



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.



Global internet search trends related to gastrointestinal symptoms predict regional COVID-19 outbreaks



Shuai Ben^{a,b,1}, Junyi Xin^{a,b,1}, Silu Chen^{a,b,1}, Yan Jiang^c, Qianyu Yuan^d, Li Su^d, David C. Christiani^{d,e}, Zhengdong Zhang^{a,b,*}, Mulong Du^{b,f,*}, Meilin Wang^{a,b,g,*}

^a Department of Genetic Toxicology, The Key Laboratory of Modern Toxicology of Ministry of Education, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China

^b Department of Environmental Genomics, Jiangsu Key Laboratory of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Personalized Medicine, Nanjing Medical University, Nanjing, China

^c Department of Epidemiology, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China

^d Departments of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA 02115, United States of America

^e Department of Medicine, Massachusetts General Hospital, Boston, MA 02114, United States of America

^f Department of Biostatistics, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China

^g Suzhou Municipal Hospital, Gusu School, The Affiliated Suzhou Hospital of Nanjing Medical University, Nanjing Medical University, Suzhou, China

ARTICLE INFO

Article history:

Accepted 7 November 2021

Available online 9 November 2021

Keywords:

Google trends

COVID-19

Gastrointestinal symptoms

Time-series analysis

SUMMARY

Background: Real-time surveillance of search behavior on the internet has achieved accessibility in measuring disease activity. In this study, we systematically assessed the associations between internet search trends of gastrointestinal (GI) symptom terms and daily newly confirmed COVID-19 cases at both global and regional levels.

Methods: Relative search volumes (RSVs) of GI symptom terms were derived from internet search engines. Time-series analyses with autoregressive integrated moving average models were conducted to fit and forecast the RSV trends of each GI symptom term before and after the COVID-19 outbreak. Generalized additive models were used to quantify the effects of RSVs of GI symptom terms on the incidence of COVID-19. In addition, dose-response analyses were applied to estimate the shape of the associations.

Results: The RSVs of GI symptom terms could be characterized by seasonal variation and a high correlation with symptoms of “fever” and “cough” at both global and regional levels; in particular, “diarrhea” and “loss of taste” were abnormally increased during the outbreak period of COVID-19, with elevated point changes of 1.31 and 8 times, respectively. In addition, these symptom terms could effectively predict a COVID-19 outbreak in advance, underlying the lag correlation at 12 and 5 days, respectively, and with mutual independence. The dose-response curves showed a consistent increase in daily COVID-19 risk with increasing search volumes of “diarrhea” and “loss of taste”.

Conclusion: This is the first and largest epidemiologic study that comprehensively revealed the advanced prediction of COVID-19 outbreaks at both global and regional levels via GI symptom indicators.

© 2021 The British Infection Association. Published by Elsevier Ltd. All rights reserved.

Introduction

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has emerged as a public health crisis worldwide. In addition to the most common

symptoms such as cough and fever,¹ symptoms of the gastrointestinal (GI) system, including nausea, vomiting, abdominal pain, diarrhea, anorexia, and loss of taste, are also of high incidence with widespread concern.^{2–5}

The internet is an important channel for disseminating health information and has been proven useful in assessing various aspects of human behavior.⁶ One previous study successfully detected influenza epidemics by monitoring health-seeking behavior in the form of queries to online search engines.⁷ Google Trends, a web-based surveillance tool for gathering search behavior on the internet, has been widely used to evaluate disease prevalence and predict infectious disease pandemics.^{8–10} The records of search terms on internet search engines from different countries and re-

* Corresponding authors at: Department of Environmental Genomics, Jiangsu Key Laboratory of Cancer Biomarkers, Prevention and Treatment, Collaborative Innovation Center for Cancer Personalized Medicine, Nanjing Medical University, 101 Longmian Avenue, Jiangning District, Nanjing 211166, China.

E-mail addresses: drzdzhang@njmu.edu.cn (Z. Zhang), drdumulong@njmu.edu.cn (M. Du), mwang@njmu.edu.cn (M. Wang).

¹ These authors contributed equally to this work.

gions may have predictive effects on the local outbreak and spread of COVID-19. However, whether the search interests of GI symptom terms on the internet could be associated with outbreaks of COVID-19 is still limited.¹¹

In this epidemiologic study, we systematically analysed and forecasted the relative search volume (RSV) trends of GI symptom terms on internet search engines before and after the outbreak of COVID-19 over time and across geographical areas. We also evaluated the associations between RSVs of GI symptom terms and daily newly confirmed COVID-19 cases at both global and regional levels. These analyses would allow us to understand the data-driven association and prediction of GI symptoms during the COVID-19 outbreak.

Methods

Data retrieval and collection

In this study, we chose four representative countries as research objects (i.e., the United States, the United Kingdom, Australia, and China) based on the population, geographic location, and internet popularity rate factors. Google is the most widely used web search engine worldwide, while Baidu is the most population search engine in China. We applied Google Trends and Baidu Index, which documented the volumes of search terms at selected regions and time periods, to scientifically review the RSV trends of GI symptom terms, including “nausea”, “vomiting”, “abdominal pain”, “diarrhea”, “anorexia”, and “loss of taste”, as well as the two most common signs of “cough” and “fever”, which were defined as positive-control COVID-19 symptom terms in this study. The RSVs for these terms of disease symptoms may be a quantitative index to reflect group symptoms associated with COVID-19. A Google Trends value represents the search interest assigned to the highest point ranging from 0 to 100. A value of 100 is the peak popularity of the term, while a value of 50 means that the term is half as popular. A Baidu Index value represents the weight of the search frequency of a search term. When analysing and comparing the correlations among RSVs of different search terms as well as the associations between RSVs of search terms and COVID-19 outbreaks across countries, the Baidu Index value was scaled on a range of 0 to 100 the same as the Google Trends value for further combined analysis.

The daily newly confirmed COVID-19 cases at the global and regional levels spanning from January 1, 2020 to December 1, 2020 were obtained from the World Health Organization (WHO) COVID-19 Dashboard (<https://covid19.who.int/>). We subjectively divided the period above into a COVID-19 outbreak period (January 1, 2020 to May 31, 2020) and a pandemic period (June 1, 2020 to December 1, 2020).

