

**VILNIUS UNIVERSITY
MEDICAL FACULTY**

The Final thesis

**Forecasting and surveillance of COVID-19 spread using Google Trends:
systematic review**

Student **Tobias Saegner, VI year, IV group**

Department/ Clinic: **Institute of Health Sciences, Department of Public Health**

Supervisor

Dr. Donatas Austys

The Head of Department/Clinic

Prof. dr. Rimantas Stukas

2022

Email of the student

tobias.saegner@mf.stud.vu.lt

t.saegner@gmail.com

Abstract: Background and Objectives: The probability of future COVID-19 waves remains high, thus COVID-19 surveillance and forecasting remains important. Online search engines harvest vast amounts of data from general population in real time and make this data publicly accessible via such tools as Google Trends (GT). Therefore, the aim of this study was to systematically review literature about possible use of GT for COVID-19 surveillance and prediction of its outbreaks. The objectives were to determine the main findings about GT, also most frequently used keywords, languages, statistical methods and time periods.

Materials and Methods: In November 2021, following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, we collected articles about possible use of GT for COVID-19 surveillance. We resulted in 44 publications that were used in this review.

Results: The majority of the studies (84.1%) included in this review showed positive results of the possible use of GT for forecasting of COVID-19 outbreaks. Most of the studies were performed in English-speaking countries (68.2%). The most frequently used keyword was “coronavirus” (47.7%), followed by “COVID-19” (27.3%) and “COVID” (20.4%). Many authors have made analyses in multiple countries (50%) and got the same results for the majority of them, thus showing the robustness of the chosen methods. Various methods including long short-term memory (4.5%), random forest regression (4.5%), Adaboost algorithm (2.3%), autoregressive integrated moving average, neural network autoregression (2.3%), and vector error correction modelling (2.3%) were used for the analysis. It was seen that most of the publications with positive results were using data from first wave. Later the search volumes reduced even though the incidence peaked.

Conclusions: In most countries, the use of GT data would be beneficial for forecasting and surveillance of COVID-19 spread. However, effectiveness of the use of it might be influenced by many factors, thus longer studies are needed in order to find the most effective solutions for prediction.

Keywords: COVID-19, forecasting, surveillance, Google Trends

1. INTRODUCTION

Coronavirus disease (COVID-19), caused by the novel acute respiratory syndrome coronavirus 2 (SARS-CoV-2), is an infectious disease with high virulence and a high proportion of asymptomatic cases, which, together with other factors like a long period from infection to the onset of the symptoms, symptoms' similarity to a regular cold, continuous social interactions, led to a world-wide virus outbreak [1–3].

Early detection of COVID-19 outbreaks is crucial for multiple reasons: i) to prepare hospitals and staff, including efficiently allocating protective gear and medical equipment [4] as well as testing tents, setting up IT infrastructure (patient registration, databases etc.); ii) to prepare governments for actions, such as imposing curfew, ordering equipment, drawing up guidelines for businesses and events; iii) to improve public messaging and warn people about the risks and their prevention; iv) to prevent further spread of infection [4] by imposing local quarantine or taking other preventive measures. The probability of future COVID-19 waves remains high [5], thus COVID-19 surveillance and forecasting remains important.

Online search engines harvest vast amounts of data from general population in real time and make this data publicly accessible. This raises the interest of using such data for surveillance and forecasting of disease outbreaks [6]. Among internet-based tools for analysis of search queries used to search for specific information, the most acclaimed one is Google Trends (GT) [7,8]. As stated by other researchers, GT can be employed to solve public health issues as it provides valuable information about current concerns and health-related problems in a general society, especially in the field of infectious diseases [7] and therefore could be used for prediction of upcoming disease waves.

GT as a prediction tool has been used for many different diseases in the past two decades, including Influenza [9], Zika virus disease [10], Middle East Respiratory Syndrome (MERS) [11], Malaria [12]. These studies provided diverging results, which makes it difficult to make generalised conclusions about a possibility to use GT for prediction and surveillance of infectious diseases. When it comes to COVID-19, it is important to assess GT ability to detect changes in numbers of people who possibly do not perform COVID-19 tests, but nonetheless feel symptoms or who suspect that they had contact with an infected person and can infect others. This could be used for prediction of COVID-19 outbreaks. Therefore, **the aim** of this study was to systematically review literature about possible use of GT for COVID-19 surveillance and prediction of its outbreaks.

