



Article COVID-19 Related Early Google Search Behavior and Health Communication in the United States: Panel Data Analysis on Health Measures

Binhui Wang ¹, Beiting Liang ², Qiuyi Chen ³, Shu Wang ^{4,5}, Siyi Wang ⁶, Zhongguo Huang ⁶, Yi Long ⁷, Qili Wu ⁸, Shulin Xu ⁹, Pranay Jinna ¹⁰, Fan Yang ¹¹, Wai-Kit Ming ^{6,12} and Qian Liu ^{8,10,*}

- ¹ School of Management, Jinan University, Guangzhou 510632, China
- ² College of Economics, Jinan University, Guangzhou 510632, China
- ³ School of Journalism, Fudan University, Shanghai 200433, China
- ⁴ Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China
- ⁵ Laboratory of Biomass and Green Technologies, Gembloux Agro-Bio Tech, University of Liège, 5030 Gembloux, Belgium
- ⁶ Department of Public Health and Preventive Medicine, School of Medicine, Jinan University, Guangzhou 510632, China
- ⁷ Law School of Artificial Intelligence, Shanghai University of Political Science and Law, Shanghai 201701, China
- ⁸ School of Journalism and Communication, Jinan University National Media Experimental Teaching Demonstration Center, Jinan University, Guangzhou 510632, China
- ⁹ School of Economic, Guangzhou College of Commerce, Guangzhou 511363, China
- ¹⁰ School of Business, University at Albany, State University of New York, Albany, NY 12222, USA
- ¹¹ Communication Department, University at Albany, State University of New York, Albany, NY 12222, USA
- ¹² Department of Infectious Diseases and Public Health, Jockey Club College of Veterinary Medicine and Life Science, City University of Hong Kong, Hong Kong SAR, China
- Correspondence: tsusanliu@jnu.edu.cn

Abstract: The COVID-19 outbreak at the end of December 2019 spread rapidly all around the world. The objective of this study is to investigate and understand the relationship between public health measures and the development of the pandemic through Google search behaviors in the United States. Our collected data includes Google search queries related to COVID-19 from 1 January to 4 April 2020. After using unit root tests (ADF test and PP test) to examine the stationary and a Hausman test to choose a random effect model, a panel data analysis is conducted to investigate the key query terms with the newly added cases. In addition, a full sample regression and two sub-sample regressions are proposed to explain: (1) The changes in COVID-19 cases number are partly related to search variables related to treatments and medical resources, such as ventilators, hospitals, and masks, which correlate positively with the number of new cases. In contrast, regarding public health measures, social distancing, lockdown, stay-at-home, and self-isolation measures were negatively associated with the number of new cases in the US. (2) In mild states, which ranked one to twenty by the average daily new cases from least to most in 50 states, the query terms about public health measures (quarantine, lockdown, and self-isolation) have a significant negative correlation with the number of new cases. However, only the query terms about lockdown and self-isolation are also negatively associated with the number of new cases in serious states (states ranking 31 to 50). Furthermore, public health measures taken by the government during the COVID-19 outbreak are closely related to the situation of controlling the pandemic.

Keywords: COVID-19; Google trend; health communication; public health measures; panel data analysis



Citation: Wang, B.; Liang, B.; Chen, Q.; Wang, S.; Wang, S.; Huang, Z.; Long, Y.; Wu, Q.; Xu, S.; Jinna, P.; et al. COVID-19 Related Early Google Search Behavior and Health Communication in the United States: Panel Data Analysis on Health Measures. *Int. J. Environ. Res. Public Health* **2023**, *20*, 3007. https:// doi.org/10.3390/ijerph20043007

Academic Editors: Paul B. Tchounwou and Beatrice Casini

Received: 26 December 2022 Revised: 20 January 2023 Accepted: 4 February 2023 Published: 9 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/).

1. Introduction

1.1. Public Health Measures in the US

First detected in Wuhan, COVID-19 was highly virulent and spread to Iran, Europe, and finally, North America, contributing to an unprecedented global health crisis [1].

However, among these areas, the coronavirus raging through the US severely affected the country and gradually accounted for the most COVID-19 cases and deaths world-wide [2]. The US is facing a major public health crisis and is trying to address this issue through a series of medical reforms [3].

Since the pandemic, many governments introduced public health measures to control the development of the pandemic and the use of medical substances [4]. Quarantine, social distancing, lockdown, and staying at home were the pivotal role of old-style public health measures in the novel coronavirus (COVID-19) outbreak [5,6]. Quarantine was one of the oldest and most effective tools for controlling communicable disease outbreaks [6]. This public health practice was widely used in fourteenth-century Italy when ships arriving at the port of Venice from plague-infected ports had to anchor and wait for 40 days before disembarking their surviving passengers [7]. Social distancing was designed to reduce interactions between people in a broader community in which individuals may be infectious but have not yet been identified as such and therefore not yet isolated [6]. In South Korea and China social distancing has played an important role in mitigating the spread of COVID-19 [8,9]. In particular, to alert European countries that they should avoid close contact at the individual level and social meetings in each country [10]. The most restrictive nonpharmaceutical interventions (NPIs) for controlling the spread of COVID-19 were mandatory stay-at-home and business closures ("lockdown") [11]. They were used in some countries to prevent and control the pandemic [12–15].

1.2. Impact of the COVID-19 Pandemic

As a disease caused by a new coronavirus called SARS-CoV-2, COVID-19 affected more than 200 countries and caused respiratory tract infections in humans ranging from mild symptoms to lethal outcomes [16]. The extremely high infection rate and relatively high mortality raised public awareness of the emergence of COVID-19 and hence led to fears, worries, and anxiety among individuals globally. Dealing with COVID-19 had become one of the emergent global health challenges in managing infectious diseases [17,18]. More and more arduous efforts by health experts and authorities were put into mitigating the spread of the virus, and more individuals were beginning to focus on searching for relevant preventive information and taking preventive actions [19].

In less than half a year (in October 2020), 200 countries had reported over 10.3 million confirmed COVID-19 cases worldwide [20], and the cumulative number of cases reported globally was over 323 million; the number of deaths exceeded 5.5 million as of 16 January 2022 [21]. The arrival of the Omicron strain led to the adoption of changes in public safety measures in many areas, but the study of early prevention measures is still relevant. Public safety measures in coping with the coronavirus could significantly ease the spread of the disease and reduce the effects of disease during the pandemic [22,23].