Time-series analyses

Prediction of the RSV trends of GI symptom terms

Autoregressive integrated moving average (ARIMA) models were used to assess the seasonal variations and model the RSV trends of candidate disease symptom terms worldwide before the COVID-19 outbreak and subsequently to predict their trends as well as calculate their abnormal changes caused by COVID-19. Model fitting and selection are described in the **Supplementary Methods**.

Weekly RSVs for candidate terms derived from Google Trends and daily RSVs for candidate terms derived from the Baidu Index from December 1, 2015 to November 30, 2019 were used to construct ARIMA models, which were then applied to forecast the RSVs of each symptom term during the COVID-19 epidemic by a standard time-series approach in separate analyses. The point change of each symptom was calculated as the fold change (FC) by

comparing the actual value with the predicted value at the same time point.

Associations between RSVs of GI symptoms and COVID-19 incidence

The associations between RSVs of each symptom with daily COVID-19 incidence were assessed in separate analyses using a standard time-series approach. The negative-binomial generalized additive model (GAM) was used to evaluate the effects of RSVs of each symptom term on COVID-19 risk. We used the search volumes for the terms of each symptom (i.e., RSVs) as the exposure of interest and daily confirmed COVID-19 cases as the outcome. To control for potential confounding effects, the following covariates were included in the model: (i) a natural cubic smooth function of calendar time with 7 degrees of freedom (df) to remove long-term and seasonal trends; and (ii) an indicator variable for “day of the week (DOW)” to account for short-term weekly variations.

Herein, we used a single-symptom model to explore the effects of RSVs for each GI symptom term on the COVID-19 outbreak and pandemic with different lag days (lag 0 to 28 days in 4 weeks). In addition, to estimate the overall shape of the associations between RSVs of GI symptom terms and COVID-19 incidence at the global or regional level, we plotted the dose-response curves by using a 3df for the smoothing function in a single-symptom model. The effects of RSVs for disease symptoms on daily newly confirmed COVID-19 cases are shown as the relative risks (RRs) and 95% confidence intervals (CIs).

Statistical analysis

Spearman’s correlation analysis was applied to investigate the associations between RSVs of different symptom terms. Three- to eight-knot restricted cubic spline (RCS) models were used to empirically assess the shape of the curve representing newly confirmed COVID-19 cases over time. All statistical analyses were conducted with R software (version 4.0.0). All tests were two-sided, and $P < 0.05$ was considered as statistical significance.

Results

Time-series characteristics of GI symptom terms during COVID-19 periods

As shown in **Fig. 1**, spanning from December 1, 2015 to November 30, 2019, the RSVs for the terms “fever”, “cough”, “diarrhea”, and “vomiting” visually presented persistent seasonal variations, and the peak search occurred during the winter season. Subsequently, four symptom terms (i.e., “fever”, “cough”, “diarrhea”, and “loss of taste”) were sharply and abnormally elevated after December 1, 2019, which might be attributed to the COVID-19 outbreak. These characteristics were also observed across four representative countries of the United States, the United Kingdom, Australia, and China (**Supplementary Fig. 1**).

The ARIMA model was then applied to quantify and forecast the abnormally increased RSVs of different symptom terms spanning the COVID-19 outbreak and pandemic periods. As positive-control symptoms, the RSVs of “fever” and “cough” were 2.31 (95% CI = 2.07–2.62) and 2.60 (95% CI = 2.33–2.95) times higher than the predictive values on March 15, 2020, respectively (**Table 1, Fig. 2**). Similar findings were observed in three GI symptoms in which the RSVs of “diarrhea” (FC = 1.31, 95% CI = 1.22–1.42) and “loss of taste” (FC = 8) were also significantly higher than their predicted values around March 15, 2020 (**Table 1, Fig. 2**), even though “loss of taste” lacked time trends. In contrast, the peak of “abdominal pain” was beyond the COVID-19 outbreak, and “vomiting”, “nausea”, and “anorexia” did not show significant increases (**Table 1**).

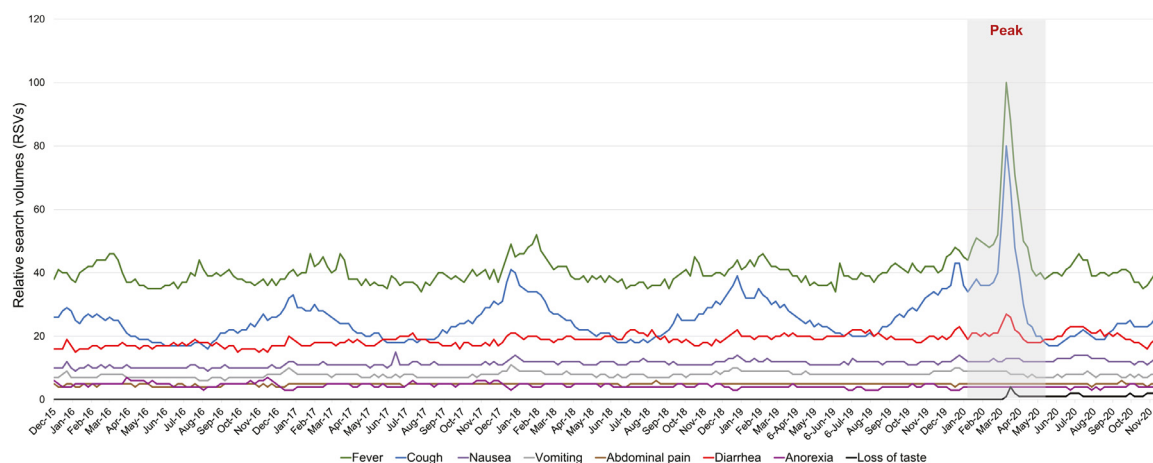


Fig. 1. The RSV trends of symptom terms on Google Trends from December 1, 2015 to December 1, 2020 at the global level. The x-axis represents the time period from December 1, 2015 to December 1, 2020, while the y-axis represents the RSVs of various symptom terms. Google Trends provides weekly search data for the time intervals. Each disease symptom is coloured differently. Abbreviations: RSV, relative search volume.