2. MATERIALS AND METHODS

This systematic review was performed following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines [13]. The PubMed search engine was used to search for scientific publications on November 2021.

Inclusion and Exclusion Criteria

The search phrase used for the search query was “Google Trends” AND “Covid-19”. For the initial search of publications, we have not used any language, publication type or other criteria filters. The initial search yielded 246 results (Figure 1). All publications were reviewed for the following inclusion criteria: i) primary original articles addressing the usage of GT for Covid-19 prediction and/or surveillance; ii) articles available in full text; and exclusion criteria: i) publications that had only part of the search phrase in the title or the snippet of the abstract which made us suppose that the publication was not about usage of GT tool; ii) publications with type review, letter, comment, correspondence or presentation, iii) publications written in any other language than English, iv) publications where data obtained from other sources than GT (e.g., WikiTrends, Twitter, etc.) was analysed.

Firstly, with respect to inclusion and exclusion

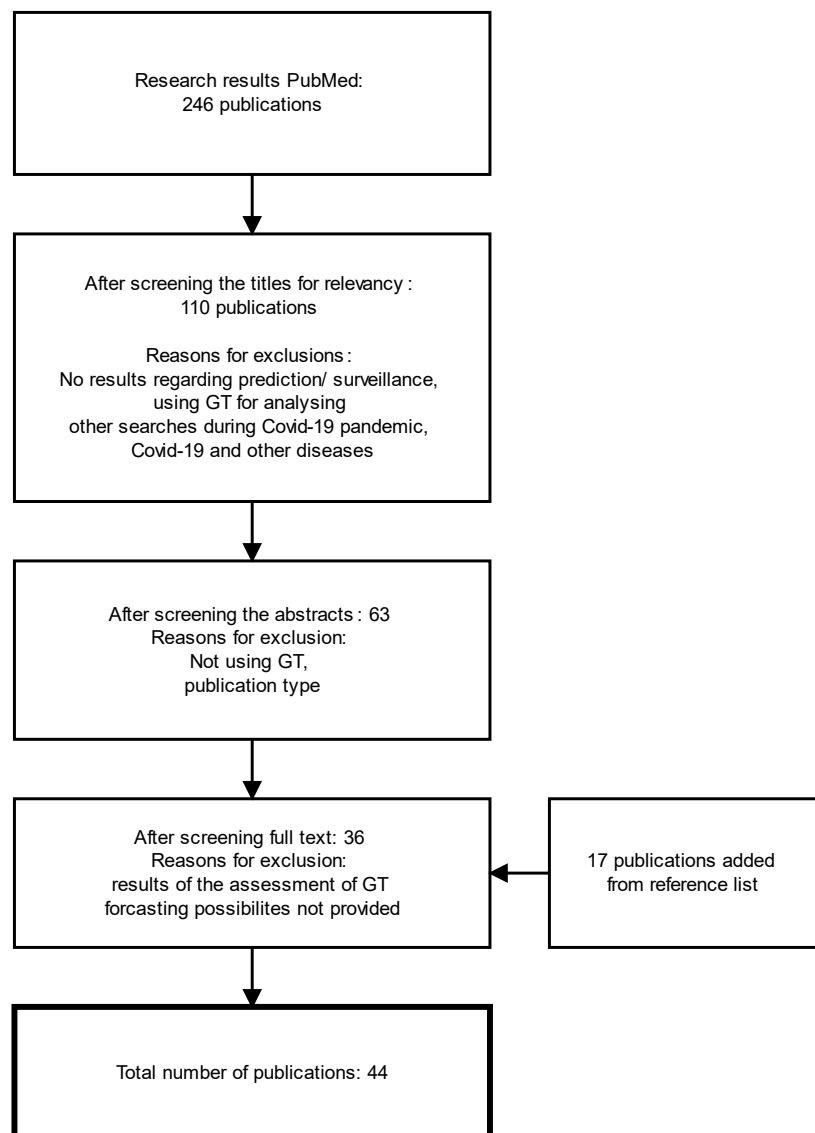


Figure 1. Selection process of the articles to review.