The public searches for medical-related information and safety measures through the internet, where search engines can record the public attention on pandemic-related topics. As one of the largest search engines in the world, Google Trends can reflect public health information-seeking behaviors during the pandemic, especially regarding safety measures in local areas.

Because of the usefulness of public measures for the spread of the disease and the universality of Google search, the objective of in this paper is to investigate and understand the correlation between public health measures and public search behaviors with the development of the disease during the COVID-19 pandemic in the United States.

1.3. Aims

The primary aim of this study is to verify whether there is a correlation between Google search query keywords and the development of the COVID-19 pandemic in the United States (research question 1, RQ 1).

The secondary objective is to investigate whether there are significant correlations between the state's anti-pandemic public health (control) measures and the development of the pandemic (research question 2, RQ 2).

2. Materials and Methods

2.1. Keyword Search Tools

Google Trends (https://trends.google.com/trends/ (accessed on 10 April 2020) is a publicly available web-based tool that provides access to a largely unfiltered sample of actual search requests (queries) made to Google. Users could analyze interest in a specific topic from around the globe or drill down to geography at the city level by using its data which are anonymized, categorized, and grouped (support.google.com/trends).

Google search queries had been used to understand health communication during the COVID-19 outbreak [18,24–28]. As the COVID-19 outbreak continued, it was plausible that people might increasingly use internet search engines such as Google search (Google Inc., Mountain View, CA, USA) to obtain information about symptoms, diagnoses, or prevention [29]. Google released these Google COVID-19 search trends as open-access datasets for public health surveillance purposes as well as to accelerate valuable insights into the spread and impact of COVID-19 [30–32].

2.2. Obtaining and Analysis of the COVID-19 Pandemic Dataset

Google Trend, one of the most popular search engines worldwide, was used in our study to track down the updated internet hit search volumes in different states in the United States. Users of Google Trends could search through the link https://trends.google.com (accessed on 10 April 2020) to obtain the corresponding search volume (SV) of the terms [22] that revealed the general concern of the citizens in a certain area about one topic. We utilized this method in our study for analyzing public concerns related to certain health topics; however, the search volume turned out to be a normalized relative search volume (RSV) instead of an exact search count. The RSV value of relative search volume varies from 0 to 100, where the value of 0 referred to the least popular search term and the value of 100 meant the most popular search term [33,34]. Therefore, we compared and scaled all search information and normalized the data.

According to previous Google search research [35], symptoms, treatments and medical resources, measures, and the virus itself were some of the major concerns reported by online media platforms during the outbreak period of COVID-19. Therefore, in this research, we chose four main topics for search terms: "Diseases," "Treatments and medical resources," "Symptoms," and "Public health measures".

2.3. Public-Health-Measures-Related Keyword Search

Google search query variables were mainly divided into diseases, symptoms, treatments and medical resources, and public health measures. They were determined according to the pandemic-related search terms mentioned in the relevant literature or the relevant important pandemic research.

It was first called "unknown" virus pneumonia and then it was called "new coronavirus pneumonia"; COVID-19 was introduced by WHO [36], after that, COVID-19 was widely used in the study of this "unknown" virus pneumonia [37,38]. Facing COVID-19, there was presumably no pre-existing immunity in the population against the new coronavirus, and everyone in the population was assumed to be susceptible [37]. Among them, we had chosen pneumonia and COVID-19 to describe the disease. Its symptoms were usually fever, cough, sore throat, breathlessness, fatigue, and malaise among others [38]. Fever was the most critical signal feature of COVID-19 infection and was usually accompanied by fatigue and malaise; the cough usually contained a sore throat, and breathlessness was a serious and obvious symptom [39]. Therefore, we chose fever, cough, and shortness of breath to describe the COVID-19 symptoms.

After COVID-19 had swept the world, the biggest concern was whether the healthcare system could handle a large number of patients. The number of hospitals and doctors was closely related to the control of the pandemic [40]. Ventilators were essential as medical supplies for the treatment of critically ill patients [41]. Masks and vaccines were very effective medical resources for preventing the spread of COVID-19 and had always been widely considered [42,43]. Therefore, we added ventilator, hospital, doctor, vaccine, and mask as the main variables to reflect the treatments. However, many public health measures also played an important role in the control of the spread of COVID-19, such as quarantine, social distancing, lockdown, stay-at-home, etc. [5]. Among them, quarantine and lockdown were general measures implemented by the government, and social distancing and staying at home were voluntary actions of the people [6]. We chose these four search terms to comprehensively discuss public health measures.

2.4. Statistical Analysis

According to the above data, we obtained four groups of data related to COVID-19 from google search queries in the United States on COVID-19 from 1 January 2020, to 4 April 2020. We then used random effect analysis to investigate and analyze the relations between the key query terms with the newly added cases.

2.4.1. Panel Unit Root Test

Before the panel dynamic regression model was fitted, the stationary of each time series needed to be tested. If there was a unit root in the autoregressive part of the model, it indicated that this series was not stationary, that is, as time progressed, it did not return to a given value (long-term average), which made the regression a spurious regression. This article employed the Augmented Dickey–Fuller test (ADF test) [44] and the Phillips–Perron test (PP test) [45] to test the stationaries of variables.

For each time series x_t (such as positive increase etc.), we structured a p-order autoregressive model AR(p) as follows:

$$x_t = \phi_1 x_{t-1} + \cdots \phi_p x_{t-p} + \varepsilon_t,$$

where $t = 1, \dots, T$. Therefore, the ADF test set $\rho = \phi_1 + \dots + \phi_p - 1$ and the null and the alternative hypotheses were as follows:

$$H_0: \rho = 0 \{x_t \text{ was a non} - \text{ stationaryseries}\}, \\ H_1: \rho < 0 \{x_t \text{ was a stationaryseries}\}.$$

The ADF test statistic was:

$$\tau = \frac{\hat{\rho}}{S(\hat{\rho})},$$

where $S(\hat{\rho})$ was the sample standard deviation of ρ .