Table 1

Parameters and predicted results of the ARIMA model for the RSV trends of symptom terms of interest worldwide.

Symptom terms	MAPE	Date ^a	Predicted RSVs (95% CI) ^b	Actual RSVs	Point change (FC, 95% CI) ^c	ARIMA model
Fever	2.58	Mar 15, 2020	43.20 (38.16–48.24)	100	2.31 (2.07–2.62)	ARIMA (2, 0, 0)
Cough	2.42	Mar 15, 2020	30.75 (27.14–34.37)	80	2.60 (2.33–2.95)	ARIMA (1, 1, 2)
Nausea	2.87	Dec 29, 2019	13.90 (12.58–15.23)	14	1.01 (0.92–1.11)	ARIMA (2, 1, 2)
Vomiting	2.90	Dec 29, 2019	10.04 (9.12–10.96)	10	1.00 (0.91–1.10)	ARIMA (2, 1, 1)
Abdominal pain	4.05	Sep 20, 2019	5.14 (4.27–6.00)	6	1.12 (1.00–1.41)	ARIMA (1, 1, 1)
Diarrhea	2.32	Mar 15, 2020	21.31 (19.70–22.92)	27	1.31 (1.22–1.42)	ARIMA (1, 1, 1)
Anorexia	6.15	Oct 4, 2020	3.46 (1.76–5.17)	5	1.44 (0.97–2.84)	ARIMA (0, 1, 2)
Loss of taste ^d	/	Mar 22, 2020	0.50	4	8	/

Abbreviations: RSV, relative search volume; MAPE, mean absolute percentage error; CI, confidence interval; FC, fold change.

The period of the training set was from December 1, 2015 to November 30, 2019 and the period of the testing set was from December 1, 2019 to December 1, 2020.

^a This date corresponds to the highest RSV of each symptom term on Google Trends for the testing set.

^b The predicted RSV of each symptom term was forecasted by the ARIMA model.

^c The point change of each symptom term was calculated by comparing the actual value with the predicted value at the same point in time.

^d The analysis of loss of taste was not suitable for the ARIMA model because this search volume had no fluctuation during the training set period.

The correlations among RSVs of different GI symptom terms

During the outbreak period (January 1, 2020 to May 31, 2020) at the global level, positive correlations were mainly observed among “fever”, “cough”, “vomiting”, “diarrhea”, and “nausea” (Spearman $r > 0.3$, $P < 0.05$; **Fig. 3A**). However, these correlations were distinctly attenuated during the pandemic period (June 1, 2020 to December 1, 2020; **Fig. 3B**) and the whole period (January 1, 2020 to December 1, 2020); **Fig. 3C**). Similar correlation patterns were observed at the country level (**Supplementary Fig. 2**), regardless of certain differences. These results collectively revealed that GI symptoms of diarrhea and loss of taste, along with fever and cough, serve as important and common clinical symptoms of COVID-19 patients.

Associations between RSVs of GI symptoms and COVID-19 outbreak

The distribution of disease symptoms of interest (RSVs) and the daily newly confirmed COVID-19 cases worldwide are shown in **Fig. 4**. Six- and five-knot RCS models were used to fit the shape of the curve representing newly confirmed COVID-19 cases during the outbreak period (January 1, 2020 to May 31, 2020; **Fig. 4A**) and pandemic period (June 1, 2020 to December 1, 2020; **Fig. 4B**). During the outbreak period, newly confirmed cases rapidly increased daily worldwide starting in early January and achieved a relatively stable peak on April 4, 2020, which was an approximately 20-day delay after the peak of abnormal clinical symptom terms of interest (March 15, 2020; **Fig. 4A**). Intriguingly, this predicted time

delay was consistent with the well-accepted incubation period for COVID-19.¹² In contrast, although the daily number of newly confirmed cases continued to increase during the pandemic period (**Fig. 4B**), the RSVs of clinical symptom terms remained smooth. Similar distributions were observed at the country level (**Supplementary Fig. 3**).

To estimate the feasibility of an internet search term as an indicator for a disease outbreak, we initially utilized a single-symptom GAM to assess the prediction effect of disease symptoms on the COVID-19 outbreak with a single lag day within 4 weeks. The two positive-control symptom terms had the strongest prediction effects (fever: $RR = 1.03$, 95% CI = 1.02–1.05; $P = 2.83 \times 10^{-6}$; cough: $RR = 1.03$, 95% CI = 1.01–1.04; $P = 2.30 \times 10^{-4}$) on daily newly confirmed COVID-19 cases on lag 9 and lag 12 days at the global level, respectively (**Table 2, Fig. 5**). In terms of GI symptoms, “diarrhea” ($RR = 1.11$, 95% CI = 1.03–1.20; $P = 6.51 \times 10^{-3}$) and “loss of taste” ($RR = 1.30$, 95% CI = 1.05–1.61; $P = 1.59 \times 10^{-2}$) displayed the highest correlation with increased daily newly confirmed COVID-19 cases on lags 12 and 5 days at the global level, respectively (**Table 2, Fig. 5**). Meanwhile, clinical symptoms presented similar lag correlation patterns with daily newly confirmed COVID-19 cases at the country level but were accompanied by different lag days across countries (**Table 2, Fig. 5**). In addition, the prediction effect of the lag day pattern was consistently attenuated during the COVID-19 pandemic period (**Supplementary Fig. 4**).

