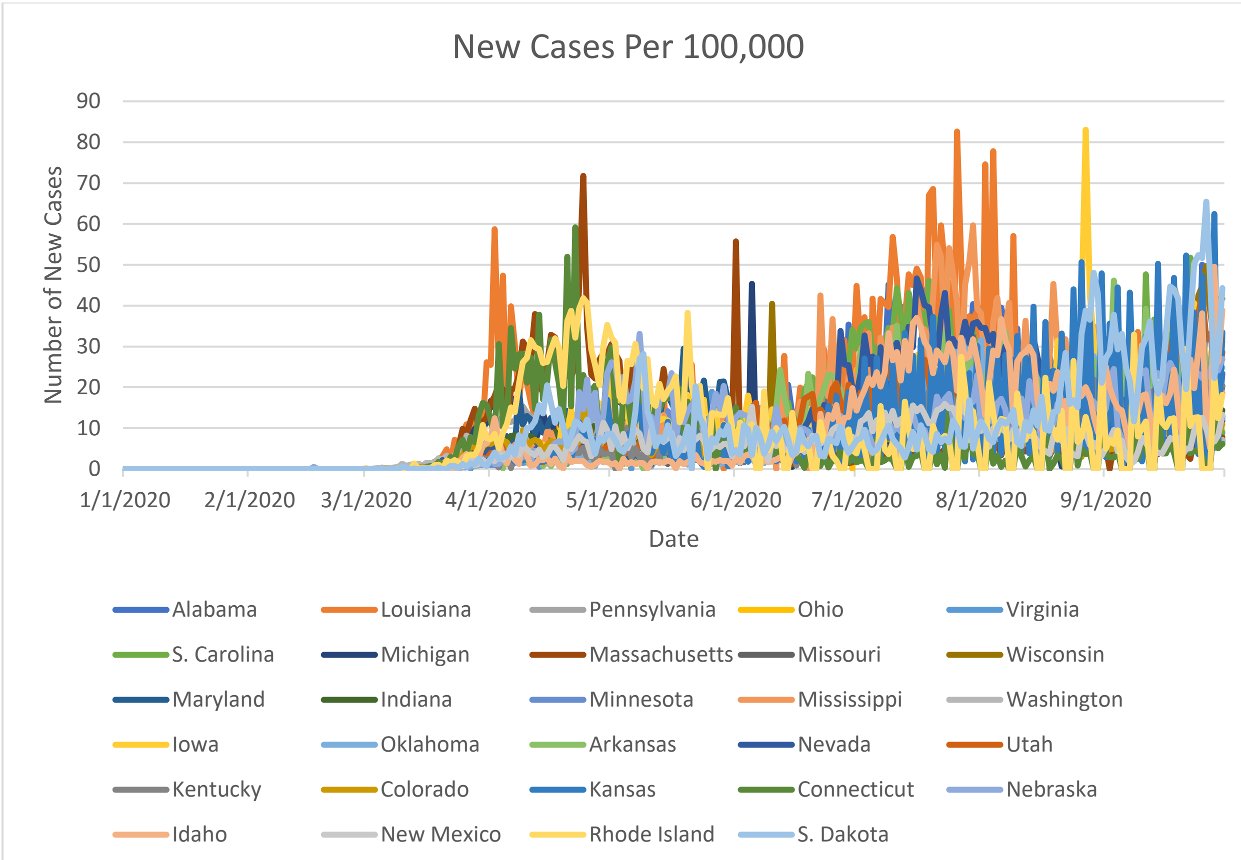
Idaho	3/15/2020	3/26/2020	3/25/2020	3/31/2002	4/3/2020	7/10/2020	8/10/2020
Illinois	3/15/2020	3/17/2020	3/16/2020	3/20/2020	3/22/2020	4/4/2020	4/16/2020
Indiana	3/16/2020	3/16/2020	3/21/2020	3/26/2020	3/28/2020	4/17/2020	5/11/2020
Iowa	3/16/2020	3/24/2020	3/23/2020	4/1/2020	4/7/2020	5/5/2020	6/19/2020
Kansas	3/15/2020	3/12/2020	3/24/2020	4/2/2020	4/8/2020	6/1/2020	7/24/2020
Kentucky	3/12/2020	3/16/2020	3/22/2020	3/31/2020	4/5/2020	6/1/2020	7/22/2020
Louisiana	3/15/2020	3/14/2020	3/15/2020	3/20/2020	3/23/2020	4/3/2020	4/22/2020
Maine	3/12/2020	3/27/2020	3/23/2020	4/7/2020	4/26/2020		
Maryland	3/12/2020	3/18/2020	3/19/2020	3/26/2020	3/28/2020	4/15/2020	5/3/2020
Massachusetts	3/12/2020	3/20/2020	3/12/2020	3/21/2020	3/24/2020	4/3/2020	4/12/2020
Michigan	3/15/2020	3/18/2020	3/19/2020	3/20/2020	3/22/2020	4/2/2020	4/13/2020
Minnesota	3/15/2020	3/21/2020	3/20/2020	3/29/2020	4/7/2020	5/8/2020	
Mississippi	3/15/2020	3/19/2020	3/21/2020	3/27/2020	4/1/2020	5/13/2020	6/26/2020
Missouri	3/15/2020	3/18/2020	3/22/2020	3/26/2020	3/30/2020	5/11/2020	7/7/2020
Montana	3/15/2020	3/27/2020	3/27/2020	5/30/2020	7/1/2020	9/19/2020	
Nebraska	3/15/2020	3/27/2020	3/27/2020	4/8/2020	4/16/2020	5/16/2020	7/28/2020
Nevada	3/15/2020	3/16/2020	3/20/2020	3/26/2020	3/30/2020	6/9/2020	7/9/2020
New Hampshire	3/15/2020	3/23/2020	3/23/2020	4/3/2020	4/13/2020		
New Jersey	3/15/2020	3/10/2020	3/16/2020	3/19/2020	3/21/2020	3/28/2020	4/2/2020
New Mexico	3/15/2020	3/25/2020	3/24/2020	4/4/2020	4/10/2020	6/17/2020	8/28/2020
New York	3/12/2020	3/14/2020	3/8/2020	3/14/2020	3/17/2020	3/21/2020	3/24/2020
North Carolina	3/15/2020	3/25/2020	3/19/2020	3/25/2020	3/29/2020	4/30/2020	5/28/2020
North Dakota	3/15/2020	3/27/2020	3/30/2020	4/18/2020	4/29/2020	8/24/2020	
Ohio	3/15/2020	3/20/2020	3/19/2020	3/24/2020	3/27/2020	4/18/2020	5/12/2020
Oklahoma	3/12/2020	3/19/2020	3/24/2020	3/31/2020	4/4/2020	6/20/2020	7/18/2020
Oregon	3/15/2020	3/14/2020	3/20/2020	3/29/2020	4/4/2020	7/5/2020	8/24/2020
Pennsylvania	3/15/2020	3/18/2020	3/17/2020	3/22/2020	3/25/2020	4/4/2020	4/14/2020
Rhode Island	3/15/2020	3/28/2020	3/23/2020	4/1/2020	4/6/2020	5/6/2020	
South Carolina	3/15/2020	3/16/2020	3/20/2020	3/27/2020	3/31/2020	5/24/2020	6/22/2020
South Dakota	3/12/2020	3/10/2020	3/30/2020	4/10/2020	4/15/2020	8/14/2020	
Tennessee	3/15/2020	3/21/2020	3/19/2020	3/23/2020	3/27/2020	4/29/2020	6/3/2020
Texas	3/12/2020	3/16/2020	3/17/2020	3/21/2020	3/25/2020	4/9/2020	4/26/2020
Utah	3/12/2020	3/22/2020	3/20/2020	3/28/2020	4/1/2020	6/1/2020	7/5/2020
Vermont	3/15/2020	3/19/2020	3/25/2020	4/5/2020	6/4/2020		
Virginia	3/15/2020	3/14/2020	3/20/2020	3/26/2020	3/30/2020	4/22/2020	
Washington	3/15/2020	2/29/2020	3/7/2020	3/13/2020	3/18/2020	4/11/2020	6/7/2020
West Virginia	3/13/2020	3/29/2020	3/28/2020	4/9/2020	4/24/2020	8/30/2020	
Wisconsin	3/14/2020	3/19/2020	3/18/2020	3/25/2020	3/28/2020	5/9/2020	6/13/2020
Wyoming	3/12/2020	4/13/2020	3/31/2020	4/26/2020	6/11/2020		

The states with the highest cumulative number of cases (Supplemental Figure 1) had two distinctly large outbreaks. The first occurred in April and the second in July (Figure 1). Interestingly, there was no secondary peak of interest of significant magnitude (Figure 2) that coincided with a second major outbreak for any of the states for search terms relating to SARS-CoV-2. States in the middle and lower population groups also experienced outbreak peaks in April, but only North Dakota had a large second outbreak among the lower quartile states, while the other states in that category had much smaller peaks. Louisiana and Iowa had the highest second outbreak peaks in the middle quartile states. Massachusetts, Michigan, and Wisconsin all had peaks in June, while other states did not experience similar peak outbreaks at that time. New Jersey did not appear to experience a second peak at all, and maintained a steady rate of new cases after its first major peak outbreak. New York, Vermont, and New Hampshire followed the same trend. All of these large outbreaks occurred after public interest had peaked according to Google Trends.[13] Delaware had its first major outbreak peak one month after interest in the state had peaked (Figures 1 and 2). This negates the hypothesis that the highest percent interest would occur after a significant outbreak.