However, the ADF test method had a basic setting in that the noise of the time series had the same variance, so the ADF test method was not effective in the time series with heterogeneity. Phillips and Perron presented the PP test statistics, which can be applied to the stationary test for heterogeneity; it also obeyed the limit distribution of the ADF test statistic. The PP test statistic was:

$$Z(\tau) = \tau(\frac{\widehat{\sigma}^2}{\sigma_{Sl}^2}) - \frac{1}{2}(\sigma_{Sl}^2 - \widehat{\sigma}^2)T \sqrt{\sigma_{Sl}^2 \sum_{t=2}^T (x_{t-1} - \overline{x}_{T-1})^2},$$

where $\widehat{\sigma}^2 = T^{-1} \sum_{t=1}^T \widehat{\varepsilon}_t^2, \widehat{\sigma}_{Sl}^2 = T^{-1} \sum_{t=1}^l \widehat{\varepsilon}_t^2 + 2T^{-1} \sum_{j=1}^t w_j(l) \sum_{t=j+1}^T \widehat{\varepsilon}_t \widehat{\varepsilon}_{t-j}, \overline{x}_{T+1} = (T-1)^{-1} \sum_{t=1}^{T-1} x_t$. We performed a panel unit root test at a given confidence level; the values 0.001, 0.01, and 0.05 for the confidence level were the most frequently used [44,45].

2.4.2. Panel Dynamic Regression Model

In order to estimate a positive increase by the panel data analysis, it was common to explain the non-observed heterogeneity by using a fixed or random effect. The general form of the panel dynamic regression model was as follows [46]:

$$y_{it} = \alpha + \sum_{g=1}^{G} \beta_g x_{git} + u_{it} + \varepsilon_{it}, \ (1)$$

where y_{it} was the response (that was a positive increase); α was the intercept term (a constant); $\beta = (\beta_1, ..., \beta_G)$ were the parameters to be estimated; $x_{it} = (x_{1it,...}, x_{git})$ were the predictors; x_{git} was the *g*-th predictor for the individual *i* into the observation value at period *t*; and u_{it} was an individual heterogeneity, which was non-observed. i = 1, ..., I, t = 1, ..., T.

In the fixed effect model, the individual heterogeneity u_{it} was as a constant; so we can rewrite the Model (1) as:

$$y_{it} = \theta + \sum_{g=1}^{G} \beta_g x_{git} + \varepsilon_{it}, \ (2)$$

where $\theta = \alpha + u_{it}$.

On the contrary, the random effect model assumed u_{it} as a random variable and rewrote Model (1) as:

$$y_{it} = \alpha + \sum_{g=1}^{G} \beta_g x_{git} + \delta_{it},$$
(3)

where $\delta_{it} = u_{it} + \varepsilon_{it}$.

Before performing the panel data regression analysis, it was necessary to choose a fixed effect model (2) or a random effect model (3) to fit the model according to the specific actual data. If the model selection was wrong, it would cause a huge estimation error and fail to analyse [47].

2.4.3. Hausman Test

In order to choose a fixed effect model (2) or a random effect model (3), we conducted a Hausman test [46]. This is because the Hausman test would lose more degrees of freedom when estimating a fixed effect model, and the Hausman test tests whether a random effect model would be appropriate. The Hausman test set the null and the alternative hypotheses as follows:

 $H_0 : \operatorname{cov}(x_{it}, u_{it}) = 0 \{u_{it} \text{ was a random variable}\},$ $H_1 : \operatorname{cov}(x_{it}, u_{it}) \neq 0 \{u_{it} \text{ was a constant}\}.$

The Hausman test statistics were:

$$W = [\widehat{\beta}_{fe} - \widehat{\beta}_{re}]^T \widehat{\Omega}^{-1} [\widehat{\beta}_{fe} - \widehat{\beta}_{re}],$$

where β_{fe} and β_{re} were the estimated results of the regression parameters β in the fixed effect model (2) and random effect model (3), respectively, and $\Omega = Var(\beta_{fe} - \beta_{re})$.

We performed a Hausman test at a given confidence level, which commonly used 0.001, 0.01, and 0.05 [46].

3. Result

In total, we collected and normalized all states' Google search query data in the United States from 1 January 2020 to 4 April 2020 and saved it into CSV files. We also added CDC newly added cases data for coronavirus from CDC into the dataset [48]. Variables and sources and the basic descriptive statistics are listed in Table 1 below.

Source	Search Query Topics	Variable	Mean	Std. Dev.	Min	Max
CDC		Positive increase	64.280	464.245	0	10,841
	Diseases	Pneumonia	19.200	11.043	0.1	124
		COVID-19	114.700	212.172	0.1	1400
		Shortness of breath	4.382	5.517	0.1	74
	Symptoms	Fever	43.790	20.768	0.1	200
		Cough	33.340	15.057	0.1	200
C 1	Treatments and medical resources	Ventilator	7.628	13.781	0.1	180
Search query		Hospital	194.800	65.158	0.1	500
		Vaccine	31.210	22.777	0.1	216
		Doctor	86.020	27.849	0.1	230
		Mask	132.800	164.294	0.1	2500
		Quarantine	56.180	70.950	0.1	700
	Dalalia haalih maaaaaaaa	Social distance	2.143	4.786	0.1	48
	r ublic nealth measures	Lockdown	23.260	38.305	0.1	400
		Stay at home	24.660	88.260	0.1	1600

Table 1. Descriptive statistics of the main variables.

3.1. Full Sample Empirical Regression

In Model (1), *I* was the number of states in the United States, I = 50; *T* was the number of days from 1 January 2020 to 4 April 2020, T = 95; *G* was the number of predictors, G = 15; the unit root test results for each variable are shown in Table 2; and the Hausman test result of the full sample is shown in Table 3.

Table 2. Unit root test results for 15 variables.