criteria the publication titles were screened to determine if the publication could possibly fit the scope of this review. As a second step, abstracts of the selected publications were screened to verify the relevance of the publication. The full text was downloaded only if an abstract

showed that the publication might be relevant to this review. Full texts were then analysed to include only those articles which provided the results of the assessment of GT forecasting possibilities for Covid-19 disease. In addition, the reference lists of included publications were reviewed according to the same criteria for those not uncovered with the initial search. After completing all these steps and removing duplicates we concluded with 44 articles meeting all the criteria (Figure 1). From each included publication, we extracted such data as year of publication, short description of the main findings, country where GT data was collected, keywords used by people in that country, period of data collection for GT analysis and the statistical analysis method(s) used to analyse the data.

Statistical analysis

With respect to the relatively large number of publications included into this review, relative frequencies (%) were counted in order to show the part of publications with different year of publication, positive findings, negative findings, country where GT data was collected, keywords, periods of data collection and the statistical analysis method(s) used.

3. RESULTS

Most of the studies (84.1%) included in this review showed positive results of the possible use of GT for forecasting of COVID-19 outbreaks (Table 1, Table 2). Most of the publications with positive results were performed in Western countries – mostly in Europe [8,14,23–32,15,33–39,16–22] (61.4%), less in USA [5,7,31–33,35,38–43,8,44–48,14,17–19,23,26,29] (56,8%), Australia [8,19,23,31–33,38] (15.9%) and Canada [19,31,33,38,46] (11.4%). The rest 50% of the studies were performed in the middle east [8,14,18,27,31–33,49] (18.2%), India [18,29,31–33,38,50,51] (18.2%) and China [8,16,18,31,33,52] (13.6%). Half (50%) of the included studies analysed data from multiple countries and half (50%) analysed GT data in single countries. Most of the studies analysed GT data in USA [5,7,40–45,47,48] (22.7%), followed by Italy [21,24,30,36,37] (11.4%), India [50,51] (4.5%), Iran [49] (2.3%), Spain [20] (2.3%), Germany [28] (2.3%), Taiwan [53] (2.3%) and China [52] (2.3%).

Table 1. Publications with positive results of GT use for COVID-19 prediction and surveillance

Author and year	The main findings about Google Trends	Country	Period	Keywords
Husnayan et.al (2020)	GT can be used for public restlessness monitoring towards Covid-19 pandemic 1-3 days before the increase of confirmed cases.	TW	12 2019 – 02 2020	Coronavirus, handwash, face masks
Walker et al. (2020)	Strong correlation between smell-related information	IT, ES, UK, US, DE, FR, NL, IR	12 2019 – 03 2020	Smell, loss of smell, anosmia, hyposmia, olfaction, taste, loss of taste, dysgeusia. The