Considering that the RSVs for the terms “fever” and “cough” as well as GI symptoms including “diarrhea” and “loss of taste” were

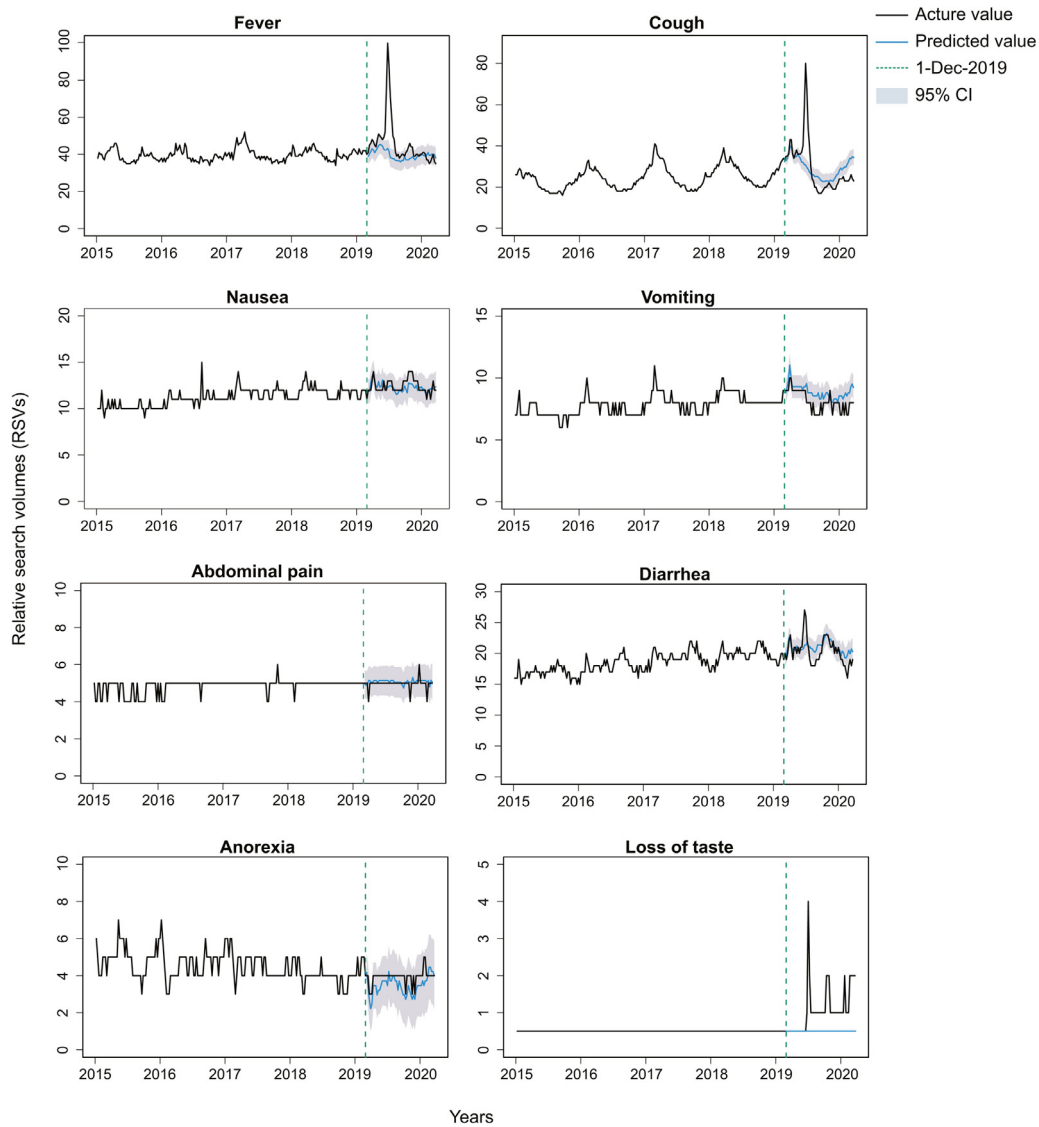


Fig. 2. Fitting and predicting the RSV trends of GI symptom terms with the ARIMA model on Google Trends at the global level. Weekly search data of disease symptom terms from December 1, 2015 to November 30, 2019 were used to construct the ARIMA models, which were used to predict weekly RSV trends of each symptom term from December 1, 2019 to December 1, 2020. The RSV trends of “fever” and “cough” were seated as positive-control symptom terms. The ARIMA model was used to fit and forecast the RSV trends of GI symptom terms. Abbreviations: RSV, relative search volume; ARIMA, autoregressive integrated moving average.

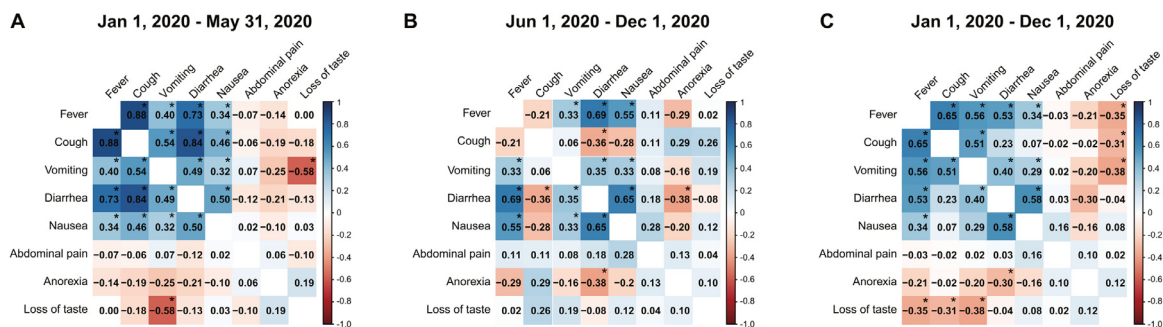


Fig. 3. Spearman’s correlation coefficients among RSVs of symptom terms at the global level. (A) Spearman’s correlation during the COVID-19 outbreak period from January 1, 2020 to May 31, 2020 at the global level; (B) Spearman’s correlation during the COVID-19 pandemic period from June 1, 2020 to December 1, 2020 at the global level; (C) Spearman’s correlation during the COVID-19 epidemic period from January 1, 2020 to December 1, 2020 at the global level. * Spearman $r > 0.3$, $P < 0.05$. The dark blue colours represent strong positive correlations (close to 1), and the dark red colours represent strong inverse correlations (close to -1). Light/white represents an absence of correlation (close to 0). Abbreviations: RSV, relative search volume.