	Variable	ADF Test	PP Test
	Positive increase	-13.080 ***	-517.240 ***
Diagona	Pneumonia	-12.160 ***	-5595.000 ***
Diseases	COVID-19	-14.940 ***	-351.900 ***
	Shortness of breath	-13.283 ***	-5307.900 ***
Symptoms	Fever	-14.181 ***	-3010.200 ***
	Cough	-12.253 ***	-5861.100 ***
	Ventilator	-15.715 ***	-1598.800 ***
	Hospital	-6.4026 ***	-2950.300 ***
Treatments and medical resources	Vaccine	-10.325 ***	-2662.900 ***
	Doctor	-7.7911 ***	-4502.700 ***
	Mask	-13.789 ***	-1667.200 ***
	Quarantine	-15.639 ***	-917.420 ***
	Social distance	-13.856 ***	-3297.300 ***
Public health measures	Lockdown	-14.085 ***	-1235.600 ***
	Stay at home	-15.909 ***	-2103.500 ***
	Self-isolation	-12.574 ***	-4659.600 ***

***: p < 0.001.

Table 3. Hausman test result of the full sample of Model (1).

Sample Type	Hausman Statistic	<i>p</i> -Value	
Full sample	4.799	0.994	

Table 2 presents the results of the two-unit root tests for each variable. If the sample sequence were non-stationary, it would need to be processed by a difference or a lag operator to make it a stationary time series. However, both the ADF test and PP test results for 15 variables rejected the null hypothesis at the 0.1% level; all 15 variables were stationary.

After the stationary test, the second step was to apply a Hausman test to assess whether the panel data Model (1) was a fixed or a random effect. The Hausman test result was shown in Table 3. The result showed that the *p*-value is 0.994, which meant the null hypothesis of the Hausman test was not rejected at the 0.1%, 1%, and 5% levels; we chose a random effect model (e.g., Model (3)) to analyze the full sample.

Based on the result of the Hausman test, the random effect model was selected for the full sample estimation in this study. The estimation results are shown in Table 4.

	Variable	Coef.	z-Value	Pr (> z)
	Constant	-315.677	-8.355	0.000 ***
Diseases	Pneumonia COVID-19	2.110 - 0.097	3.670 -1.522	0.000 ***
2	Shortness of breath	1.576	1.360	0.174
Symptoms	Fever Cough	1.972 0.543	5.029 1.197	0.000 *** 0.231
Treatments and medical resources	Ventilator Hospital Vaccine Doctor	8.424 0.770 -0.433 -0.080	$12.750 \\ 6.046 \\ -1.283 \\ -0.330$	0.000 *** 0.000 *** 0.199 0.742
	Mask	0.574	11.474	0.000 ***
Public health measures	Quarantine Social distance Lockdown Stay at home Self-isolation	$\begin{array}{c} 0.587 \\ -2.697 \\ -2.172 \\ -0.067 \\ -6.250 \end{array}$	$\begin{array}{r} 3.369 \\ -1.655 \\ -8.415 \\ -0.846 \\ -2.507 \end{array}$	0.000 *** 0.098 0.000 *** 0.398 0.012 *

Table 4. Random effect results of the full sample of Model (3).

***: p < 0.001; *: p < 0.05

To answer RQ1, we listed different Google search query keywords with their random effect results of the full sample of Model (3) in Table 4.

With concern about changing the name from coronavirus to COVID-19, only pneumonia (2.11) was found to have a positive coefficient with newly added cases for disease name-related search queries.

In terms of symptom-related queries, fever was found to have a significant positive coefficient (1.972), which indicated that the most significant symptom for predicting infection was high temperature. Others were not significant, such as shortness of breath (1.576).

For searches related to treatments and medical resources, positive correlations were found between infected cases and searches for a ventilator (8.424), hospital (0.77), and mask (0.574). This meant that more people searching for medical resources of these kinds could predict the increase in the number of positive cases. This was a warning sign to authorities that medical resources such as masks, hospitals, and especially ventilators could be generally insufficient.

With public-health-measures-related search terms, quarantine had a significant positive coefficient (0.587), while lockdown (-2.172) and self-isolation (-6.25) had significant negative coefficients. Public health measures towards passive individual control such as quarantine could positively predict newly added cases, which meant that when citizens searched for public controls at the passive individual level, infections increased rapidly. Compared to collective-level search terms, such as lockdown, the coefficient was negative, which meant that when people noticed a lockdown in an area, it was associated with a decrease in new cases. Search terms on social distancing and self-isolation were significantly associated with a drop in newly added COVID-19 cases. Relevant medical management departments needed to increase public concern about preventive measures at the individual active/initiative level and further carried out collective level control measures, while the effect of passive individual control measures may not be significant. The results showed that to answer RQ1, the increase or decrease in COVID-19 cases was partly related to these search variables in the United States on the Google platform. The query terms related to treatments and medical resources, ventilator, hospital, and mask were positive predictors of COVID-19 cases. This was probably because COVID-19 was the most harmful to the respiratory system and ventilators were very expensive and scarce. In terms of symptoms, people could find and compare symptoms online to see whether they have COVID-19. This allowed potential patients to be detected and confirmed as early as possible. They could thereby reduce travel, ensure safe social distancing and self-isolate, and seek medical treatment at hospitals. Public safety measures such as lockdown and self-isolation were negative predictors of the number of new cases, reflecting that these measures could have reduced the number of new cases.

3.2. Sub-Sample Empirical Regression

To answer RQ2, we discussed the impact of public health measures between states, this study collected data from 50 states in the United States, focusing on the effect of the significant variables (quarantine, lockdown, and self-isolation) of public health measures in Table 4 on the increase in COVID-19 cases.

This article defined the first day with a non-zero increase in new cases as the day when the outbreak period began and set the number of days in the outbreak period in each state as H_i ; the sum of daily positive increase of each state was L_i and the average daily positive increase during the outbreak period was K_i , where: $K_i = L_i/H_i$, i = 1, ..., 50. Additionally, this study used the average daily positive increase K_i to sort 50 states from least to most, as shown in Table 5.

Table 5. State rankings of the average daily positive increase (K_i).