	search frequency and onset of Covid-19 infection.			keywords were automatically translated to national languages of study countries.
Mavragani (2020)	Significant correlations between online interest of coronavirus and Covid-19 cases and deaths.	IT, ES, FR, DE, UK	01 2020 – 03 2020	Coronavirus
Venkatesh and Gandhi (2020)	Google Web, together with other internet-based tools might be useful in predicting Covid-19 outbreaks 2-3 weeks earlier than conventional disease surveillance.	IN	01 2020 – 04 2020	coronavirus, COVID, COVID 19, corona, virus
Kurian et al. (2020)	The information obtained from US GT precedes Covid-19 outbreaks. This information could allow better preparation and planning of health care systems.		01 2020 – 04 2020	COVID symptoms, coronavirus symptoms, sore throat+shortness of breath+fatigue+cough, coronavirus testing center, loss of smell, Lysol (sanitizer), antibody, face mask, coronavirus vaccine, COVID stimulus check
Panuganti et al. (2020)	Google search of fever and shortness of breath are better indicators of Covid-19 incidence than anosmia.	US	01 2020 – 04 2020	COVID, coronavirus, COVID-19, SARS-CoV2 and COVID19, nonsmell symptoms of (excluding COVID-19 (shortness of breath, fatigue, a short cough and fever) loss of smell, anosmia, lose smell, sense of smell, cannot smell, can't smell and hyposmia, nasal irrigation and sinus rinse, (dysgeusia, taste change and taste loss, COVID, coronavirus, COVID-19, SARS-CoV2 and COVID19), (shortness of breath, fatigue, cough and fever), and smell loss (anosmia, loss of smell, reduced smell, decreased smell, lose your sense of smell, lost sense of smell, decreased sense of smell, decrease your sense of smell, decreased my sense of smell, reduce your sense of smell, reduced my sense of smell, reduced sense of smell, loss of sense of smell, loss of smell, hyposmia
Mavragani and Gkillas (2020)	Significant correlation found between GT search queries and Covid-19 incidence.	US	03 2020 – 04 2020	coronavirus (virus) and coronavirus (search term)
Higgins et al. (2020)	Many search terms showed significant correlations with Covid-19 cases and mortality rate.	CN, US, IT, ES	01 2020 – 04 2020	Real world deaths, Coronavirus, COVID-19, Fever, SOB, Cough, Sputum, Anosmia, Dys/ageusia, Nasal congestion, Rhinorrhea, Sneezing, Sore throat, Headache, Myalgia, Chest pain, Eye pain, Diarrhea
Ahmad et al. (2020)	Google searches for gastrointestinal symptoms preceded the increase in Covid-19 cases in a predictable manner.	US	01 2020 – 04 2020	ageusia, abdominal pain, loss of appetite, anorexia, diarrhea, and vomiting
Cherry et al. (2020)	GT data containing searches related to loss of smell could potentially identify Covid-19 outbreaks.	IT, ES, FR, BR, US	02 2020 – 05 2020	loss of sense of smell, loss of sense of taste, sense of smell, sense of taste
Cousins et al. (2020)	Identifiable patterns in internet searches could predict Covid-19 outbreaks, although stochastic changes in search intensity can alter these predictions.	US	01 2020 – 04 2020	463 unique search queries. Appendix A.

Sharma and Sharma (2020)	A positive correlation between Covid-19 cases and GT values has been recorded.	US, ES, IT, FR, UK, CN, IR, IN	03 2020 – 04 2020	COVID-19
Schnoell et al. (2021)	Clear correlation found between GT data and Covid-19 incidence. GT data might be useful in selecting the best timing for web-based Covid-19 specific information and prevention measures.	AU, BR, CA, DE, IT, ZA, ROK, ES, UK, US	01 2020 – 06 2020	Coronavirus, corona
Jimenez et al. (2020)	Significant correlation found between the rise of Covid-19 incidences and GT search queries with a lag of 11 days.	ES	02 2020 – 05 2020	cansancio, which translates as fatigue; coronavirus, COVID 19, covid 19, and COVID19; diarrea, which translates as diarrhea; dolor de garganta, which translates as sore throat; fiebre, which translates as fever; neumonia, which translates as pneumonia and was searched without an accent due to being more relevant; perdida de olfato, which translates as lost sense of smell and was also searched without an accent; and tos, which translates as cough.
Lippi et al. (2020)	Significant correlations found between GT search data and newly diagnosed Covid-19 cases with a 3-week lag.	IT	02 2020 – 05 2020	tosse (i.e., cough), febbre (i.e., fever) and dispnea (i.e., dyspnea)
Strzelecki et al. (2020)	There was a correlation between Covid-19 spread and GT search data for personal protective gear and hand hygiene.	PL, PT	01 2020 – 06 2020	máscara cirúrgica (face mask), desinfetante (sanitizer) and álcool (alcohol)
Grau et al. (2020)	Strong correlations found between Covid-19 related search terms and cases and mortality rates from Covid-19.	AU, DE, IT, ES, UK, US	11 2019 – 04 2020	keywords used in three categories and four languages: Government Policy, Medical Interventions and Misinformation
Rajan et al. (2020)	GT data could be used to identify active disease transmission areas in the beginning of new outbreaks.	US	10 2019 – 05 2020	diarrhea, nausea, vomiting, and abdominal pain. The terms fever and cough were included as positive controls. The term constipation was included as a negative control.
Xie et al. (2020)	Monitoring internet search activity could prevent and control the epidemic and rumors around it.	CN	01 2020 – 02 2020	Coronavirus
Hartwell et al. (2020)	GT data relating to the public interest of Covid-19 preventative measures correlated with stay-at-home expiration dates and decreased new Covid-19 cases after that expiration. States with higher interest in preventative measures had higher Covid-19 related deaths per capita and higher case-fatality rates.	US	05 2020	hand sanitizer, social distancing, COVID testing, contact tracing
Effenberger et al. (2020)	Significant correlations were found between GT data relating to coronavirus and new Covid-19 cases across studied countries. The time lag was 11.5 days.	KR, JP, IR, IT, AT, DE, UK, US, EG, AU, BR, CN	12 2019 – 04 2020	Coronavirus (virus)

