

Journal Pre-proof

Association of the COVID-19 pandemic with Internet Search Volumes: A Google TrendsTM Analysis

Maria Effenberger Andreas Kronbichler Jae Il Shin Gert Mayer
Herbert Tilg Paul Perco



PII: S1201-9712(20)30249-6
DOI: <https://doi.org/doi:10.1016/j.ijid.2020.04.033>
Reference: IJID 4121

To appear in: *International Journal of Infectious Diseases*

Received Date: 24 March 2020

Revised Date: 9 April 2020

Accepted Date: 10 April 2020

Please cite this article as: Effenberger M, Kronbichler A, Shin JI, Mayer G, Tilg H, Perco P, Association of the COVID-19 pandemic with Internet Search Volumes: A Google TrendsTM Analysis, *International Journal of Infectious Diseases* (2020), doi: <https://doi.org/10.1016/j.ijid.2020.04.033>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier.

Association of the COVID-19 pandemic with Internet Search Volumes: A Google Trends™ Analysis

Maria Effenberger^{1*}, Andreas Kronbichler^{2*}, Jae Il Shin^{3,4,5}, Gert Mayer², Herbert Tilg¹, Paul Perco²

¹ Department of Internal Medicine I (Gastroenterology, Hepatology, Endocrinology and Metabolism), Medical University Innsbruck, Anichstrasse 35, 6020 Innsbruck, Austria

² Department of Internal Medicine IV (Nephrology and Hypertension), Medical University Innsbruck, Anichstrasse 35, 6020 Innsbruck, Austria

³ Department of Pediatrics, Yonsei University College of Medicine, Seoul, Republic of Korea

⁴ Department of Pediatric Nephrology, Severance Children's Hospital, Seoul, Republic of Korea

⁵ Institute of Kidney Disease Research, Yonsei University College of Medicine, Seoul, Republic of Korea

* These authors contributed equally to this work.

Corresponding author:

Correspondence should be addressed to Paul Perco, Ph.D., P.D.,

Department of Internal Medicine IV (Nephrology and Hypertension), Medical University Innsbruck, Anichstrasse 35, 6020 Innsbruck, Austria

Phone: +43 512 504 25855

E-Mail: paul.perco@i-med.ac.at

Abstract

Objectives

To assess the association of public interest in coronavirus infections with the actual number of infected cases for selected countries across the globe.

Methods

We performed a Google Trends™ search for “Coronavirus” and compared Relative Search Volumes (RSV) indices to the number of reported COVID-19 cases by the European Center for Disease Control (ECDC) using time-lag correlation analysis.

Results

Worldwide public interest in Coronavirus reached its first peak end of January when numbers of newly infected patients started to increase exponentially in China. The worldwide Google Trends™ index reached its peak on the 12th of March 2020 at a time when numbers of infected patients started to increase in Europe and COVID-19 was declared a pandemic. At this time the general interest in China but also the Republic of Korea has already been significantly decreased as compared to end of January. Correlations between RSV indices and number of new COVID-19 cases were observed across all investigated countries with highest correlations observed with a time lag of -11.5 days, i.e. highest interest in coronavirus observed 11.5 days before the peak of newly infected cases. This pattern was very consistent across European countries but also holds true for the US. In Brazil and Australia, highest correlations were observed with a time lag of -7 days. In Egypt the highest correlation is given with a time lag of 0, potentially indicating that in this country, numbers of newly infected patients will increase exponentially within the course of April.

Conclusions

Public interest indicated by RSV indices can help to monitor the progression of an outbreak such as the current COVID-19 pandemic. Public interest is on average highest 11.5 days before the peak of newly infected cases.

Key Words

Public awareness, COVID-19, coronavirus, Google Trends, SARS-CoV-2

Introduction

A novel coronavirus, the acute respiratory syndrome coronavirus 2 (SARS-CoV-2), causes a new disease named Corona Virus Disease 2019 (COVID-19). It was first detected in December 2019 in Wuhan (Hubei, China)(Wang et al., 2020). Due to a high virulence and a high proportion of asymptomatic cases, the outbreak spreads all over the world. On April 5th 2020 the World Health Organization (WHO) reported 1 133 758 confirmed cases. Today, a cumulative mortality rate of 5.5 % (62 784) has been reported.