Ranking	State	Ranking	State	Ranking	State
1	Alaska	18	Minnesota	35	Ohio
2	North Dakota	19	Arkansas	36	Colorado
3	Wyoming	20	Oregon	37	Indiana
4	South Dakota	21	Oklahoma	38	Connecticut
5	Montana	22	Idaho	39	Texas
6	Nebraska	23	Utah	40	Georgia
7	Hawaii	24	Mississippi	41	Washington
8	West Virginia	25	South Carolina	42	Illinois
9	Vermont	26	Arizona	43	Pennsylvania
10	Maine	27	Alabama	44	Florida
11	New Hampshire	28	Nevada	45	Massachusetts
12	New Mexico	29	North Carolina	46	California
13	Delaware	30	Wisconsin	47	Louisiana
14	Rhode Island	31	Missouri	48	Michigan
15	Kansas	32	Virginia	49	New Jersey
16	Kentucky	33	Maryland	50	New York
17	Iowa	34	Tennessee		

As shown in Table 5, we found that the average daily positive increase K_i of the state of Alaska was the lowest, while the average daily positive increase K_i of New York was the highest in the 50 states of the United States. To better discuss the 50 states of the United States, we divided them into five categories according to the state rankings and calculated the mean values of H_i (i.e., the number of days in the outbreak period) and K_i (i.e., the average daily positive increase during the outbreak period) in each category, as shown in Table 6.

	Mean of <i>H_i</i>	Mean of K _i
States ranking 1 to 10	24.6	11.648
States ranking 11 to 20	27.7	26.498
States ranking 21 to 30	26.1	63.907
States ranking 31 to 40	28.6	141.366
States ranking 41 to 50	29.8	793.319

Table 6. Mean of the number of days in the outbreak period (H_i) and the average daily positive increase (K_i).

According to Table 6, both the mean values of H_i (i.e., the number of days in the outbreak period) and K_i (i.e., the average daily positive increase during the outbreak period) presented a gradual increase in the five categories. Additionally, the mean values of H_i were between 24–30 days in each category but the mean values of K_i were quite different, with the lowest being 11.648 and the highest being 793.319. Then, we used Model (1) to perform a panel data regression on the top 20 (Alaska to Oregon) and the last 20 cities (Missouri to New York); the Hausman test results are shown in Table 7.

Table 7. Hausman test results of two sub-samples of Model (1).

Sample Type	Hausman Statistic	<i>p</i> -Value
States ranking 1 to 20	1.008	1.000
States ranking 31 to 50	9.569	0.846

Table 7 presents the Hausman tests for assessing whether the two sub-sample panel data models were fixed effect models or random effect models. The results showed that the *p*-value of states ranking 1 to 20 was 1.000, and the *p*-value of states ranking 31 to 50 was 0.846, which meant the null hypotheses of the two Hausman tests were not rejected at the 0.1%,1%, and 5% levels. Based on the results of the Hausman tests, the random effect model was selected for both sub-sample estimations in this study. The partial estimation results of the significant variables of public health measures in full sample regression were shown in Tables 8 and 9, respectively.

Table 8. Random effect partial results of states ranking 1 to 20.

	Variable	Coef.	z-Value	Pr (> z)
Public health measures	Quarantine Lockdown Self-isolation	$-0.024 \\ -0.035 \\ -0.185$	-3.514 -2.918 -2.272	0.000 *** 0.004 ** 0.023 *
*** < 0.001. ** <	0.01.*			

***: p < 0.001; **: p < 0.01; *: p < 0.05.

Table 9. Random effect partial results of states ranking 31 to 50.

	Variable	Coef.	z-Value	Pr (> z)
Public health measures	Quarantine Lockdown Self-isolation	0.244 -3.208 -20.735	0.577 -5.644 -2.383	0.564 0.000 *** 0.017 *

***: p < 0.001; *: p < 0.05.

Table 8 shows that in the states ranking 1 to 20, with public-health-measures-related search terms, quarantine (-0.024), lockdown (-0.035), and self-isolation (-0.185) have negative coefficients, and the significant variables were the same as the full sample. It was particularly important to note that search queries such as quarantine and self-isolation were individual passive and active behaviors in public health measures, while lockdown was an overall government-imposed or suggested public health measure, which meant that in mild states (states ranking 1 to 20), such as Alaska, North Dakota, Wyoming, South Dakota,

Montana and so on, all public health measures search queries (i.e., quarantine, lockdown, and self-isolation) were significantly associated with the reduction of COVID-19 situations.

However, as Table 9 shows, in the states ranking 31 to 50, the variables lockdown (-3.208) and self-isolation (-20.735) had significant negative coefficients but the quarantine was not significant at 0.1%, 1%, and 5% levels, which meant in serious states (states ranking 31 to 50), such as Washington, Illinois, Pennsylvania, Florida, Massachusetts and so on, only government-imposed or suggested quarantine and self-isolation search queries were significantly associated with the reduction of COVID-19 situations; especially the self-isolation search query had a large negative correlation to COVID-19 cases.

To answer RQ2, the searching behaviors suggested that the public health measures taken by the government in response to COVID-19 were closely related to the situation of the pandemic. In mild states (states ranking 1 to 20), all public health measures search queries were significantly associated with the reduction of COVID-19 situations. In serious states (states ranking 31 to 50), only quarantine and self-isolation search queries were significantly associated with the reduction of COVID-19 situations; especially the self-isolation search query had a large negative correlation to COVID-19 cases.

4. Discussion

4.1. Interpretation of the Results

Given the continuing pandemic, Google search keywords could reflect the trend of health information that the public seeks. Thus, we discussed the correlation between public health measures and public search behaviors with the development of the disease during the COVID-19 pandemic in the United States using panel data analysis. The four main query topics we chose—"Diseases," "Treatments and medical resources," "Symptoms," and "Public health measures"—represented people's major concern for the outbreak during search activities with 3–5 query variables each. We tested for the stationarity of data and found that all variables were stationary (tested by ADP test and PP test). Applying the Hausman test we also found that the random effect model is appropriate for our data.