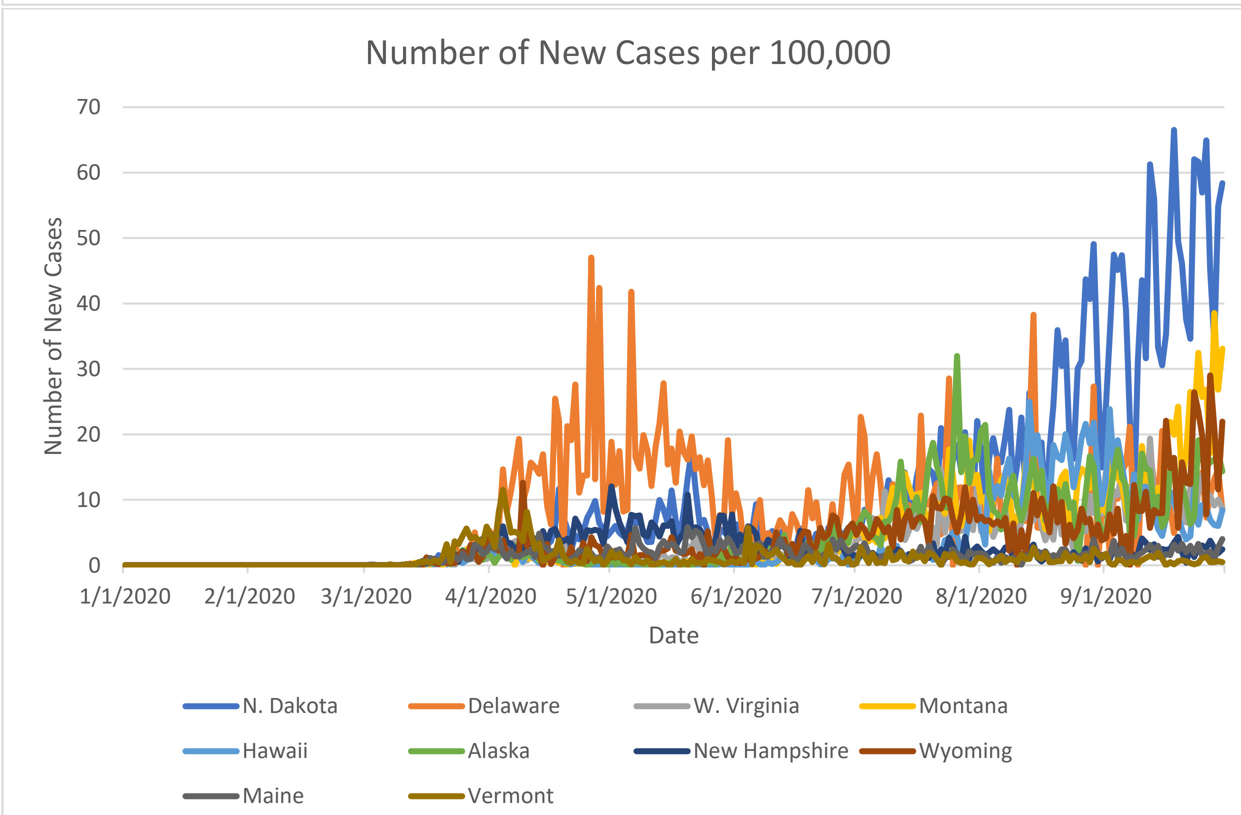
Figure 1. New Cases Per 100,000 of SARS-CoV-2. New cases daily per 100,000 individuals per state of SARS-CoV-2. A. states in the upper quartile (highest cumulative number of positive confirmed cases). B. states in the middle quartile (middle cumulative number of positive confirmed cases). C. states in the lower quartile (lowest cumulative number of positive confirmed cases).

a.




New Cases Per 100,000
Number of New Cases
0
10
20
30
40
50
60
70
80
90
1/1/2020
2/1/2020
3/1/2020
4/1/2020
5/1/2020
6/1/2020
7/1/2020
8/1/2020
9/1/2020
Date
Alabama
Louisiana
Pennsylvania
Ohio
Virginia
S. Carolina
Michigan
Massachusetts
Missouri
Wisconsin
Maryland
Indiana
Minnesota
Mississippi
Washington
Iowa
Oklahoma
Arkansas
Nevada
Utah
Kentucky
Colorado
Kansas
Connecticut
Nebraska
Idaho
New Mexico
Rhode Island
S. Dakota

b.