Lin et al. (2020)	Google searches for “wash hands” from January to February correlated with lower Covid-19 spread from February to March in 21 countries.	IT, IR, KR, FR, ES, DE, US, CH, NL, SE, NO, AT, AU, CA, JP, UK, BE, SG, HK, TW, TH	01 2020 – wash hands, face mask 02 2020
Brunori and Resce (2020)	Significant positive correlation found between google search queries of Covid-19 symptoms and reported Covid-19 deaths.	IT	02 2020 – ‘fever’, ‘dry cough’, ‘cough’, ‘sore throat’, 03 2020 ‘loss of sense of smell’, and ‘loss of sense of taste’.
Sulyok et al. (2021)	String positive correlation found between Google search queries for coronavirus and Covid-19 cases in Europe.	BE, FE, DE, HU, IE, IT, NL, NO, ES, SE, CH, UK	01 2020 – Coronavirus 03 2020
Abbas et al. (2021)	The dynamics of the correlations found between GT data Covid-19 cases and deaths suggest that it would be possible to make predictions of Covid-19 cases and mortality rates up to 3 weeks in advance.	US	Dataset released 09 2020, accessed 11 2020 422 symptoms and conditions dataset. Appendix B.
Pellegrini et al. (2021)	GT data on conjunctivitis reveals significant correlations with Covid-19 new cases with a lag of 14-18 days.	IT, FR, UK, US	01 2020 – conjunctivitis” and the translation in Italian 04 2020 (“congiuntivite”) and French language (“conjonctivite”)
Yousefinaghani et al. (2021)	GT data allowed to identify starts and peaks of Covid-19 waves 1 and 3 weeks earlier, respectively. Strong correlation was found between Twitter/GT data and the number of Covid-19 cases in Canada with 3–5-week lags.	CA, US	01 2020 - Shortness of breath, cough, fever, sore throat, 09 2020 loss of smell, loss of taste, Face mask, quarantine, wearing mask, wash hand, covid-19 vaccine, covid-19 vaccine, covid vaccine, corona vaccine, coronavirus vaccine, physical distancing, social distancing
Cinarka et al. (2021)	Online interest shown in Covid-19 pulmonary symptoms can reliably predict later reported cases of the first Covid-19 wave.	TR, IT, ES, FR, UK	01 2020 - fever, cough, dyspnea 08 2020
Husnayain et al. (2021)	Significant correlations between Covid-19 and GT data reached their highest point in June and decreased as the outbreak progressed.	US	01 2020 – Data retrieved for Covid-19 related terms, 12 2020 topics and disease; the top related queries; and most-searched COVID-19 terms in 2020 with a lag of 7 days
Kristensen et al. (2021)	Significant correlations found between term “RKI” and increase in Covid-19 cases (2-12 days lag). Similar pattern was observed for the term “corona”. Searches for “protective mask” peaked 6-12 days after the peak of Covid-19 cases.	DE	02 2020 – ‘RKI’ (Robert Koch Institut), ‘Mundschutz’ 04 2020 (protective mask) and ‘corona’

TW – Taiwan, IT – Italy, ES – Spain, UK – United Kingdom, US – United States, DE – Germany, FR – France, NL – Netherlands, IR – Iran, IN – India, CN – China, BR – Brazil, AU – Australia, CA – Canada, ZA – South Africa, PL – Poland, PT – Portugal, KR – Republic of Korea, JP – Japan, AT – Austria, EG – Egypt, CH – Switzerland, SE – Sweden, NO – Norway, BE – Belgium, SG – Singapore, HK – Hong Kong, TH – Thailand, HU – Hungary, TR – Turkey, IE – Ireland

Differences between countries

Most of GT Covid-19 related analyses were performed in English-speaking countries: mostly USA [5,7,31–33,35,38–43,8,44–48,14,17–19,23,26,29] (56.8%), as well as United Kingdom [8,14,32–34,38,39,54,15,18,19,23,26,27,29,31] (36.4%), Australia [8,19,23,31–33,38] (15.9%) and Canada [19,31,33,38,46] (11.4%). Also, more studies were performed in bigger countries, i.e., those with many residents, as opposed to smaller ones. Moreover, it is seen that GT data analysis was performed more in high-income countries compared to medium and low-income ones.