The internet is increasingly used as a source of health care information. Infodemiology and infoveillance are essential public health informatics methods which are used to analyze search behavior on the internet. Infodemiology is defined as *“science of distribution and determinants of information in an electronic medium, specifically the internet, or in a population, with the ultimate aim to inform public health and public policy”*, while the primary aim of infoveillance is surveillance (Eysenbach, 2009). Infodemiology and infoveillance of epidemiological data are important to increase situational awareness and make suitable interventions(Rivers et al., 2019).

The analysis of relative internet search volumes (RSV) gives information on the extent of public attention (Arora et al., 2019, Kaleem et al., 2019, Ling and Lee, 2016) with Google TrendsTM being one of the most widely used tools for this purpose. RSV are used for real-time analyses for transmissibility, severity, and natural history of an emerging pathogen, as observed with severe acute respiratory syndrome (SARS), the 2009 influenza pandemic, and Ebola (Chowell et al., 2009, Cleaton et al., 2016). The analyses of confirmed cases are particularly useful to infer key

epidemiological parameters, such as the incubation and infectious periods and ongoing outbreaks or an outbreak probability. In addition, Google Trends™ data might be used to forecast an increase in infected cases. A linear time series pattern with official dengue reports, indicating a potential use to monitor public interest before an increase of cases and during the outbreak (Husnayain et al., 2019). Beside infectious diseases, Google Trends™ have been successfully used to forecast the suicide risk increase (Barros et al., 2019).

In this study, we investigated the public interest in COVID-19 since December 31st 2019 comparing Google Trends™ data to data of newly infected COVID-19 cases.

Methods

Retrieving outbreak and confirmed cases numbers from the WHO

Data on confirmed COVID-19 cases were retrieved on the 5th of March from the European Center for Disease Control (ECDC) for the time from the 31st of December 2019 until the 1st of April 2020 (<https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide>).

Worldwide data were retrieved as well as data for the following countries, namely China, Republic of Korea, Japan, Iran, Italy, Austria, Germany, the United Kingdom (UK), the United States (US), Egypt, Australia, and Brazil.

Retrieving Google Trends™ data on COVID-19

The Google Trends™ tool was used to retrieve data on internet user search activities in the context of COVID-19. Google Trends™ enables researchers to study trends and patterns of Google™ search queries (Arora et al., 2019). It was implemented in 2004 and data on internet search queries are available on a monthly basis. Google

Trends™ expresses the absolute number of searches relative to the total number of searches over the defined period of interest (Arora et al., 2019). The retrieved Google Trends™ index ranges from 0 to 100, with 100 being the highest relative search term activity for the specified search query in the time period of interest. Further information on Google Trends™ can be found on the respective help page (<https://support.google.com/trends/>).

Nation-specific Google Trends™ indices were retrieved from December 31st 2019 to April 1st 2020 using the search term “Coronavirus (Virus)”. We retrieved Google Trends™ indices for China, Republic of Korea, Japan, Iran, Italy, Austria, Germany, the UK, the US, Egypt, Australia, and Brazil. The search was performed in English.

Analysis

Daily Google Trends™ indices were used for generation of the lineplots with the ggplot2 package of the statistical software R (version 3.4.1). The ccf function of the tseries R package was used to calculate time-lag correlations between Google Trends™ indices and new cases. A time lag between -35 and +35 days was used in the analysis with the Pearson correlation coefficient serving as correlation measure.

Results

Worldwide interest in coronavirus started on January 20th and reached its first peak on January 31st, a few days after the word was spread on the outbreak in Wuhan, China. The increasing numbers of cases over the globe prompted the WHO to declare the coronavirus outbreak as a pandemic on March 11th, leading to an increase in public interest currently peaking on March 12th 2020 (Figure 1). The data on newly confirmed cases, overall confirmed cases, and overall death worldwide as

well as for the afore-mentioned countries under study are summarized in Table 1. There are two peaks, one sharp increase in numbers when cases were counted based on clinical diagnosis and not from a confirmatory laboratory test in China and the other peak on March 16th due to cases around the globe.