The result showed that the search queries were related to the development of the disease. As we adopted all samples, the search query terms had an increase or a decrease in COVID-19 cases at different levels. Search query terms that had a positive correlation with the COVID-19 pandemic included ventilator, hospital, and mask, and pneumonia, fever, quarantine, lockdown, and self-isolation had a negative correlation with the COVID-19 pandemic. For further study on the impact of the state's anti-pandemic public health measures on the COVID-19 pandemic, we ranked the 50 states by the average daily positive increase from lowest to highest. The panel data analysis of the mild states (states ranking 1 to 20) showed that the significant variables were the same with the full sample (i.e., quarantine, lockdown, and self-isolation), but quarantine search query terms had a negative relationship with COVID-19 pandemic. However, quarantine search query terms had no relationship with the COVID-19 pandemic in the serious states (states ranking 31 to 50), the analysis of lockdown and self-isolation still had a negative correlation with the COVID-19 pandemic. Comparing the results of the full sample and two sub-samples, we found that lockdowns and self-isolation related to COVID-19 cases negatively and significantly. Quarantine had a positive coefficient in full sample regression but had a negative coefficient in mild states regression and was not significant in serious states regression. This meant that policies might increase health communication and discussions on the measures such as quarantine, lockdown, and self-isolation to confront COVID-19.

4.2. Future Perspectives

Appropriate public health measures can limit the spread of COVID-19. The symptoms of COVID-19 are similar to the flu (such as fever, cough, or sore throat). COVID-19 occurs during respiratory disease season in America and its symptoms such as fever or feeling feverish/chills, cough, and shortness of breath are similar to influenza. In addition, the incidence of the disease in each state is mainly related to the sequence of

the disease. States with a later onset of the disease can learn from the experience of states with an earlier onset of the disease and adopt better isolation and lockdown measures. Therefore, the government could take more and better measures to control the spread of the outbreak (e.g., all people aged 6 months or older should be vaccinated annually) [49–51]. Additionally, in a study based on the data of 1700 locations deployed during the pandemic across China, Korea, Italy, Iran France, and the United States, as well as national non-drug interventions, researchers show that the measures have prevented or delayed the 6.1 million confirmed cases and avoided about 495 million infectious cases [52]. Proper compulsory public health measures can avoid infection to a certain degree [53,54]. The government should suggest effective public measures according to its national conditions to cope with coronavirus as a disaster that challenges mankind.

4.3. Limitations

This study has certain limitations. First, we only used the Google search engine, and the scope of the query, therefore, is limited to the United States and English language users; therefore, can't fully reflect people's choices in other countries or regions of the world. In addition, the Google search users' demographics in our study such as age and gender are not available, and hence we cannot do further analysis. Moreover, the effect of these prevention and control measures is related to the original strain and Delta strain in the earlier period. If it is an Omicron strain, the effect of the prevention and control measures is almost ineffective. Future studies could be improved in these aspects.

5. Conclusions

In this study of the US Google search query sample, public search behaviors and the development of the COVID-19 pandemic are analyzed and found to be related. Public-healthmeasure searches, including quarantines, lockdowns, and self-isolation, are negatively correlated to outbreaks with varying degrees across states. These results underscore the need for governments to take more and better health measures to mitigate the development of COVID-19.

Author Contributions: B.W. and Q.L. conceived the original idea and designed the whole research process; Q.L., B.L., Q.C., Y.L. and S.W. (Shu Wang) collected and cleaned the data; B.W., B.L., Q.C., S.W. (Siyi Wang), Z.H. and Q.L. conducted the data analysis and data interpretation and wrote the draft of the paper with input from all co-authors; B.L., Q.C., Q.W. and S.X. made the figures; Q.L., F.Y., P.J. and W.-K.M. contributed to the administration of the project, data analysis, and data interpretation; P.J. and F.Y. reviewed the manuscript. All authors contributed to the interpretation of the results and the final manuscript. All authors discussed and agreed on the implications of the study findings and approved the final version to be published. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable, this study doesn't involve humans or animals.

Data Availability Statement: The data used in this study is available upon request from the corresponding author.

Acknowledgments: We are grateful for the help in the data-collecting process from Bence Hoston and Leon Ferri, and we thank Barry M. Loneck for valuable suggestions and discussions on this paper.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

 Fong, Z.V.; Qadan, M.; McKinney, R., Jr.; Griggs, C.L.; Shah, P.C.; Buyske, J.; Sachdeva, A.K.; Callery, M.P.; Altieri, M.S. Practical Implications of Novel Coronavirus COVID-19 on Hospital Operations, Board Certification, and Medical Education in Surgery in the USA. J. Gastrointest. Surg. 2020, 24, 1232–1236. [CrossRef] [PubMed]