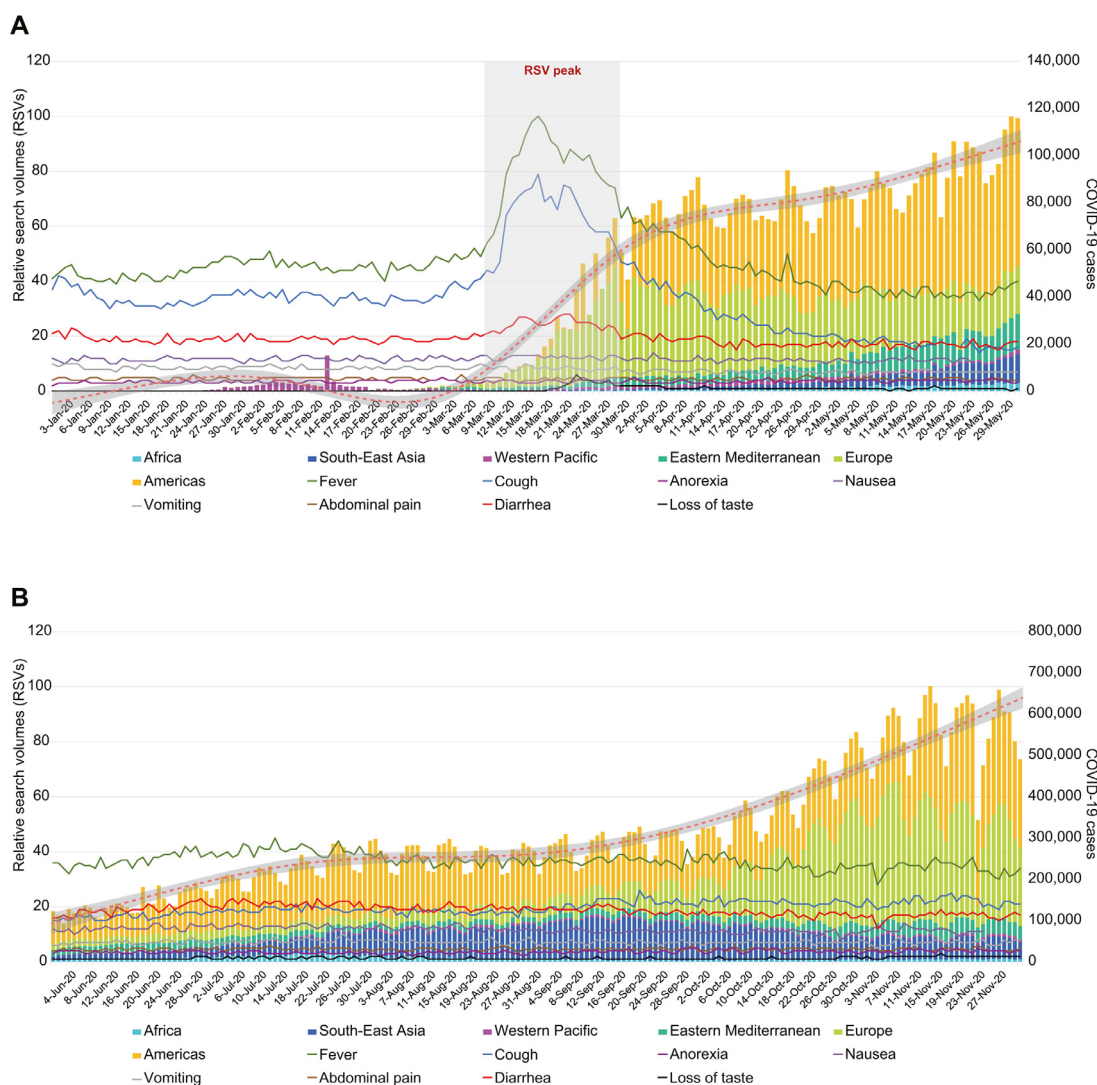


Fig. 4. Time-series trends of RSVs of symptom terms and daily newly confirmed COVID-19 cases at the global level. (A) Time-series plot from January 1, 2020 to May 31, 2020 of the COVID-19 outbreak period. (B) Time-series plot from June 1, 2020 to December 1, 2020 of the COVID-19 pandemic period. The x-axis indicates the time span. Time-series plot presenting both RSV trends of symptom terms of interest (left y-axis) and daily newly confirmed COVID-19 cases at the global level (right y-axis). Google Trends provides daily data of RSV assigned to each symptom term. RSVs of each symptom term were obtained from Google Trends, and newly diagnosed COVID-19 cases were obtained from the WHO COVID-19 Dashboard. Each disease symptom and region are coloured and shaped differently. The dashed line represents the trend of diagnosed cases over time calculated by RCS models, and the shadow represents the 95% CIs of the calculated values. Abbreviations: RSV, relative search volume; WHO, World Health Organization; RCS, restricted cubic spline; CI, confidence interval.

positively and significantly associated with the COVID-19 outbreak, we then assessed the ability of different GI symptom terms to predict the outbreak of COVID-19. At the global level, the RSVs for the GI symptom terms combined with “fever” and “cough” could predict COVID-19 outbreaks approximately 1–18 days in advance (**Supplementary Fig. 5**). At the country level, the predictive ability in the United States, United Kingdom, Australia, and China was up to 19 days, 13 days, 20 days, and 21 days, respectively (**Supplementary Fig. 5**).

Dose-response associations between RSVs of GI symptoms and COVID-19 risk

Moreover, there were significant dose-response associations of daily RSVs for GI symptom terms including “diarrhea” and “loss of taste” with COVID-19 risk at the strongest effect lag days at the global level (**Fig. 6**). The “diarrhea” curves showed an increase with no discernible thresholds. However, the slopes of the “loss of taste” curves were steeper at RSVs lower than one, but seemed to flat-

ten at high ranges. In addition, positive associations were still detectable in country-specific dose-response curves (**Supplementary Fig. 6**).