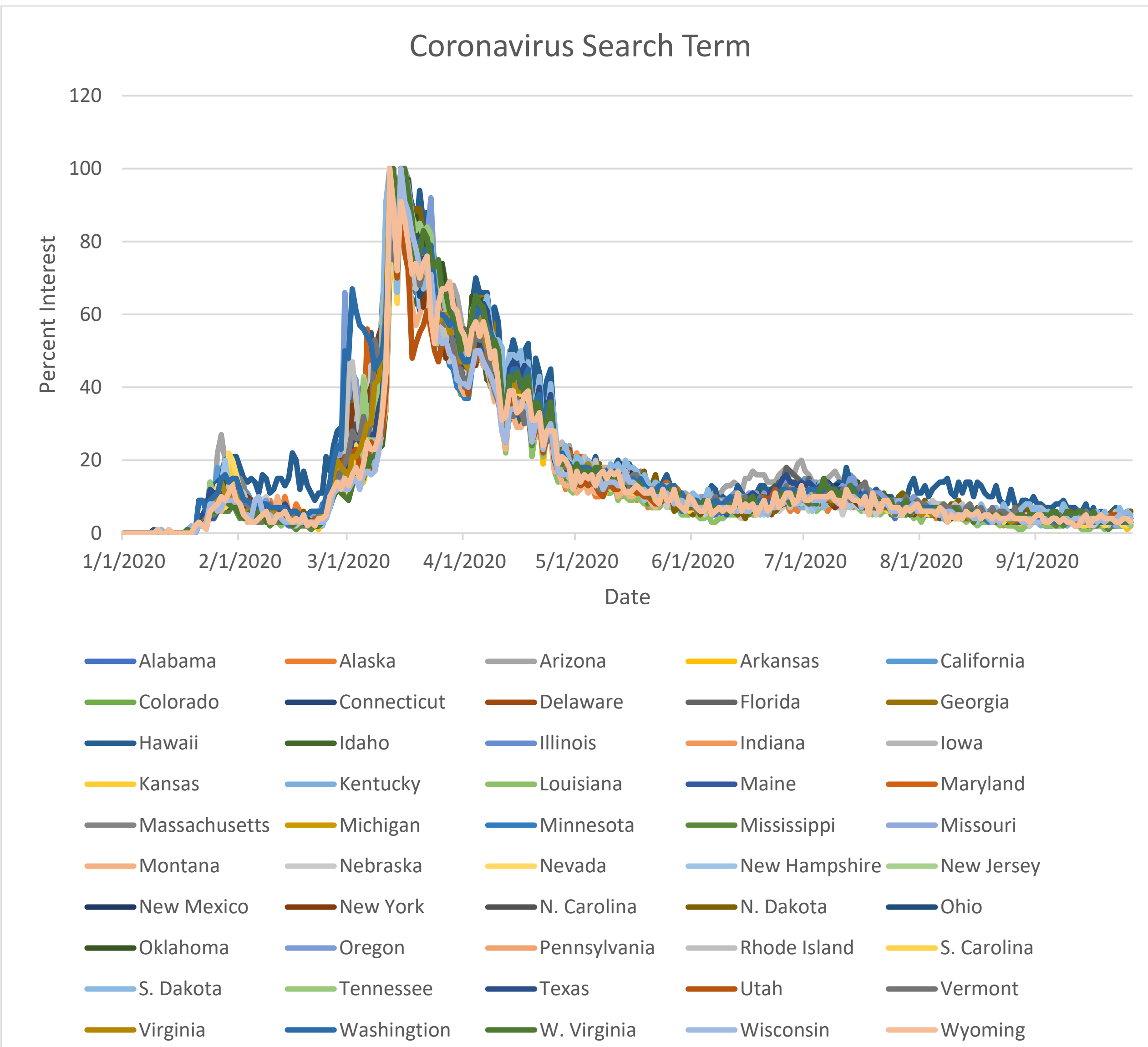

Number of New Cases per 100,000
Number of New Cases
0
10
20
30
40
50
60
70
1/1/2020
2/1/2020
3/1/2020
4/1/2020
5/1/2020
6/1/2020
7/1/2020
8/1/2020
9/1/2020
Date
N. Dakota
Delaware
W. Virginia
Montana
Hawaii
Alaska
New Hampshire
Wyoming
Maine
Vermont

c.

All states had achieved at least 100 cases by the end of March, and all states (with the exceptions of Alaska and Montana) had achieved 500 cases by the end of April (Table 1). Interestingly, interest peaked in all states even before 500 cases were achieved with the exception of Washington, regardless of state population size, governance, geographic location, or access to media (Supplemental Figures 2-3). Almost all states had peaked interest before 100 cases were achieved, with the exception of California, Colorado, New York, and Washington states. Florida, Louisiana, and Massachusetts interest peaked on the same day in which 100 cumulative cases were achieved. It was expected that interest would peak after each state reached these thresholds, but they actually peaked before these thresholds were met. This negated the hypothesis that significant spread of the virus played a role in increasing public interest in SARS-CoV-2.

The majority of states also had interest peak before the date of first death (Table 1). Only 10 out of 50 of the states had interest peak after first death. It was surprising because it was hypothesized that interest would peak after a death in the state had occurred. This negated the hypothesis that public interest in SARS-CoV-2 would peak after the first death occurred in local communities.

Figure 2. Coronavirus Search Interest. Percent interest over time in the term "coronavirus" on Google Trends.



To ensure that the study covered all possible searches relating to SARS-CoV-2, the symptoms relating to the disease were also searched. Terms such as “headache” and “vomiting” showed 1% or 0% interest throughout the pandemic, despite being listed as a potential symptom of SARS-CoV-2 on the CDC website. When terms such as “virus” or “corona” were compared to the specific symptoms, they greatly outweighed all of the possible symptom searches. This was surprising because it showed that individuals were significantly more interested in searching for SARS-CoV-2 itself rather than for symptoms relating to the virus. While symptoms are more vague and could possibly be resulting from other infectious agents, SARS-CoV-2 still had significantly captured public interest by close to 99% more than any other possible search terms (Figures 3 and 4).

Figure 3. Symptoms Search Interest. A. The search terms of symptoms related to SARS-CoV-2 as listed by the CDC. Only the four search terms listed showed significant interest at some point during the pandemic. The full list of possible SARS-CoV-2 symptoms according to the CDC includes fever, chills, cough, shortness of breath, difficulty breathing, fatigue, muscle or body aches, headaches, ageusia, hyposmia or anosmia, sore throat, congestion, rhinitis, nausea, vomiting, diarrhea, trouble breathing, persistent chest pain, a feeling of pressure in or on the chest, confusion, sleepiness, or blueish lips or face. Both scientific terminology and layman terminology were compared to generate these results.

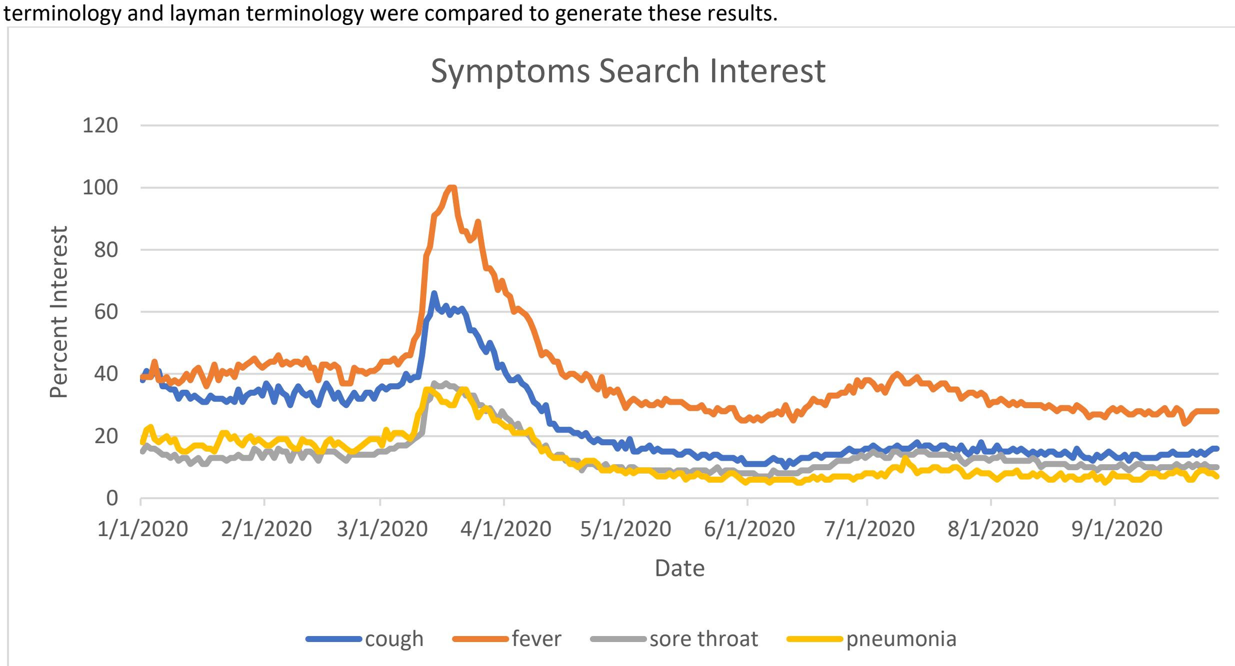
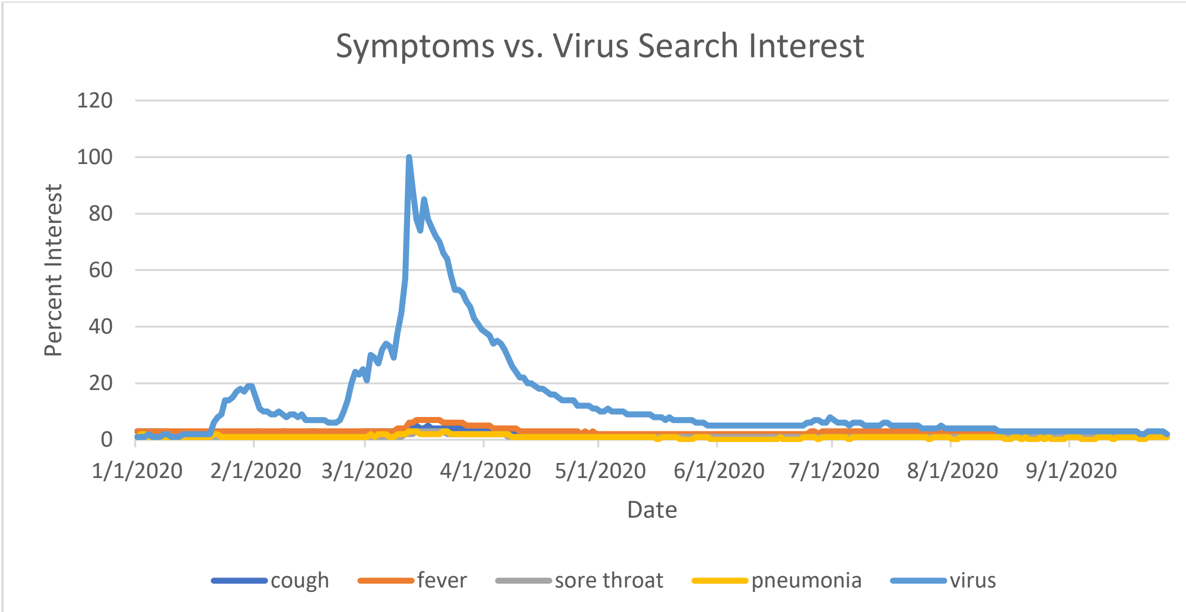
a.