- Post, L.A.; Issa, T.Z.; Boctor, M.J.; Moss, C.B.; Murphy, R.L.; Ison, M.G.; Achenbach, C.J.; Resnick, D.; Singh, L.N.; White, J.; et al. Dynamic Public Health Surveillance to Track and Mitigate the US COVID-19 Epidemic: Longitudinal Trend Analysis Study. J. Med. Internet Res. 2020, 22, e24286. [CrossRef] [PubMed]
- 3. Feng, Z.; Xiao, C.; Li, P.; You, Z.; Yin, X.; Zheng, F. Comparison of spatio-temporal transmission characteristics of COVID-19 and its mitigation strategies in China and the US. *J. Geogr. Sci.* **2020**, *30*, 1963–1984. [CrossRef]
- 4. Wu, H.-L.; Huang, J.; Zhang, C.J.P.; He, Z.; Ming, W.-K. Facemask shortage and the novel coronavirus disease (COVID-19) outbreak: Reflections on public health measures. *EClinicalMedicine* **2020**, *21*, 100329. [CrossRef]
- 5. Cetron, M.; Landwirth, J. Public health and ethical considerations in planning for quarantine. Yale J. Biol. Med. 2005, 78, 329–334.
- 6. Wilder-Smith, A.; Freedman, D.O. Isolation, quarantine, social distancing and community containment: Pivotal role for old-style public health measures in the novel coronavirus (2019-nCoV) outbreak. *J. Travel Med.* **2020**, *27*, taaa020. [CrossRef] [PubMed]
- Cetron, M.; Simone, P. Battling 21st-century scourges with a 14th-century toolbox. *Emerg. Infect. Dis.* 2004, 10, 2053–2054. [CrossRef]
- 8. Park, S.W.; Sun, K.; Viboud, C.; Grenfell, B.T.; Dushoff, J. Potential roles of social distancing in mitigating the spread of coronavirus disease 2019 (COVID-19) in South Korea. *medRxiv* 2020. [CrossRef]
- 9. Kupferschmidt, K.; Cohen, J. Can China's COVID-19 strategy work elsewhere? Science 2020, 367, 1061–1062. [CrossRef]
- 10. Huynh, T.L.D. Does culture matter social distancing under the COVID-19 pandemic? Saf. Sci. 2020, 130, 104872. [CrossRef]
- 11. Varalakshmi, R.; Swetha, R. COVID-19 lock down: People psychology due to law enforcement. *Asian J. Psychiatry* **2020**, *51*, 102102. [CrossRef]
- 12. Aleta, A.; Martín-Corral, D.; Pastore, Y.P.A.; Ajelli, M.; Litvinova, M.; Chinazzi, M.; Dean, N.E.; Halloran, M.E.; Longini, I.M., Jr.; Merler, S.; et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. *Nat. Hum. Behav.* **2020**, *4*, 964–971. [CrossRef]
- 13. Wells, C.R.; Townsend, J.P.; Pandey, A.; Moghadas, S.M.; Krieger, G.; Singer, B.; McDonald, R.H.; Fitzpatrick, M.C.; Galvani, A.P. Optimal COVID-19 quarantine and testing strategies. *Nat. Commun.* **2021**, *12*, 356. [CrossRef]
- 14. Ahmed, M.Z.; Ahmed, O.; Aibao, Z.; Hanbin, S.; Siyu, L.; Ahmad, A. Epidemic of COVID-19 in China and associated Psychological Problems. *Asian J. Psychiatr.* **2020**, *51*, 102092. [CrossRef]
- 15. Wei, Y.; Ye, Z.; Cui, M.; Wei, X. COVID-19 prevention and control in China: Grid governance. J. Public Health 2021, 43, 76–81. [CrossRef]
- 16. Li, Y.D.; Chi, W.Y.; Su, J.H.; Ferrall, L.; Hung, C.F.; Wu, T.C. Coronavirus vaccine development: From SARS and MERS to COVID-19. *J. Biomed. Sci.* **2020**, *27*, 104. [CrossRef]
- 17. Ahorsu, D.K.; Lin, C.Y.; Imani, V.; Saffari, M.; Griffiths, M.D.; Pakpour, A.H. The Fear of COVID-19 Scale: Development and Initial Validation. *Int. J. Ment. Health Addict.* 2020, 20, 1537–1545. [CrossRef]
- Zitting, K.M.; Lammers-van der Holst, H.M.; Yuan, R.K.; Wang, W.; Quan, S.F.; Duffy, J.F. Google Trends reveals increases in internet searches for insomnia during the 2019 coronavirus disease (COVID-19) global pandemic. *J. Clin. Sleep Med.* 2021, 17, 177–184. [CrossRef]
- 19. Ahorsu, D.K.; Lin, C.Y.; Pakpour, A.H. The Association Between Health Status and Insomnia, Mental Health, and Preventive Behaviors: The Mediating Role of Fear of COVID-19. *Gerontol. Geriatr. Med.* **2020**, *6*, 2333721420966081. [CrossRef]
- 20. WHO. Coronavirus Disease (COVID-19) Situation Report—163. Available online: https://www.who.int/docs/default-source/ coronaviruse/situation-reports/20200701-covid-19-sitrep-163.pdf?sfvrsn=c202f05b_2 (accessed on 1 July 2020).
- WHO. Weekly Epidemiological Update on COVID-19—18 January 2022. Available online: https://www.who.int/publications/ m/item/weekly-epidemiological-update-on-covid-19---18-january-2022 (accessed on 18 January 2022).
- Carneiro, H.A.; Mylonakis, E. Google trends: A web-based tool for real-time surveillance of disease outbreaks. *Clin. Infect. Dis.* 2009, 49, 1557–1564. [CrossRef]
- Teng, Y.; Bi, D.; Xie, G.; Jin, Y.; Huang, Y.; Lin, B.; An, X.; Feng, D.; Tong, Y. Dynamic Forecasting of Zika Epidemics Using Google Trends. *PLoS ONE* 2017, 12, e0165085. [CrossRef] [PubMed]
- Ming, W.K.; Huang, F.; Chen, Q.; Liang, B.; Jiao, A.; Liu, T.; Wu, H.; Akinwunmi, B.; Li, J.; Liu, G.; et al. Understanding Health Communication Through Google Trends and News Coverage for COVID-19: Multinational Study in Eight Countries. *JMIR Public Health Surveill.* 2021, 7, e26644. [CrossRef] [PubMed]
- McCarthy, A.D.; McGoldrick, D. Analyzing Public Interest in Metabolic Health-Related Search Terms During COVID-19 Using Google Trends. *Cureus* 2021, 13, e15715. [CrossRef] [PubMed]
- Alicino, C.; Bragazzi, N.L.; Faccio, V.; Amicizia, D.; Panatto, D.; Gasparini, R.; Icardi, G.; Orsi, A. Assessing Ebola-related web search behaviour: Insights and implications from an analytical study of Google Trends-based query volumes. *Infect. Dis. Poverty* 2015, 4, 54. [CrossRef] [PubMed]
- Rovetta, A.; Bhagavathula, A.S. Global Infodemiology of COVID-19: Analysis of Google Web Searches and Instagram Hashtags. J. Med. Internet Res. 2020, 22, e20673. [CrossRef]
- Sousa-Pinto, B.; Anto, A.; Czarlewski, W.; Anto, J.M.; Fonseca, J.A.; Bousquet, J. Assessment of the Impact of Media Coverage on COVID-19-Related Google Trends Data: Infodemiology Study. J. Med. Internet Res. 2020, 22, e19611. [CrossRef]
- 29. Pier, M.M.; Pasick, L.J.; Benito, D.A.; Alnouri, G.; Sataloff, R.T. Otolaryngology-related Google Search trends during the COVID-19 pandemic. *Am. J. Otolaryngol.* 2020, *41*, 102615. [CrossRef]