The worldwide initial peak associates with a strong increase of confirmed cases in China. In China, a maximum of Google Trends™ RSV was observed at the end of January with a 5.47-fold increase of cases between January 24th and January 28th. Afterwards, with rigorous measures the relative increase in new cases was slower, and a decrease of new cases was firstly reported to the WHO on February the 6th, with the exception of a sharp increase as mentioned above. The RSV trend followed a similar path, with a steady number of search enquiries around 25% of the maximal interest during the last weeks. (Figure 2). Correlation analysis indicates highest public interest in COVID-19 on average around 11.5 days before the maximum of newly infected cases was reported (Figure 3). In countries with proximity to China such as the Republic of Korea or Japan a high volume of search queries was observed during or closely after the peak was reached in China. A non-comparable smaller peak was observed in countries in the European Union or the US (Figure 2).

In the Republic of Korea, a first Google Trends™ index peak was observed end of January only slightly shifted as compared to the peak in China with a second peak being observed on February 23rd (Figure 2). This second peak in Korea proceeded the peak in newly infected cases by 7 days (Figure 3). Japan's RSV started to increase on February 24th, with a peak on February 27th, also followed by an increase in confirmed COVID-19 cases. In Iran, the most affected country in the Middle East, a strong increase of RSV could be observed on February 18th with a peak between 20th and 22nd of February. The Iranian increase of RSV was five days

before the first confirmed cases in Iran, with also a strong association and prediction of the outbreak, which followed five to seven days later. Egypt, the first country on the African continent with a confirmed COVID-19 case, showed a small RSV peak during the outbreak in China. Furthermore, the RSV started to steadily increase since February 20th with an observed leap in interest on April 1st. Australia showed a similar pattern with an increase in RSV during the first outbreak in China, followed by a decrease afterwards and again an increase since February 23rd, followed by increasing new COVID-19 cases 10 days later (Figure 2).

In European countries, especially in Italy, a small peak in the Google Trends™ analysis was found during the outbreak in China and a climax was found on February 23rd 2020, a few days before the numbers of newly COVID-19 started to increase exponentially. Similar trends were observed in Austria, Germany and the UK with a delay of several days and a second peak, which was accompanied by an increase in numbers in the following days. The highest RSV peak was reached mid of March, which is in line with rigorous policies by the government regarding the rapid spread. The UK and Australia show very similar patterns with highest correlations between RSV indices and newly diagnosed cases found with time lags of -12 and -7 days respectively (Figure 3).

In the US, a steady increase of Google™ search queries since February 27th was observed followed by an outbreak since March 2nd. The peak of search queries was reached on March 12th while the numbers of newly infected patients are still increasing (Figure 2). High correlation with a time lag of -12 might indicate that the curve of newly infected cases might flatten out soon if measures taken by the government are comparable to European countries (Figure 3). Google Trends™ analysis in Brazil showed a peak during COVID-19 outbreak in China and on

February 24th, where the first case of COVID-19 was confirmed in Brazil. Since March 3rd a new increase in RSV is found in Brazil, followed by increasing numbers of newly confirmed cases of COVID-19 (Figure 2).

Discussion

In our study, we found a significant increase in RSV using Google Trends™ for COVID-19 worldwide with a peak of RSVs around 11.5 days prior to the peak in newly diagnosed cases in different countries all over the world. As such, Google Trends™ can be used to associate and predict outbreaks worldwide and provides a valuable picture of the outbreak of COVID-19 in real time.