Figure 4. Virus search interest compared to symptom search interest. Chart showing the result when the search terms relating to SARS-CoV-2 are compared to "virus" as a search term.



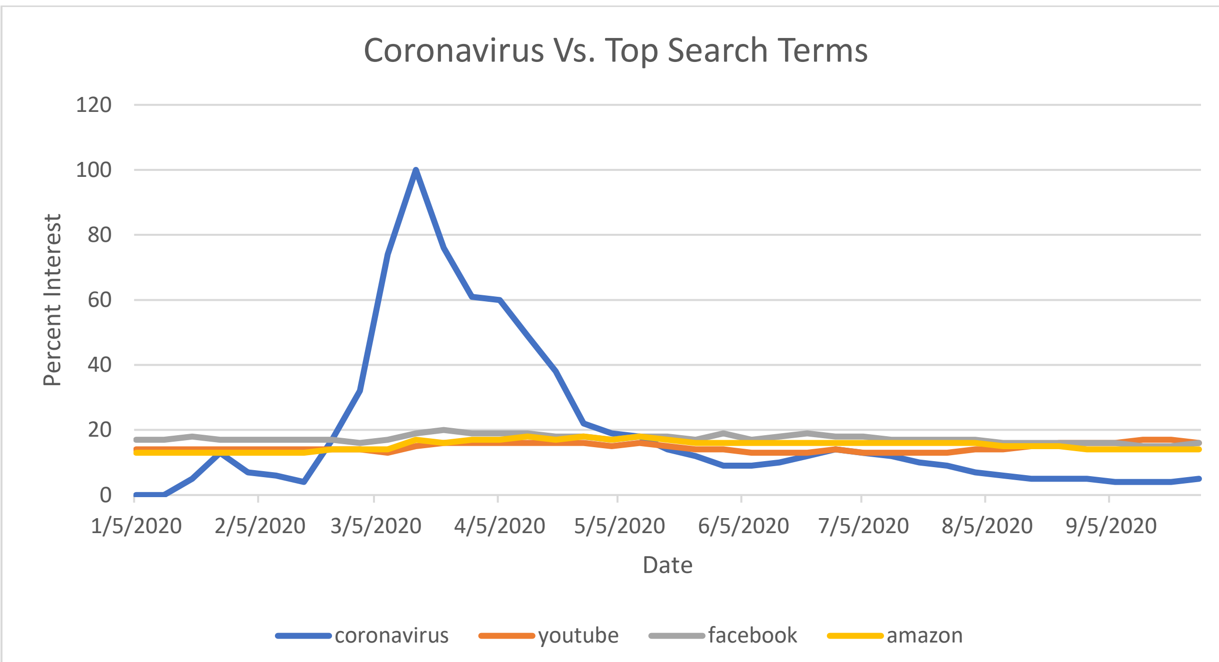
b.

The terms "coronavirus" and "virus" both peaked in March and continued through April. The same was true for the term "symptoms." It is possible that these peaks correlated because the public searched for some of these terms simultaneously (same combination of terms in one search). However, despite the term "symptoms" having the capability of covering virtually any infectious disease, it only achieved up to 20% interest in any given state, while the search term "coronavirus" achieved 100% in the same time period. Therefore, it is reasonable to conclude that the majority of searches for "symptoms" were attached to search terms relating to SARS-CoV-2 rather than other infectious diseases. This was surprising because other infectious diseases such as influenza were to date still more deadly than SARS-CoV-2 disease according to the CDC,[25] but did not get nearly as high spikes when compared with terms relating to SARS-CoV-2 on Google Trends from January 1, 2020 to September 30, 2020 (although both peaked in April).[15] It was expected that the public would find more concern over deadlier diseases, but this was clearly not the case. The public interest in SARS-CoV-2 far exceeded interest in seasonal flu. Therefore, comparative mortality rate did not play a role in determining search interest, and negated the hypothesis that number of deaths would increase public search interest.

When the term "coronavirus" was searched alone, it achieved a significant peak in March and April, but there were no other major peaks before or after. This was interesting because it did not align with the outbreak trends in most states. In fact, the interest peaked for all states around the same time. (Figure 2). Interestingly, despite the obvious differences in spread of the disease over time as shown by the date in which 100, 500, or 1000 cases were achieved in each state (Table 1), as well as the total cumulative number of cases, all states generally peaked interest in SARS-CoV-2 at the same relative timepoints (Figure 2). Washington, and Illinois all had their first case confirmations in January, but their peak interest did not occur until March 12-16, 2020. The same was true for Massachusetts, Wisconsin, Texas, Nebraska, Utah, and Oregon, which all had their first positive cases confirmed in February but did not show peak interest until March 12-16, 2020 (Supplemental Figure 4). All states had peak interest between March 12, 2020 and March 16, 2020. Arizona had the highest peak interest in SARS-CoV-2 in February 2020, and was the only state to have such a peak only shortly after their first case on January 26, 2020. Washington never had a peak interest that exceeded the interest of any other state throughout the pandemic, despite being the first state to confirm a positive case. California's interest did not peak until the last week of March, despite having its first confirmed case on January 25, 2020 and its first confirmed death on March 4, 2020. This data negates the hypothesis that an increase in cumulative number of cases led to an interest in SARS-CoV-2. In fact, the opposite had occurred. Interest had peaked prior to a rise in the number of cases in most all states, and before most outbreaks.

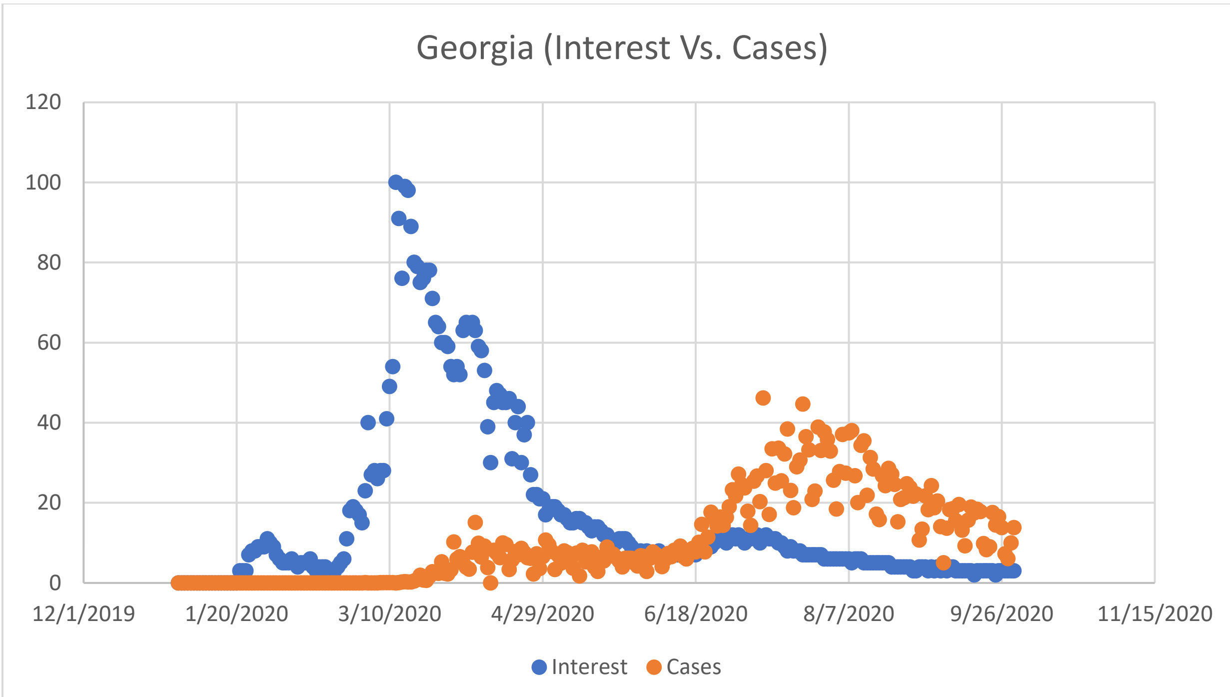
To find out how coronavirus searches compared to other unrelated search terms, Google Trends was consulted to determine the most popular searches on Google that were not related to the pandemic. According to Google Trends, the most popular searches not relating to SARS-CoV-2 included Youtube, Facebook, and Amazon. When each of these search terms were compared together, "coronavirus" was the only term that reached peak interest from January 1, 2020 to September 30, 2020. However, the other three search terms were more popular for the majority of the investigated time period. Coronavirus only exceeded interest over the other search terms from the end of February to the beginning of May (Figure 5). This peak of interest in SARS-CoV-2 was consistent regardless if the other search terms related to SARS-CoV-2, other infectious diseases, or completely unrelated terms. Therefore, public interest in the virus was not influenced or changed due to any other events or items of public interest over the course of the pandemic.