- 30. Alruily, M.; Ezz, M.; Mostafa, A.M.; Yanes, N.; Abbas, M.; El-Manzalawy, Y. Prediction of COVID-19 transmission in the united states using google search trends. *Comput. Mater. Contin.* **2022**, *71*, 1751–1768.
- Li, C.; Chen, L.J.; Chen, X.; Zhang, M.; Pang, C.P.; Chen, H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. Euro Surveill. 2020, 25, 2000199. [CrossRef]
- 32. Rovetta, A.; Bhagavathula, A.S. COVID-19-Related Web Search Behaviors and Infodemic Attitudes in Italy: Infodemiological Study. *JMIR Public Health Surveill.* **2020**, *6*, e19374. [CrossRef]
- 33. Mavragani, A.; Ochoa, G. Google Trends in Infodemiology and Infoveillance: Methodology Framework. *JMIR Public Health Surveill.* **2019**, *5*, e13439. [CrossRef]
- Ayyoubzadeh, S.M.; Ayyoubzadeh, S.M.; Zahedi, H.; Ahmadi, M.; Niakan Kalhori, S.R. Predicting COVID-19 Incidence Through Analysis of Google Trends Data in Iran: Data Mining and Deep Learning Pilot Study. *JMIR Public Health Surveill.* 2020, 6, e18828. [CrossRef]
- 35. Yu, W.; Liu, X.; Zhao, F.; Li, M.; Zhang, L. Control of unreasonable growth of medical expenses in public hospitals in Shanghai, China: A multi-agent system model. *BMC Health Serv. Res.* **2020**, *20*, 490. [CrossRef]
- 36. Dong, H.; Zhou, M.; Che, D.; Zhang, H.; Bodomo, A. Analysis of COVID-19 Name Varieties in Chinese Mass and Social Media. *Int. J. Environ. Res. Public Health* **2021**, *18*, 9850. [CrossRef]
- 37. Eurosurveillance Editorial Team. Updated rapid risk assessment from ECDC on the novel coronavirus disease 2019 (COVID-19) pandemic: Increased transmission in the EU/EEA and the UK. *Euro Surveill.* **2020**, *25*, 2003121. [CrossRef]
- 38. Singhal, T. A Review of Coronavirus Disease-2019 (COVID-19). Indian J. Pediatr. 2020, 87, 281–286. [CrossRef]
- 39. Li, L.-Q.; Huang, T.; Wang, Y.-Q.; Wang, Z.-P.; Liang, Y.; Huang, T.-B.; Zhang, H.-Y.; Sun, W.; Wang, Y. COVID-19 patients' clinical characteristics, discharge rate, and fatality rate of meta-analysis. *J. Med. Virol.* **2020**, *92*, 577–583. [CrossRef]
- 40. Xie, L.; Yang, H.; Zheng, X.; Wu, Y.; Lin, X.; Shen, Z. Medical resources and coronavirus disease (COVID-19) mortality rate: Evidence and implications from Hubei province in China. *PLoS ONE* **2021**, *16*, e0244867. [CrossRef]
- 41. Auld, S.; Caridi-Scheible, M.; Blum, J.M.; Robichaux, C.J.; Kraft, C.S.; Jacob, J.T.; Jabaley, C.S.; Carpenter, D.; Kaplow, R.; Hernandez, A.C.; et al. ICU and ventilator mortality among critically ill adults with COVID-19. *medRxiv* 2020. [CrossRef]
- Tirupathi, R.; Bharathidasan, K.; Palabindala, V.; Salim, S.A.; Al-Tawfiq, J.A. Comprehensive review of mask utility and challenges during the COVID-19 pandemic. *Infez. Med.* 2020, 28, 57–63.
- 43. Soleimanpour, S.; Yaghoubi, A. COVID-19 vaccine: Where are we now and where should we go? *Expert Rev. Vaccines* **2021**, 20, 23–44. [CrossRef] [PubMed]
- 44. Said, S.E.; Dickey, D.A. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* **1984**, 71, 599–607. [CrossRef]
- Perron, P. Trends and random walks in macroeconomic time series: Further evidence from a new approach. *J. Econ. Dyn. Control.* 1988, 12, 297–332. [CrossRef]
- 46. Frondel, M.; Vance, C. Fixed, random, or something in between? A variant of Hausman's specification test for panel data estimators. *Econ. Lett.* **2010**, *107*, 327–329. [CrossRef]
- 47. Jiang, J.; Luo, L. Influence of population mobility on the novel coronavirus disease (COVID-19) epidemic: Based on panel data from Hubei, China. *Glob. Health Res. Policy* **2020**, *5*, 30. [CrossRef]
- Centers for Disease Control and Prevention. Trends in Number of COVID-19 Cases and Deaths in the US Reported to CDC, by State/Territory. Available online: https://covid.cdc.gov/covid-data-tracker/#trends_dailycases (accessed on 17 April 2020).
- Patel, A.; Jernigan, D.B. Initial Public Health Response and Interim Clinical Guidance for the 2019 Novel Coronavirus Outbreak— United States, December 31, 2019-February 4, 2020. MMWR Morb. Mortal. Wkly. Rep. 2020, 69, 140–146. [CrossRef]
- 50. Jernigan, D.B. Update: Public Health Response to the Coronavirus Disease 2019 Outbreak—United States, 24 February 2020. MMWR Morb. Mortal. Wkly. Rep. 2020, 69, 216–219. [CrossRef]
- 51. Xu, S.; Zhu, Q.; Yang, Z. Influencing factors of environmental efficiency of strategic emerging industries and their power cooperation mechanism design. *Environ. Sci. Pollut. Res. Int.* 2022, 1–26. [CrossRef]
- Hsiang, S.; Allen, D.; Annan-Phan, S.; Bell, K.; Bolliger, I.; Chong, T.; Druckenmiller, H.; Huang, L.Y.; Hultgren, A.; Krasovich, E.; et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature* 2020, 584, 262–267. [CrossRef]
- 53. Su, D.; Xu, S.; Tong, Z. Credit availability and corporate risk-taking: Evidence from China's green credit policy. *Post-Communist Econ.* **2023**, 1–35. [CrossRef]
- 54. Jebran, K.; Yang, Z.; Chen, S.; Ali, S.T. Does the famine experience of board chair hamper innovation? J. Int. Financ. Manag. Account. 2023, 1–51. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.