Time periods

GT seemed to have a higher prediction capability during the first wave of the Covid-19 pandemic (most of the studies (77.2%) took GT data from 01 2020 to 05 2020). All studies reviewed in this article used GT data obtained in 2020 (some starting December 2019) with only two extending their GT data collection to previous years [32,38] for comparison.

Keywords

The most frequently used keyword was “coronavirus” (47.7%), followed by “COVID-19” (27.3%) and “COVID” (20.4%). Other variations included “corona”, “SARS-CoV2”, “COVID19” (or other variations). Specific symptoms showed to be less frequently used in Google searches, however searches relating to loss of smell and taste (31.8%) were rather common as well.

Table 2. Publications where GT data was incorporated into complex analytical models

Author and year	The main findings about Google Trends	Country	Period	Keywords
Ayyoubzadeh et. al. (2020)	Data mining algorithms (linear regression and long short-term memory) can predict Covid-19 outbreak trends.	IR	02 2020 – 03 2020	Corona, COVID-19, Coronavirus, Antiseptic selling, Antiseptic buying, Handwashing, Hand sanitizer, Ethanol, Antiseptic
Prasanth et al. (2021)	Data obtained from GT significantly improved deep learning model (long short-term memory optimized with Grey Wolf optimization) for forecasting Covid-19 numbers.	IN, US, UK	02 2020 – 05 2020	Coronavirus symptoms, Coronavirus, Covid, Handwash, Healthcenter, Mask, Positive cases, Sanitizer, Coronavirus vaccine
Niu et al. (2021)	GT data combined with Adaboost algorithm had strong predictive ability of Covid-19 infection with hopes to further enhance the online prediction system.	IT	02 2020 – 03 2020	40 keywords. Appendix C.
Peng et al. (2020)	A model with GT data and Random Forest Classification, developed from 20 countries	202 countries.	01 2020 – 04 2020	Coronavirus, Pneumonia, Cough, Diarrhea, Fatigue, Fever, Nasal congestion and Rhinorrhea

	worldwide, can be used for epidemic alert level prediction.	Appendix D.		
Rabiolo et al. (2021)	GT data could improve statistical models (ERS, ARIMA and NNA models fitted on the first two principal components) of nowcasting and forecasting Covid-19 incidence with a 15-day time lag and could be used as one of surveillance systems for this disease.	AU, BR, FR, IN, IR, ZA, UK, US	01 2015 – 07 2020 (weekly data) and 01 2020 – 12 2020 (daily data)	Abdominal pain, ageusia, anorexia, anosmia, bone pain, chills, conjunctivitis, cough, diarrhea, eye pain, fatigue, fever, headache, myalgia, nasal congestion, nausea, rhinorrhea, shortness of breath, sore throat, and tearing.
Turk et al. (2021)	GT data was incorporated in a vector error correction model, which showed very good results in forecasting regional Covid-19 hospital census.	US	02 2020 - 08 2020	Coronavirus, covid testing + covid test + covid19 Testing + covid19 test + covid 19 Testing + covid 19 test, headache, pneumonia, “shortness of breath” + “trouble breathing” + “difficulty breathing”, CDC
Peng et al. (2021)	Random forest regression algorithm with integrated previous incidence and GT data was able to accurately predict increase in Covid-19 cases in most countries 7 days in advance.	215 countries. Appendix E.	01 2020 – 07 2020	14 terms, including coronavirus, pneumonia, and Covid-19; 6 symptom-related terms (cough, diarrhea, fatigue, fever, nasal congestion, and rhinorrhea); and 5 prevention-related terms (hand washing, hand sanitizer, mask, social distance, and social isolation)

IR – Iran, IN – India, US – United States, UK – United Kingdom, IT – Italy, AU – Australia, BR – Brazil, FR – France, ZA – South Africa