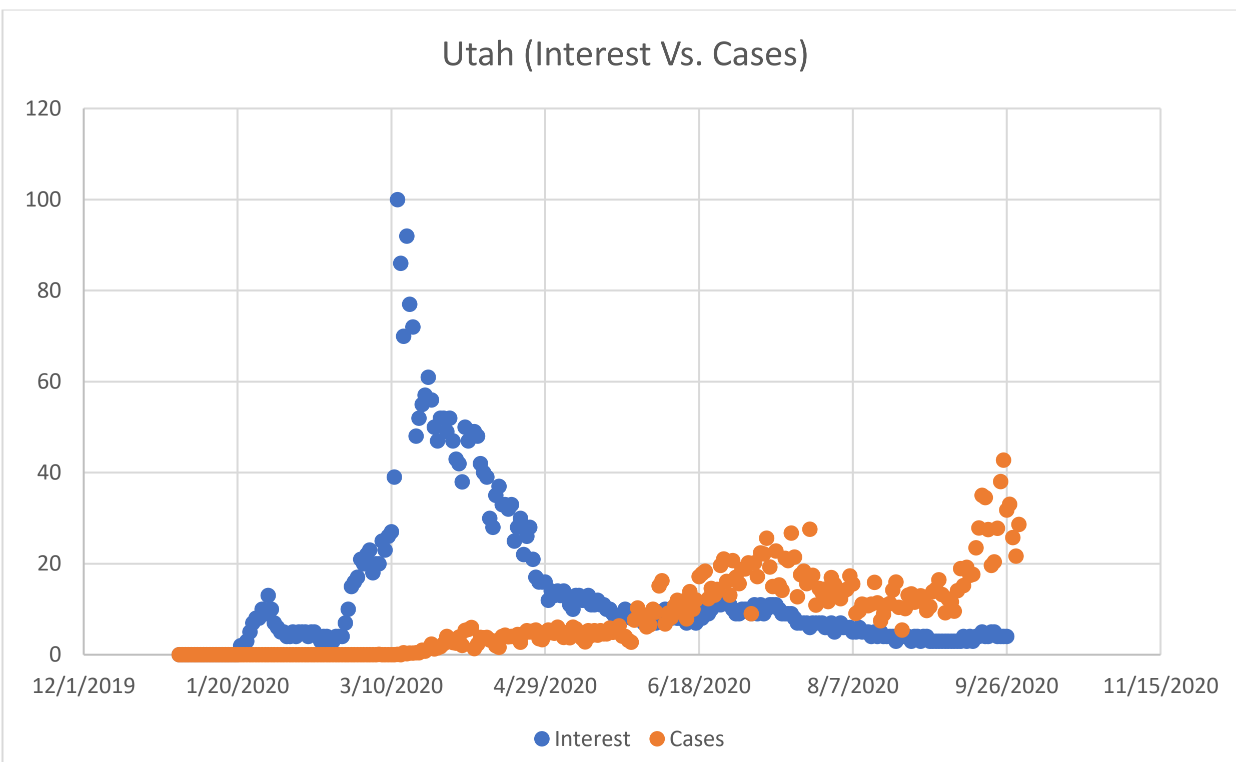
Figure 5. Search Term Coronavirus Compared to Top Search Terms. "Coronavirus" search term compared to other popular search terms of 2020.[13]

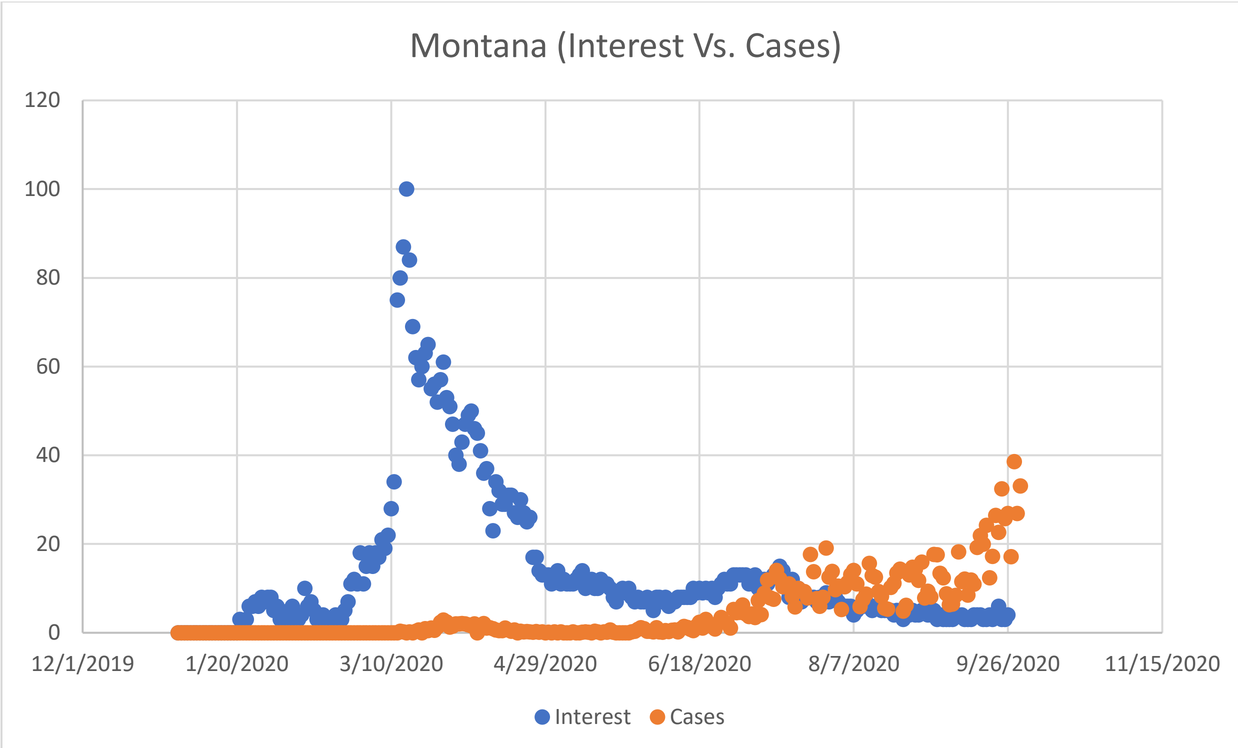


Lastly, the percent interest was compared to the number of cases per 100,000 individuals per state. The number of cases per state were normalized to a scale of 100 to match the scale of interest (percent out of 100). Both data sets were plotted together. There was a clear peak in interest that occurred before case increases in each state (Figure 6).

Figure 6. State Cases of SARS-CoV-2 Versus Interest. Representative graphs of state cases versus interest. The peak in interest occurred before the peak in cases in each state. A. Georgia state trend representative of states in the upper quartile for most cumulative confirmed cases of COVID-19. B. Utah state trend representative of states in the middle quartile for most cumulative confirmed cases of COVID-19. C. Montana state trend representative of states in the lower quartile for most cumulative confirmed cases of COVID-19.







Discussion

This study is the first to show that public interest in an infectious disease such as SARS-CoV-2 had come before a) significant outbreaks of the infectious disease, B) significant spread in the local population had occurred, and C) deaths had occurred and were reported to the public. It was reasonable to hypothesize that public interest would occur after scenarios A, B, or C (or multiple of the above) had occurred, however that is not what was shown by the data.