Discussion

We comprehensively analysed the RSVs of GI symptom terms on internet search engines using time-series analysis and found abnormal trends during the COVID-19 epidemic. Data on daily newly confirmed COVID-19 cases were obtained to analyze the associations between the RSVs of GI symptom terms and the COVID-19 outbreak and pandemic at both global and regional levels. This study also quantitatively evaluated the ability of RSVs of GI symptom terms to predict the COVID-19 outbreak.

In the early stage of the COVID-19 outbreak, its epidemiological and pathological features were not fully understood; thus, it was difficult to take effective measures to stop its outbreak and spread, which ultimately caused a global public health crisis with millions of fatalities. However, early detection of unknown infectious

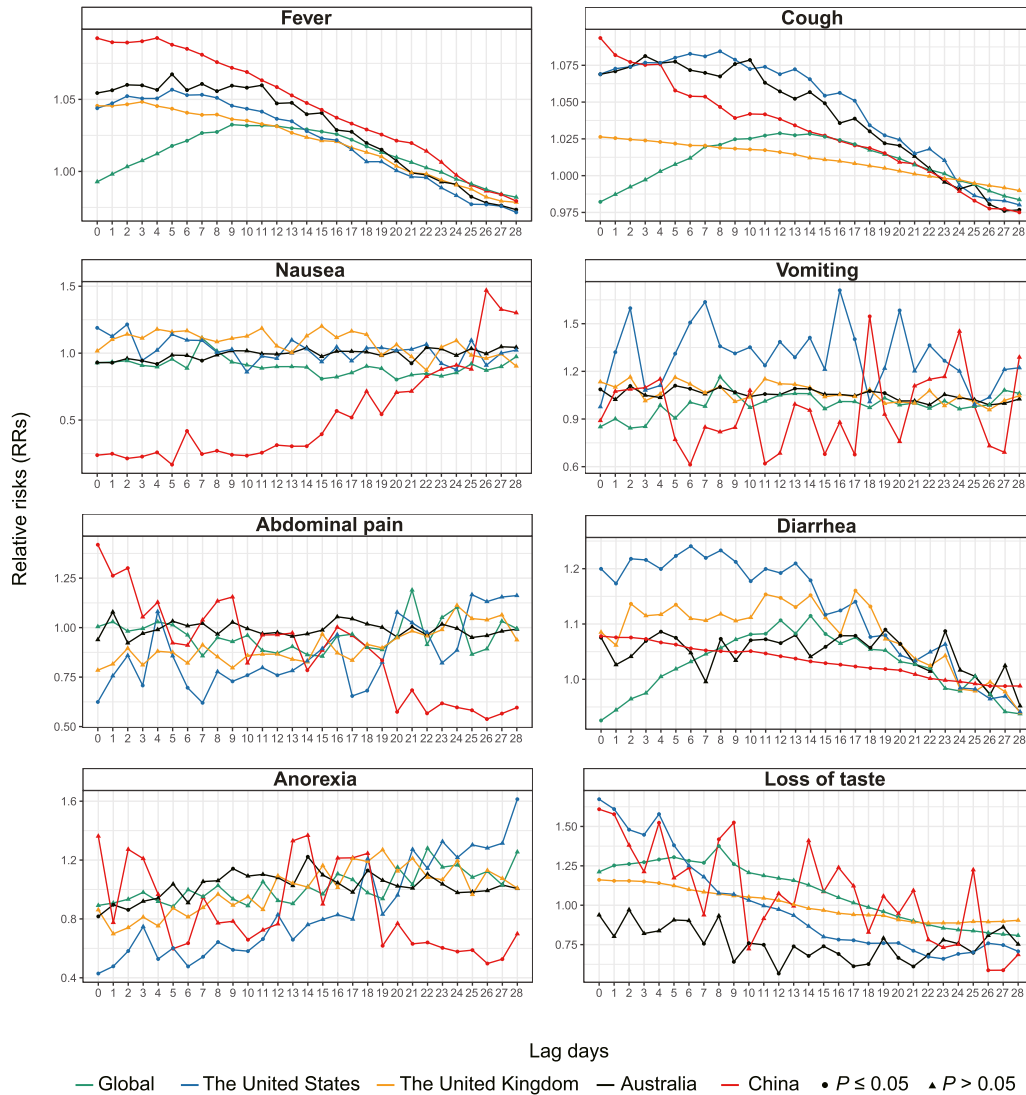


Fig. 5. The estimated effects of RSVs of GI symptom terms on COVID-19 incidence with a single-symptom model during the COVID-19 outbreak period. Estimated RRs and 95% CIs of daily newly confirmed COVID-19 cases for one search volume increase of symptom terms on internet search engines, with different lag days in the single-symptom model during the COVID-19 outbreak period from January 1, 2020 to May 31, 2020 at the global and country levels. Abbreviations: RSV, relative search volume; GI, gastrointestinal; RR, relative risk; CI, confidence interval.

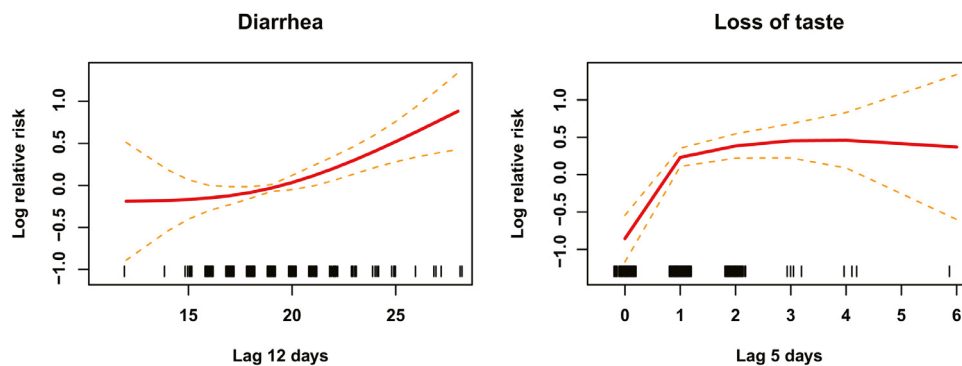


Fig. 6. Dose-response curves between RSVs of GI symptom terms (lag days with the strongest effects) and daily COVID-19 risk at the global level during the COVID-19 outbreak period. The x-axis represents RSVs of GI symptom terms. The y-axis can be interpreted as the relative risk from the one RSV of GI symptom terms increase on COVID-19 incidence. Abbreviations: RSV, relative search volume; GI, gastrointestinal.