Usage of GT in complex analytical models

There were 7 publications (Table 2) found where GT data was fed into long short-term memory [49,55] (4.5%), random forest regression [31,33] (4.5%), Adaboost algorithm [30] (2.3%), autoregressive integrated moving average (ARIMA), error, trend and seasonality (ERS) and feed-forward neural network autoregression (NNA) models [32] (2.3%) and vector error correction model [48] (2.3%) rather than analyzed itself. The findings of those studies showed that GT significantly improved the predictive capability of the methods used in the analysis and could be used in future with even higher predictability as more data becomes available [33,49].

Negative results of GT use for Covid-19 prediction and surveillance

7 publications (Table 3) showed negative results of GT use in Covid-19 surveillance and/or prediction. Most of them [34,36,39,51] stated that the correlations between GT search queries and Covid-19 cases in those countries were present because of media coverage [39,51] or announcements by governments and/or WHO [34,36]. A high variation in correlations between Covid-19 incidence and internet searches were identified as well [35,37], showing that GT data might not a reliable source for Covid-19 prediction and surveillance.

Table 3. Publications with negative results of GT use for COVID-19 prediction and surveillance

Author and year	The main findings about Google Trends	Country	Period	Keywords
Szmuda et al. (2020)	GT data did not correlate with Covid-19 incidence and mortality; however, they had a strong correlation with international WHO announcements.	40 European countries.	12 2019 – 04 2020	Coronavirus Appendix F.
Asseo et al. (2020)	The correlation between internet searches for symptoms and new Covid-19 cases varied significantly over time. High fluctuations show that relying only on GT data to monitor the spread of Covid-19 is not a viable strategy.	IT, US	03 2020 – 04 2020	taste loss, smell loss, sight loss (control), hearing loss (control), COVID symptoms (and the same in Italian)
Muselli et al. (2021)	The volume of Google searches did not reflect the actual epidemiological situation. It has been seen that official communications and government activity has more impact on public interest in the disease.	IT	12 2019 – 03 2020	coronavirus, coronavirus symptoms (in Italian), coronavirus news (in Italian), and coronavirus Italy (in Italian).
Rovetta (2021)	Big number of anomalies seen in multiple cities' relative search volumes (RSVs) made this data unusable for statistical inference. Furthermore, correlations varied greatly depending on the day RSVs were collected.	IT	02 2020 – 12 2020 and 02 2020 – 05 2020	coronavirus + covid
Satpathy et al. (2021)	Correlations found between GT queries and Covid-19 cases maybe either because of media-coverage-induced curiosity, or health-seeking curiosity.	IN	01 2020 – 05 2020	88 terms in Hindi and English. Appendix G.
Sato et al. (2021)	Results suggest that search keywords, previously identified as candidates for Covid-19 prediction, might be unreliable.	JP, AU, CA, UK, IE, IN, SG, US, ZA	10 2017 – 10 2020	54 English keywords and the corresponding 60 Japanese keywords. Appendix H.
Dagher et al. (2021)	Google searches for chilblain were influenced by media coverage and government policies during the Covid-19 pandemic, showing that GT, as a monitoring tool for emerging infectious diseases, should be used with caution.	US, UK, FR, IT, ES, DE	01 2020 – 05 2020	(1) toe or chilblains and (2) coronavirus,

IT – Italy, US – United States, IN – India, JP – Japan, AU – Australia, CA – Canada, UK – United Kingdom, IE – Ireland, SG – Singapore, ZA – South Africa, FR – France, ES – Spain, DE – Germany

4. DISCUSSION

Differences between countries

One possible reason why there were more studies done in high-income countries compared to low-income ones could be the lack of IT infrastructure – only 50% of individuals in low and

middle income countries use internet [56] as opposed to almost 90% in high-income countries [57], thus allowing people to search for information easily. For example, even though India is the second country in the world in internet user numbers, only 36% percent of its population uses internet monthly [58] as opposed to over 90% in USA [59,60] or 92% in Europe [61].