Overall, it is unclear as to why the public interest in SARS-CoV-2 spiked in March outside of major outbreaks or before significant spread had occurred in each state or before a death had occurred. It was also interesting to find that each state had the same timeline of peak interest. It would be interesting to explore the psychology of collecting information on infectious disease by use of public social media that is not peer reviewed. Perhaps laymen in the general population are not sure of where to collect information and go to whatever shows up first on their devices and need more education as to where to find valid information. Or, perhaps with our modern knowledge of infectious diseases, people are more ready and willing to research and understand it to better protect themselves from infection that could lead to serious symptoms ranging from a sneeze to pneumonia or death.

Google Trends, while helpful, does not report the exact number of searches per day that have been executed by public users. Rather, it normalizes the data and compares the number of searches over time. Google Trends calculates percent interest relative to the highest point on the chart for the given region and time. A value of 0 indicates that there was not enough data for the specified timeframe, and a value of 100 indicates when the search term was the most popular within that timeframe.[13] In addition, Google Trends only allows for a comparison of five search terms per run. Google Trends determined that the most common search terms associated with SARS-CoV-2 were COVID-19, corona virus, coronavirus cases, coronavirus symptoms, COVID, COVID 19 symptoms, coronavirus map, coronavirus update, coronavirus news, and COVID testing between January 1, 2020 and September 30, 2020. Because of the limited number of search terms, the top three terms relating to the novel coronavirus were selected for the study: coronavirus, coronavirus testing, and COVID-19. It would have been helpful if the data from Google Trends was not normalized and if the exact number of searches could be determined and worked with. Due to limitations of search term number, a true comparison between all of the possible search terms relating to the novel coronavirus could not be compared. However, by choosing the top three most common terms associated with the pandemic, it is likely that the data is still fairly representative of the search trends toward SARS-CoV-2 as made by public users.

The exact number of searches per search engine is debated in both internet blogs and scientific literature. For example, Internet Live Stats states that over 4.5 billion searches have been conducted on Google as of October 28, 2020,[22] while Statistica showed that over 12 billion Google searches have been conducted in the same timeframe.[23] Regardless of these differences, it is clear that the public is using search engines to gain information more so than it is using peer reviewed sources.

It is possible that as the number of cases rose globally, public interest in various states also increased. Other countries demonstrated significant outbreaks, spread, and deaths due to SARS-CoV-2 before it reached the United States.[1]

It is clear based on this evidence that there is a significant trend in the spread of information as well as the prevalence of disease through testing across the United States, whether accurate or not. It is unfortunate that the majority of the public gains information through search engines and blogs rather than scientific publications. Perhaps it is that we have become so strongly adhered to scientific language and literature that we have forgotten in some respects how to communicate our findings to the general public and laymen. While it is beneficial to be able to communicate on a highly intellectual level, it does not negate the fact that we must also be able to communicate our findings to the public in order to secure the future of public national health.

References

1. John Hopkins University. 2020. Center for Systems Science and Engineering. COVID-19 Dashboard. Retrieved from https://coronavirus.jhu.edu/map.html.
2. Mizumoto K, Kagaya K, Zarebski A, Chowell G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. Euro Surveill. 2020 Mar;25(10):2000180. doi: 10.2807/1560-7917.ES.2020.25.10.2000180. PMID: 32183930; PMCID: PMC7078829.
3. Rivett L et al. Screening of healthcare workers for SARS-CoV-2 highlights the role of asymptomatic carriage in COVID-19 transmission. Elife. 2020 May 11;9:e58728. doi: 10.7554/eLife.58728. PMID: 32392129; PMCID: PMC7314537.
4. Kakodkar P, Kaka N, Baig M (April 06, 2020) A Comprehensive Literature Review on the Clinical Presentation, and Management of the Pandemic Coronavirus Disease 2019 (COVID-19). Cureus 12(4): e7560. DOI 10.7759/cureus.7560
5. Centers for Disease Control and Prevention. 2020. Coronavirus Disease 2019 (COVID-19). Retrieved from https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html.
6. Struyf, T., Deeks, J. J., Dinnes, J., Takwoingi, Y., Davenport, C., Leeflang, M. M., Spijker, R., Hooft, L., Emperador, D., Dittrich, S., Domen, J., Horn, S., Van den Bruel, A., & Cochrane COVID-19 Diagnostic Test Accuracy Group (2020). Signs and symptoms to determine if a patient presenting in primary care or hospital outpatient settings has COVID-19 disease. *The Cochrane database of systematic reviews*, *7*(7), CD013665. https://doi.org/10.1002/14651858.CD013665
7. Stawicki, S. P., Jeanmonod, R., Miller, A. C., Paladino, L., Gaieski, D. F., Yaffee, A. Q., De Wulf, A., Grover, J., Papadimos, T. J., Bloem, C., Galwankar, S. C., Chauhan, V., Firstenberg, M. S., Di Somma, S., Jeanmonod, D., Garg, S. M., Tucci, V., Anderson, H. L., Fatimah, L., Worlton, T. J., … Garg, M. (2020). The 2019-2020 Novel Coronavirus (Severe Acute Respiratory Syndrome Coronavirus 2) Pandemic: A Joint American College of Academic International Medicine-World Academic Council of Emergency Medicine Multidisciplinary COVID-19 Working Group Consensus Paper. *Journal of global infectious diseases*, *12*(2), 47–93. https://doi.org/10.4103/jgid.jgid_86_20
8. Larici, A. R., Cicchetti, G., Marano, R., Merlino, B., Elia, L., Calandriello, L., Del Ciello, A., Farchione, A., Savino, G., Infante, A., Larosa, L., Colosimo, C., Manfredi, R., & Natale, L. (2020). Multimodality imaging of COVID-19 pneumonia: from diagnosis to follow-up. A comprehensive review. *European journal of radiology*, *131*, 109217. https://doi.org/10.1016/j.ejrad.2020.109217
9. Jardine, C. G., Boerner, F. U., Boyd, A. D., & Driedger, S. M. (2015). The More the Better? A Comparison of the Information Sources Used by the Public during Two Infectious Disease Outbreaks. *PloS one*, *10*(10), e0140028. https://doi.org/10.1371/journal.pone.0140028
10. Tang, X., Du, R. H., Wang, R., Cao, T. Z., Guan, L. L., Yang, C. Q., Zhu, Q., Hu, M., Li, X. Y., Li, Y., Liang, L. R., Tong, Z. H., Sun, B., Peng, P., & Shi, H. Z. (2020). Comparison of Hospitalized Patients With ARDS Caused by COVID-19 and H1N1. *Chest*, *158*(1), 195–205. https://doi.org/10.1016/j.chest.2020.03.032
11. Osztovits J, Balázs C, Fehér J. H1N1-influenza - pandémia, 2009 [H1N1 influenza - pandemic, 2009]. Orv Hetil. 2009 Dec 13;150(50):2265-73. Hungarian. doi: 10.1556/OH.2009.28766. PMID: 19951858.