Close monitoring and continued evolution of enhanced communication strategies is needed that provide general populations and vulnerable populations most at risk with actionable information for self-protection, including identification of symptoms^(Heymann et al., 2020). The application of internet data in health care research, also known as infodemiology, is a promising new field and it may complement and extend the current data sources and foundations(Mavragani and Ochoa, 2019). The attention to COVID-19 increased days to weeks before the actual peak outbreak, not only worldwide, but also in most of the investigated countries in this study. This strongly supports our finding that the RSV is a useful tool to monitor local and global outbreaks of infectious diseases.

The internet is the biggest platform for search engines and social media for real time data and outbreaks. RSV has been used before to detect outbreaks, like the recent severe influenza outbreak in 2009(Cook et al., 2011). Close monitoring and continued evolution of enhanced communication strategies is needed that provide general populations and vulnerable populations most at risk with actionable

information for self-protection, including identification of symptoms(Heymann et al., 2020).

Most countries and the WHO provide awareness – raising and educational programs on COVID-19 via internet. The strong association between RSV and increasing outbreak numbers may be due to implementation of such programs in the different countries.

The impact of web based research continuously grows since the past decade(Jun et al., 2018). Google Trends™ is the only unbiased approach including millions of users and has widely been used in health issues. Public attention in different fields has been published recently (e.g. osteoarthritis, breast cancer or COPD)(Boehm et al., 2019, Jellison et al., 2018, Kaleem et al., 2019). Furthermore, infodemiology and Google Trends™ is used to generate awareness profiles and is a suitable substitute for classical data collection, such as surveys(Jun et al., 2018). Far mostly, Google Trends™ is used to monitor disease control and awareness in cancer, HIV or stroke, but also in rare diseases like antiphospholipid syndrome or systemic lupus erythematosus(Ling and Lee, 2016, Mahroum et al., 2019, Sciascia and Radin, 2017, Sciascia et al., 2018). Definitely, Google Trends™ can be used to detect success rates of awareness programs and predict infectious outbreaks worldwide(McLean et al., 2019, Patel et al., 2020).

There are also some potential limitations of this study. There is no information about the individual searches for the analyzed topics. The selections of spelling/terms might affect the results and conclusions. The importance of accuracy in defining the search queries is exemplified by searching Google Trends™ for the topic “pneumonia”. Pneumonia is associated with COVID-19, although not specifically representing COVID-19. Thus, using the query “pneumonia” may be useful to analyze symptom-

related curiosity, but does not sufficiently represent COVID-19 outbreaks. The number of studies based on Google Trends™ is increasing, but so far there is no standardized procedure for data collection. More guidance by Google™ should be warranted in order to assist researchers to establish an optimal search strategy (Nutti et al., 2014). Despite the fact the Google search is accessible worldwide, the use of different search tools in certain countries like for example Baidu in China might lead to more accurate estimations of public interest. It was for example shown that a high Baidu Search Index (BSI) predicted dengue fever outbreaks in Guangzhou and to a lesser degree in Zhongshan, indicating that BSI might complement traditional dengue fever surveillance in China (Liu et al., 2016). In our study we decided to make use of data from one common framework.

In conclusion, infodemiology and RSV provide a tool to anticipate COVID-19 outbreaks and of other infectious diseases. Information on public interest could be used to monitor the outbreak in northern European countries, Africa or The Americas.

Table 1: Illustration of cases (overall, O), newly diagnosed ones (on a daily basis, N), and deaths (D) is represented. We have chosen a time frame of four days to ensure a comprehensive view. Notably, in some countries there was a steady state of cases (i.e. Italy) before a rapid increase was observed.

Figure legends

Figure 1: Worldwide Relative Search Volume (RSV) for “Coronavirus” and newly confirmed COVID-19 cases.

Figure 2: Relative Search Volume (RSV) for “Coronavirus” and newly confirmed COVID-19 cases for selected countries under study.

Figure 3: Time-lag correlations of Relative Search Volume (RSV) for “Coronavirus” and newly confirmed COVID-19 cases for selected countries under study.

Authors' contributions

Literature search: ME, AK, JIS, GM, HT, PP; Figures: PP; Study design: ME, AK, JIS, GM, HT, PP; Data collection: ME, AK, PP; Data analysis: PP; Data interpretation: ME, AK, JIS, GM, HT, PP; Writing: ME, AK, JIS, GM, HT, PP.