Table 2
The effects of RSVs of GI symptom terms on COVID-19 incidence during the COVID-19 outbreak from January 1, 2020 to May 31, 2020.

Region	Symptoms	Lag days ^a	RR (95% CI)	P value ^b
Global level	Fever	9	1.03 (1.02–1.05)	2.83×10^{-6}
	Cough	12	1.03 (1.01–1.04)	2.30×10^{-4}
	Diarrhea	12	1.11 (1.03–1.20)	6.51×10^{-3}
	Loss of taste	5	1.30 (1.05–1.61)	1.59×10^{-2}
The United States	Fever	5	1.06 (1.04–1.07)	2.44×10^{-14}
	Cough	8	1.08 (1.06–1.10)	2.30×10^{-13}
	Diarrhea	6	1.24 (1.17–1.32)	7.79×10^{-13}
	Loss of taste	0	1.67 (1.33–2.10)	8.30×10^{-6}
The United Kingdom	Fever	3	1.05 (1.02–1.07)	1.51×10^{-4}
	Cough	0	1.03 (1.01–1.04)	4.07×10^{-5}
	Loss of taste	0	1.16 (1.03–1.31)	1.54×10^{-2}
Australia	Fever	5	1.07 (1.05–1.08)	6.15×10^{-23}
	Cough	3	1.08 (1.06–1.10)	3.13×10^{-21}
China	Fever	4	1.09 (1.08–1.11)	5.20×10^{-27}
	Cough	0	1.09 (1.08–1.11)	1.74×10^{-27}
	Abdominal pain	0	1.42 (1.14–1.76)	1.66×10^{-3}
	Diarrhea	0	1.08 (1.06–1.09)	8.91×10^{-33}

Abbreviations: RSV, relative search volume; RR, relative ratio; CI, confidence interval.

Note: The effect estimates are presented as the RRs and related 95% CIs in daily newly confirmed COVID-19 cases per search volume increase of GI symptoms.

^a The lag day with the strongest effect.

^b A P value of less than 0.05 was considered to be statistically significant for the effects.

disease activity before large-scale human-to-human transmission will be of great significance to the prevention and control of epidemics. Web-based epidemiology, which mainly focuses on scanning the internet for user-contributed health-related content, has been used to surveil and predict infectious disease outbreaks, including influenza, Middle East Respiratory Syndrome (MERS), and Zika virus.^{13–15}

Here we present a method of analysing large numbers of internet search queries about GI symptom terms to track the COVID-19 outbreak in a population. Clinically, COVID-19 patients present GI symptoms accompanied by symptoms typical of other types of pneumonia, such as fever and cough.^{16,17} Moreover, up to 28% of those with GI symptoms do not have respiratory symptoms⁴. Therefore, GI symptoms are important indicators of SARS-CoV-2 infection in susceptible populations. Symptomatic cases usually initially use internet search engines, such as Google or popular search engines in respective countries, to evaluate whether their symptoms are related to COVID-19 prior to contacting physicians. Google Trends and Baidu Index analyses have been widely used to evaluate disease prevalence and predict infectious disease pandemics.^{8–10,18}

By accurately fitting the RSV trends of GI symptom terms on the internet search engines with the ARIMA model in the past 5 years, we observed that the trends of “diarrhea” and “loss of taste” formed consistent peaks rapidly after December 2019, indicating that RSVs abnormally increased during the COVID-19 epidemic worldwide. However, these abnormally increased peaks were only observed in the early stage of the COVID-19 outbreak period but were not continuously observed during the pandemic period. A possible explanation is that during the outbreak period, potentially susceptible individuals searched for symptoms on the internet, as they did not fully understand the disease. However, with the spread of media information, the public has generally understood the characteristics of COVID-19 infection, and related symptom term searches declined during the pandemic period. The correlation analyses among abnormally searched symptom terms during the COVID-19 epidemic revealed that the RSV trends of “diarrhea” were highly correlated with the positive symptoms of “fever” and “cough”. However, “loss of taste” seems to be an independent symptom, which may suggest that some studies that some infected individuals present anosmia in the absence of other symptoms.^{19,20}

We found that the abnormal increases in RSVs for positive-control symptom terms, including “fever” and “cough”, as well as GI symptom terms, including “diarrhea” and “loss of taste”, were significantly associated with the increase in daily newly confirmed COVID-19 cases worldwide. Specifically, these associations had significant symptom heterogeneity across four representative countries with different lag correlation patterns. Despite these interesting associations, the differences in effective symptoms and lag days across diverse regions remain difficult to explain, which may be attributed to the population density, ecological factors, testing availability, or racial differences of symptoms in different countries.^{21,22}

Traditional surveillance systems such as the US Centers for Disease Control and Prevention publish national and regional data that typically lag behind real-time by 1–2 weeks.²³ These data have no predictive value for the outbreak of infectious diseases. The lag correlation patterns represent different predictable values of GI symptoms for the COVID-19 outbreak. Specifically, in our study, the RSVs of GI symptom terms preceded the outbreak of COVID-19 by about 1–3 weeks, which was slightly longer than the 1–2 week lag time observed in influenza.⁷ In addition, the dose-response curves between RSVs of GI symptom terms and COVID-19 risk showed a consistent increase at the global level but had country-specific curves of different GI symptom terms. Setting RSV thresholds of different symptom terms according to these dose-response results can provide a risk assessment for future COVID-19 outbreaks.

In the context of the COVID-19 outbreak and pandemic, our study emphasizes the powerful advantages of epidemiology based on internet search engine data in the prevention and control of infectious diseases, such as: (1) internet monitoring and big data analysis of unknown infectious diseases can obtain real-time epidemiological data without delay; (2) it can predict the outbreak of infectious diseases in advance compared with traditional case reporting systems; (3) data from the internet can be obtained for free; (4) there is no need for large-scale personnel screening or laboratory testing, avoiding the spread of infectious disease; it can accurately analyze the time, season, and geographical features of infectious disease.