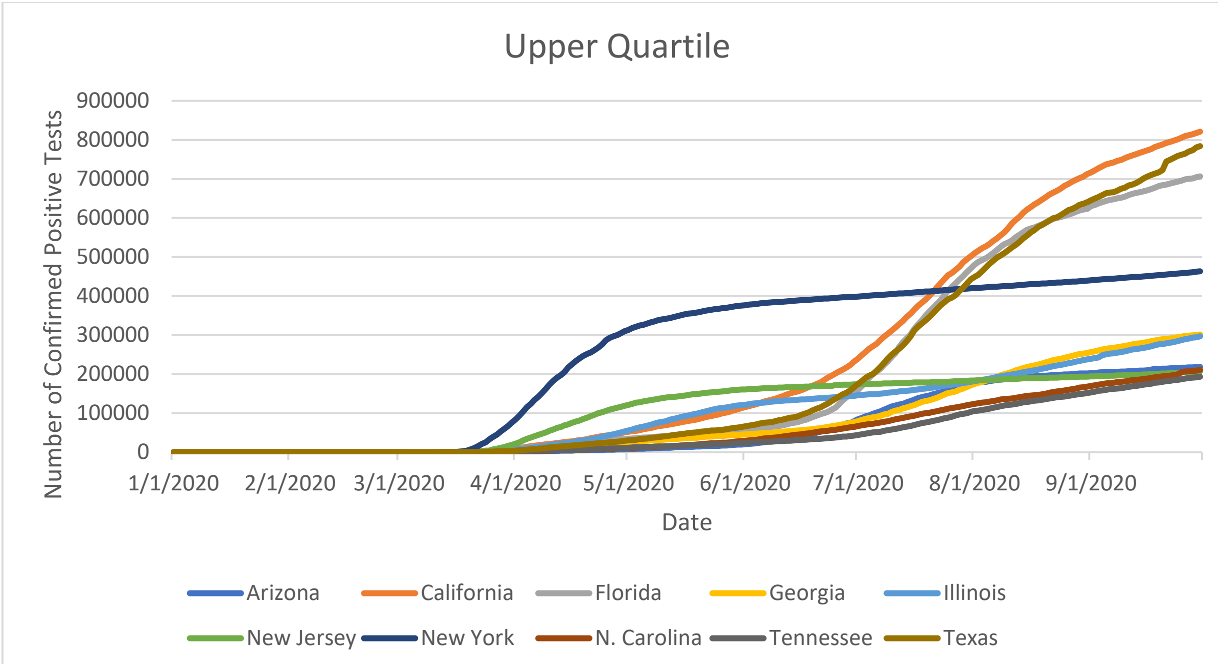
12. Banoei, M. M., Vogel, H. J., Weljie, A. M., Kumar, A., Yende, S., Angus, D. C., Winston, B. W., & Canadian Critical Care Translational Biology Group (CCCTBG) (2017). Plasma metabolomics for the diagnosis and prognosis of H1N1 influenza pneumonia. *Critical care (London, England)*, *21*(1), 97. https://doi.org/10.1186/s13054-017-1672-7
13. Google. 2020. Google Trends. Retrieved from: https://trends.google.com/trends/?geo=US.
14. Stokes EK, Zambrano LD, Anderson KN, et al. Coronavirus Disease 2019 Case Surveillance — United States, January 22–May 30, 2020. MMWR Morb Mortal Wkly Rep 2020;69:759–765. DOI: http://dx.doi.org/10.15585/mmwr.mm6924e2.
15. United States Census Bureau. 2019. Annual estimates of the resident population for the United States, regions, states, and Puerto Rico: April 2010 to July 1 2019. Retrieved from: https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html
16. Google. From the garage to the Googleplex. Retrieved from: https://about.google/our-story/.
17. Hosch, William. Google: American company. 2020. Retrieved from: https://www.britannica.com/topic/Google-Inc/additional-info#history.
18. Esposito S, Principi N. To mask or not to mask children to overcome COVID-19. Eur J Pediatr. 2020 Aug;179(8):1267-1270. doi: 10.1007/s00431-020-03674-9. Epub 2020 May 9. PMID: 32388722; PMCID: PMC7210459.
19. Swain ID. Why the mask? The effectiveness of face masks in preventing the spread of respiratory infections such as COVID-19 - a home testing protocol. J Med Eng Technol. 2020 Aug;44(6):334-337. doi: 10.1080/03091902.2020.1797198. Epub 2020 Jul 27. PMID: 32716230.
20. Li DTS, Samaranayake LP, Leung YY, Neelakantan P. Facial protection in the era of COVID-19: A narrative review. Oral Dis. 2020 Jun 7:10.1111/odi.13460. doi: 10.1111/odi.13460. Epub ahead of print. PMID: 32506757; PMCID: PMC7300840.
21. Google. Privacy Policy. 2020. Accessed from: https://policies.google.com/privacy?hl=en-US.
22. Internet Live Stats. Google search statistics. 2020. Retrieved from: https://www.internetlivestats.com/google-search-statistics/
23. Statistica. U.S. search engines ranked by number of cores searches 2008-2020. 2020. Retrieved from: https://www.statista.com/statistics/265796/us-search-engines-ranked-by-number-of-core-searches/
24. Centers for Disease Control and Prevention. 2020. Coronavirus Disease 2019 (COVID-19) Hospitalization. Retrieved from https://covidtracking.com/data/national/hospitalization
25. Centers for Disease Control and Prevention. 2020. A lifetime committed to eliminating the world's deadliest disease. https://www.cdc.gov/globalhivtb/who-we-are/features/elimintating_tb.html

Supplemental Data

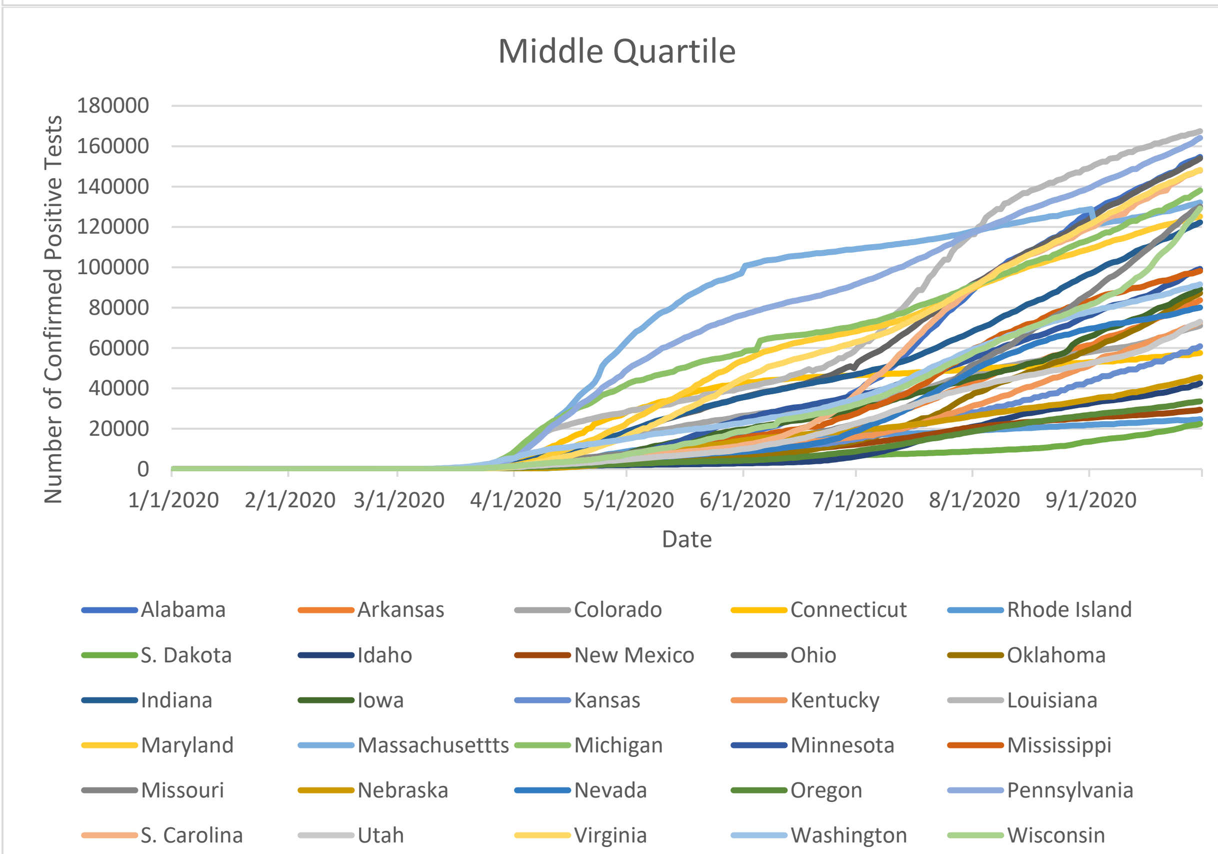
Supplemental Table 1. **Recovered Cases.** The first column lists the state name and the second column is the number that corresponds to the number of documented recovered cases according to John Hopkins University COVID-19 dashboard as of October 23, 2020[1]. States that have a result of zero did not have any recovered cases reported on the John Hopkins COVID-19 dashboard on this accession date. The total number of recovered cases is 3316297 for the United States. United States territories were not included in the study and are not represented in this table or the total recovered number.