Conflict of interest

The authors declare no conflicts of interest.

Funding source

None

Ethical approval

Not applicable

References

- Arora VS, McKee M, Stuckler D. Google Trends: Opportunities and limitations in health and health policy research. *Health Policy* 2019;123(3):338-41.
- Barros JM, Melia R, Francis K, Bogue J, O'Sullivan M, Young K, et al. The Validity of Google Trends Search Volumes for Behavioral Forecasting of National Suicide Rates in Ireland. *Int J Environ Res Public Health* 2019;16(17).
- Boehm A, Pizzini A, Sonnweber T, Loeffler-Ragg J, Lamina C, Weiss G, et al. Using Google Trends to investigate global COPD awareness. *Eur Respir J* 2019;54(2).
- Chowell G, Bertozzi SM, Colchero MA, Lopez-Gatell H, Alpuche-Aranda C, Hernandez M, et al. Severe Respiratory Disease Concurrent with the Circulation of H1N1 Influenza. *New Engl J Med* 2009;361(7):674-9.
- Cleaton JM, Viboud C, Simonsen L, Hurtado AM, Chowell G. Characterizing Ebola Transmission Patterns Based on Internet News Reports. *Clin Infect Dis* 2016;62(1):24-31.
- Cook S, Conrad C, Fowlkes AL, Mohebbi MH. Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. *PLoS One* 2011;6(8):e23610.
- Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *J Med Internet Res* 2009;11(1):e11.
- Heymann DL, Shindo N, Scientific WHO, Technical Advisory Group for Infectious H. COVID-19: what is next for public health? *Lancet* 2020;395(10224):542-5.
- Husnayain A, Fuad A, Lazuardi L. Correlation between Google Trends on dengue fever and national surveillance report in Indonesia. *Glob Health Action* 2019;12(1):1552652.
- Jellison SS, Bibens M, Checketts J, Vassar M. Using Google Trends to assess global public interest in osteoarthritis. *Rheumatol Int* 2018;38(11):2133-6.

- Jun S, Yoo H, Choi S. Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change* 2018;130:69-87.
- Kaleem T, Malouff TD, Stross WC, Waddle MR, Miller DH, Seymour AL, et al. Google Search Trends in Oncology and the Impact of Celebrity Cancer Awareness. *Cureus* 2019;11(8):e5360.
- Ling R, Lee J. Disease Monitoring and Health Campaign Evaluation Using Google Search Activities for HIV and AIDS, Stroke, Colorectal Cancer, and Marijuana Use in Canada: A Retrospective Observational Study. *JMIR Public Health Surveill* 2016;2(2):e156.
- Liu K, Wang T, Yang Z, Huang X, Milinovich GJ, Lu Y, et al. Using Baidu Search Index to Predict Dengue Outbreak in China. *Sci Rep* 2016;6:38040.
- Mahroum N, Bragazzi NL, Brigo F, Waknin R, Sharif K, Mahagna H, et al. Capturing public interest toward new tools for controlling human immunodeficiency virus (HIV) infection exploiting data from Google Trends. *Health Informatics J* 2019;25(4):1383-97.
- Mavragani A, Ochoa G. Google Trends in Infodemiology and Infoveillance: Methodology Framework. *JMIR Public Health Surveill* 2019;5(2):e13439.
- McLean S, Lennon P, Glare P. Internet search query analysis can be used to demonstrate the rapidly increasing public awareness of palliative care in the USA. *BMJ Support Palliat Care* 2019;9(1):40-4.
- Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. *PLoS One* 2014;9(10):e109583.
- Patel MS, Halpern JA, Desai AS, Keeter MK, Bennett NE, Brannigan RE. Success of Prostate and Testicular Cancer Awareness Campaigns Compared to Breast Cancer Awareness Month According to Internet Search Volumes: A Google Trends Analysis. *Urology* 2020.
- Rivers C, Chretien JP, Riley S, Pavlin JA, Woodward A, Brett-Major D, et al. Using "outbreak science" to strengthen the use of models during epidemics. *Nat Commun* 2019;10.
- Sciascia S, Radin M. What can Google and Wikipedia can tell us about a disease? Big Data trends analysis in Systemic Lupus Erythematosus. *Int J Med Inform* 2017;107:65-9.
- Sciascia S, Radin M, Unlu O, Erkan D, Roccatello D. Infodemiology of antiphospholipid syndrome: Merging informatics and epidemiology. *Eur J Rheumatol* 2018;5(2):92-5.

Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern.

Lancet 2020;395(10223):470-3.

	20-Jan-20 O/N/D	24-Jan-20 O/N/D	28-Jan-20 O/N/D	01-Feb-20 O/N/D	05-Feb-20 O/N/D	09-Feb-20 O/N/D	13-Feb-20 O/N/D	17-Feb-20 O/N/D	21-Feb-20 O/N/D	25-Feb-20 O/N/D	29-Feb-20 O/N/D	03-Mar-20 O/N/D	07-Mar-20 O/N/D	11-Mar-20 O/N/D
China	280/74/3	830/265/25	4537/1796/106	11821/2102/259	24363/3893/491	37251/2657/812	46550/1820/1368	70635/2051/1772	75569/894/2239	77780/518/2666	79394/435/2838	80304/130/2946	80813/102/3073	80955/3162
Republic of Korea	1/0/0	2/1/0	4/0/0	12/1/0	18/2/0	27/3/0	28/0/0	30/1/0	204/100/1	977/214/10	3150/813/17	4812/600/6	6767/483/44	7755/60
Japan	1/0/0	1/0/0	6/1/0	17/3/0	33/13/0	26/1/0	29/0/0	59/6/1	93/8/1	157/13/1	230/20/5	268/14/0	408/59/6	568/5
Iran	-	-	-	-	-	-	-	-	5/3/2	61/18/12	388/143/34	1501/523/66	4747/1234/124	8042/291
Italy	-	-	-	2/0/0	2/0/0	3/0/0	3/0/0	3/0/0	3/0/0	229/105/6	888/238/21	2036/347/52	4636/778/197	10149/631
Austria	-	-	-	-	-	-	-	-	-	-	5/1/0	18/8/0	66/19/0	182/5
Germany	-	-	1/0/0	7/2/0	12/0/0	14/0/0	16/0/0	16/0/0	16/0/0	16/0/0	57/31/0	157/28/0	639/105/0	1296/2
UK	-	-	-	-	2/0/0	3/0/0	9/1/0	9/0/0	9/0/0	13/0/0	20/4/0	39/3/0	167/49/1	373/51
US	-	1/1/0	5/0/0	7/1/0	11/0/0	12/0/0	14/1/0	15/0/0	15/0/0	53/18/0	62/3/0	64/2/2	213/65/11	696/25
Egypt	-	-	-	-	-	-	-	1/0/0	1/0/0	1/0/0	1/0/0	2/0/0	3/0/0	59/0/
Australia	-	-	5/1/0	12/3/0	13/1/0	15/0/0	15/0/0	15/0/0	17/2/0	22/0/0	24/1/0	33/6/1	62/5/2	112/21
Brazil	-	-	-	-	-	-	-	-	-	-	1/0/0	2/0/0	13/6/0	34/9/1
Worldwide	282/74/3	846/265/25	4593/1795/106	11953/2128/259	24554/3925/492	37558/2676/813	46997/1826/1369	71429/2162/1775	76769/1021/2247	80239/908/2700	85403/1753/2924	90869/1922/3112	101927/3735/3486	11831/4620/4292

Highlights

- Worldwide public interest in COVID-19 increased when numbers of newly infected patients started to increase exponentially in China
- The peak of worldwide interest was mid of March 2020 shortly after COVID-19 was declared a pandemic
- Public interest in COVID-19 correlates with the number of newly reported COVID-19 cases, with highest interest observed on average 11.5 days before the peak of newly reported COVID-19 cases