However, several limitations also exist in this study. First, our research methods are applicable only to countries or regions where the internet is popular and not to underdeveloped areas without

the internet. Second, in-country transmission heterogeneities were not captured, as we mainly focused on global and country level effects. As the geographic accuracy of internet records improves, a similar analysis should be performed at smaller scales. Third, detailed information about data processing methods and the user characteristics of search engines are unknown, which may introduce selection bias.

Conclusions

Our global and regional time-series study comprehensively uncovers the association of RSVs of GI symptom terms with daily newly confirmed COVID-19 cases. Strikingly, GI symptoms, including diarrhea and loss of taste, would be good indicators for surveillance of SARS-CoV-2 infection and helpful for the early prediction of COVID-19 outbreaks for up to three weeks. Real-time monitoring of the RSV trends of disease symptom terms on internet search engines and fitting disease prediction models will help policy-makers to make rapid and accurate risk assessments and strategic health care resource allocation before disease outbreaks. The prediction models can be generalized to broader countries and diseases with the widespread global usage of internet search engines.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Declaration of Interest

None.

Acknowledgments

We acknowledged Qianyu Yuan, Li Su, and David C. Christiani from Departments of Environmental Health, Harvard T.H. Chan School of Public Health, who helped to acquire the volumes of search terms from the Google search engine.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.jinf.2021.11.003](https://doi.org/10.1016/j.jinf.2021.11.003).

References

- Guan WJ, Ni ZY, Hu Y, Liang WH, Ou CQ, He JX, et al. Clinical characteristics of coronavirus disease 2019 in China. *N Engl J Med* 2020;**382**(18):1708–20.
- Chen C, Gao G, Xu Y, Pu L, Wang Q, Wang L, et al. SARS-CoV-2–positive sputum and feces after conversion of pharyngeal samples in patients with COVID-19. *Ann Intern Med* 2020;**172**(12):832–4.
- Liang W, Feng Z, Rao S, Xiao C, Xue X, Lin Z, et al. Diarrhoea may be underestimated: a missing link in 2019 novel coronavirus. *Gut* 2020;**69**(6):1141–3.
- Jin X, Lian JS, Hu JH, Gao J, Zheng L, Zhang YM, et al. Epidemiological, clinical and virological characteristics of 74 cases of coronavirus-infected disease 2019 (COVID-19) with gastrointestinal symptoms. *Gut* 2020;**69**(6):1002–9.
- Cholankeril G, Podboy A, Aivaliotis VI, Tarlow B, Pham EA, Spencer S, et al. High prevalence of concurrent gastrointestinal manifestations in patients with SARS-CoV-2: early experience from California. *Gastroenterology* 2020;**159**(2):775–7.
- Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. *J Med Internet Res* 2017;**19**(6):e232.
- Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. *Nature* 2009;**457**(7232):1012–14.
- Halford EA, Lake AM, Gould MS. Google searches for suicide and suicide risk factors in the early stages of the COVID-19 pandemic. *PLoS ONE* 2020;**15**(7):e0236777.
- Yang S, Santillana M, Kou SC. Accurate estimation of influenza epidemics using Google search data via ARGO. *Proc Natl Acad Sci* 2015;**112**(47):14473–8.
- Senecal C, Widmer RJ, Lerman LO, Lerman A. Association of search engine queries for chest pain with coronary heart disease epidemiology. *JAMA Cardiol* 2018;**3**(12):1218–21.
- Ahmad I, Flanagan R, Staller K. Increased internet search interest for GI symptoms may predict COVID-19 cases in US hotspots. *Clin Gastroenterol Hepatol* 2020;**18**(12):2833.
- Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Ann Intern Med* 2020;**172**(9):577–82.
- Shin SY, Seo DW, An J, Kwak H, Kim SH, Gwack J, et al. High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea. *Sci Rep* 2016;**6**(1):1–7.
- Majumder MS, Santillana M, Mekaru SR, McGinnis DP, Khan K, Brownstein JS. Utilizing nontraditional data sources for near real-time estimation of transmission dynamics during the 2015–2016 Colombian Zika virus disease outbreak. *JMIR Public Health Surveill* 2016;**2**(1):e30.
- Woo H, Cho Y, Shim E, Lee JK, Lee CG, Kim SH. Estimating influenza outbreaks using both search engine query data and social media data in South Korea. *J Med Internet Res* 2016;**18**(7):e177.
- Sultan S, Altayar O, Siddique SM, Davitkov P, Feuerstein JD, Lim JK, et al. AGA institute rapid review of the gastrointestinal and liver manifestations of COVID-19, meta-analysis of international data, and recommendations for the consultative management of patients with COVID-19. *Gastroenterology* 2021.
- Zhou Z, Zhao N, Shu Y, Han S, Chen B, Shu X. Effect of gastrointestinal symptoms in patients with COVID-19. *Gastroenterology* 2020;**158**(8):2294.
- Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. *Eurosurveillance* 2020;**25**(10):2000199.
- Eliezer M, Hautefort C, Hamel AL, Verillaud B, Herman P, Houdart E, et al. Sudden and complete olfactory loss of function as a possible symptom of COVID-19. *JAMA Otolaryngol Head Neck Surg* 2020;**146**(7):674–5.
- Menni C, Valdes AM, Freidin MB, Sudre CH, Nguyen LH, Drew DA, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med* 2020;**26**(7):1037–40.
- Team CCR, Team CCR, Team CCR, Bialek S, Bowen V, Chow N, et al. Geographic differences in COVID-19 cases, deaths, and incidence—United States, February 12–April 7, 2020. *Morb Mortal Weekly Rep* 2020;**69**(15):465–71.
- Jia JS, Lu X, Yuan Y, Xu G, Jia J, Christakis NA. Population flow drives spatio-temporal distribution of COVID-19 in China. *Nature* 2020;**582**(7812):389–94.
- <https://www.cdc.gov/flu/weekly/overview.htm>.