Alabama	74439
Alaska	6812
Arizona	39089
Arkansas	92288
California	0
Colorado	7352
Connecticut	9800
Delaware	12410
Florida	0
Georgia	0
Hawaii	11188
Idaho	26916
Illinois	0
Indiana	4092
Iowa	85697
Kansas	2485
Kentucky	17627
Louisiana	165282
Maine	5269
Maryland	7999
Massachusetts	122856
Michigan	109539
Minnesota	113976
Mississippi	97675
Missouri	113976
Montana	16266
Nebraska	39905
Nevada	2346
New Hampshire	8692
New Jersey	36213
New Mexico	20332
New York	78753
North Carolina	218541
North Dakota	28271
Ohio	155181
Oklahoma	96245
Oregon	5870
Pennsylvania	148804
Rhode Island	2595
South Carolina	84761
South Dakota	26397
Tennessee	212555
Texas	744283
Utah	73586
Vermont	1718
Virginia	19321
Washington	0
West Virginia	16166
Wisconsin	145509
Wyoming	7220

Supplemental Figure 1. Positive Tests for SARS-CoV-2. State confirmed positive SARS-CoV-2 cases from January 1, 2020 to September 30, 2020. A. top ten states that had 180,000 cases or more. B. 30 states that had more than 20,000 and less than 180,000 cumulative positive cases C. Lower 10 states that had less than 20000 cases.[5]

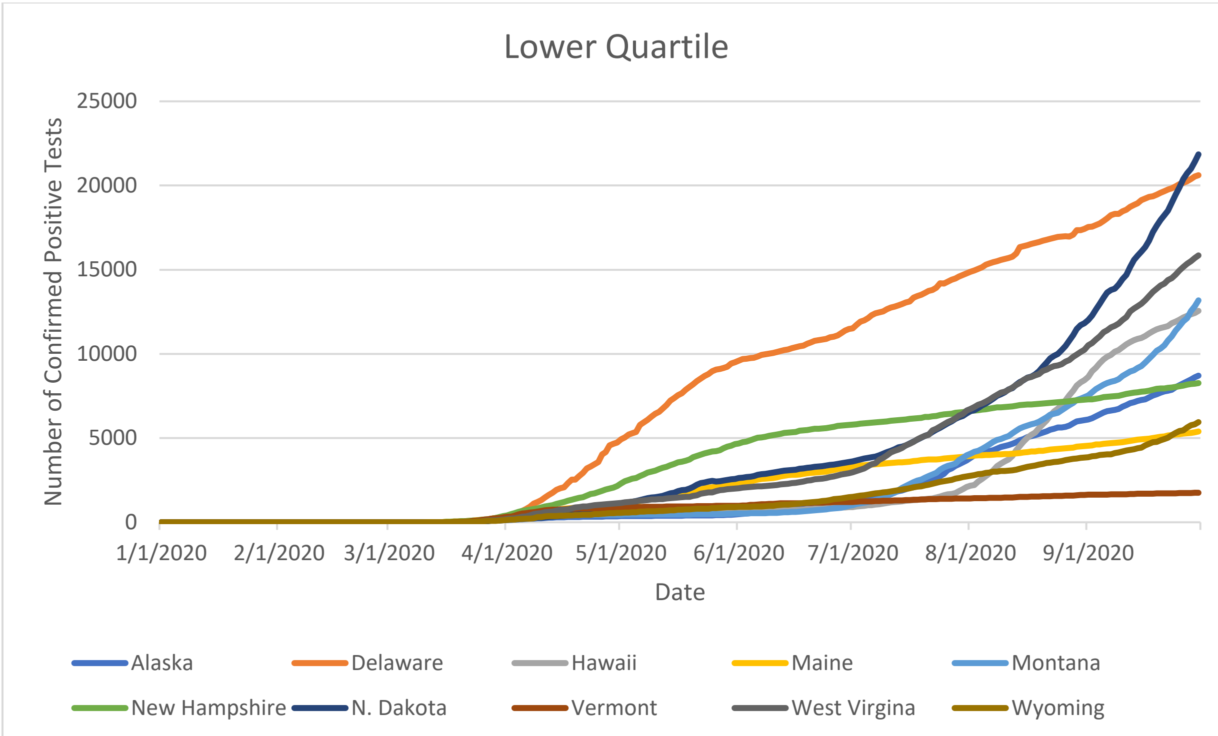


A.



B.

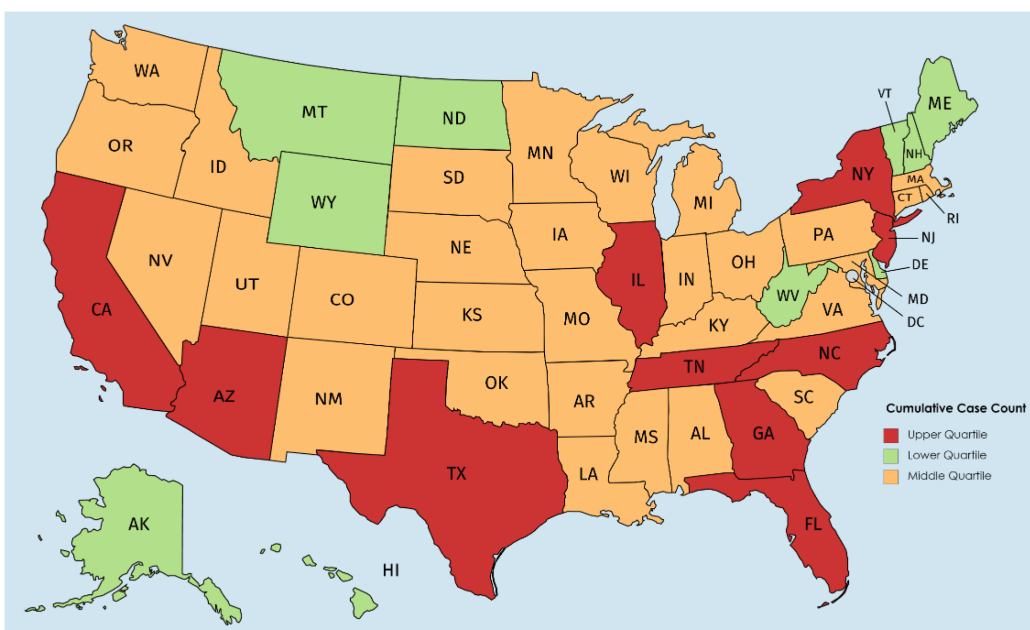
C.



Supplemental Figure 2. **Political Parties.** Political parties of states that achieved the lower quartile, middle quartile, and upper quartile rank in terms of cumulative confirmed positive case counts. The upper quartile had states with a cumulative case count over 190,000. The middle quartile included states with cumulative case counts between 20,000 and 189,999 thousand. The lower quartile included states with less than 20,000 cumulative positive cases.

	Republican	Democrat
Lower	6	4
Middle	15	15
Upper	5	5

Supplemental Figure 3. Number of Cumulative Cases by State. State status regarding cumulative number of positive cases. Red states had 190,000 cases or more. Orange states had between 20,000 and 189,999 and green states had less than 20,000. There is no significant correlation between geographic location and cumulative case count.



Supplemental Figure 4. Date of first confirmed case for each state.

Date of First Confirmed Positive Case	
Date	**State**
1/21/2020	Washington
1/24/2020	Illinois
1/25/2020	California
1/26/2020	Arizona
2/1/2020	Massachusetts
2/5/2020	Wisconsin
2/12/2020	Texas
2/17/2020	Nebraska
2/25/2020	Utah
2/28/2020	Oregon
3/1/2020	Florida
3/1/2020	New York
3/1/2020	Rhode Island
3/2/2020	Georgia
3/2/2020	New Hampshire
3/3/2020	North Carolina
3/4/2020	New Jersey
3/5/2020	Colorado
3/5/2020	Maryland
3/5/2020	Nevada
3/5/2020	Tennessee
3/6/2020	Hawaii
3/6/2020	Indiana
3/6/2020	Kentucky
3/6/2020	Minnesota
3/6/2020	Oklahoma
3/6/2020	Pennsylvania
3/6/2020	South Carolina
3/7/2020	Kansas
3/7/2020	Missouri
3/7/2020	Vermont
3/7/2020	Virginia
3/8/2020	Connecticut
3/8/2020	Iowa
3/9/2020	Louisiana
3/9/2020	Ohio
3/10/2020	Michigan
3/10/2020	South Dakota
3/11/2020	Arkansas
3/11/2020	Delaware
3/11/2020	Mississippi
3/11/2020	New Mexico
3/11/2020	North Dakota
3/11/2020	Wyoming
3/12/2020	Alaska
3/12/2020	Maine
3/13/2020	Alabama
3/13/2020	Idaho
3/13/2020	Montana
3/17/2020	West Virginia