Time periods

It was seen that most of the publications with positive results were using data from first wave. Later the search volumes reduced [62] even though the incidence peaked. This could be explained by people's initial fear and lack of knowledge about the disease – symptoms, as well as information about the measures of protection was more searched during the first wave. Later such information became more widely known – not only people learned while searching themselves, but there were plenty of announcements from the governments as well as WHO. Naturally, people lost interest in following such news [63] in addition to getting “tired” of lockdowns.

The strong public interest decline in Covid-19 related issues might cause a big public health challenge to distribute relevant information regarding the newest developments in disease treatment and prevention measures throughout the whole pandemic [62].

Risk communication

Four publications [19,62,64,65] identified during the PubMed database search were not about prediction or surveillance of Covid-19 using GT data, rather about public interest in the pandemic and risk communication during the outbreaks. Those studies have shown increased amount of search queries after first case announcement [62,65] and such events like local Covid-19 transmission, approval and implementation of testing, social-distancing campaign, face mask shortage and announcements by WHO [64].

As people's interest peaked, it would be sensible to spread scientific information and promote preventative measures, as well as prevent misinformation in this exact time period. It would be beneficial to target social media, where misinformation spreads the fastest and people feel properly informed while reading non-expert opinions and statements. In addition, the decline in interest should be met with informational campaigns to ensure proper information spread [19] as well as showing people where to search for information and how to distinguish facts discovered by scientists, instead of relying on non-expert opinions.

Language

Our study reviews publications made in many different countries, which results in different search terms. Several studies [19] indicated the importance of ‘related query analysis’ prior to further analysis since it can point out the most relevant search terms.

Furthermore, there were many multi-country studies where the search terms were translated, thus potentially resulting in lost nuances of the meaning as well as some overlap [14].

Usage of GT data in models

Several studies (Table 2) incorporated GT data in their machine learning algorithms. Results of these studies show that such method was able to successfully predict an increase in Covid-19 cases in a very large number of countries 7 days in advance [31,33]. Furthermore, data of previous incidence of Covid-19 and GT were combined, which showed improved performance of the prediction models compared to previous ones which used incidence data alone [32].

Many studies have shown correlations between GT data and Covid-19 incidence (Table 1); however, machine learning models use large amounts of data and are able to not only show correlations, but also predict increases or decreases in cases over time. When conventional metrics (numbers of cases and deaths) were combined with interest-over-time values, the prediction ability of the models increased further [32]. Rabiolo et al. have identified two principal components, which allowed to reduce data dimensionality and summarise the information into two components, thus providing a flexible approach which allows to change the variables of interest and use the same models to investigate different research questions in the future [32]. Moreover, an additional advantage is that the performance of these models can be further improved as more data becomes available over time and can reflect the current situation [33,49]. In addition, the models could have other uses than predicting Covid-19 numbers, e.g. assessing people's awareness and engagement, thus allowing health authorities to use these data for measuring the effectiveness of the information spread [49], which is crucial especially when information fatigue is present [63,66].

Negative findings

Few studies showed that GT data could not be used for Covid-19 prediction and/or surveillance. According to the authors of those studies, WHO and/or local government announcements had a major influence on search trends [34,51] and that GT is more efficient in tracking a new disease outbreak when media coverage of that disease is absent [34]. However, this was not possible to test during this pandemic since WHO, as well as governments and officials started communication regarding the novel coronavirus even before WHO announced it as pandemic. Furthermore, the authors suggest that online searches simply overlap with an increase in Covid-19 cases and related deaths since big media announcements are made at the same time as increase in incidence happens [34] or were a result of information-seeking curiosity [51].

Strengths

The authors of the studies included in this review have indicated many strengths of the methods proposed for prediction and surveillance of Covid-19 disease throughout the world. Many studies have made analyses in multiple countries and got the same results for the majority of them [8,14,18,19,54], thus showing the robustness of the chosen methods. Furthermore, Google search data is easy to obtain, is more dynamic and available compared to traditional data sources, such as data from governmental institutions, health authorities etc. as well as represents current moods of the population and can be obtained during multiple periods [49]. As the relevant search terms can change over time, it is possible to investigate GT data over and over again and incorporate the new terms and newly available data into the prediction models, thus improving the outcome. Even more improvement in prediction can be reached when search terms with higher correlation values are used for the analysis [55].

Limitations

One of the main limitations mentioned by the study authors was the short timeframe taken for the analysis [19,33]. The positive results obtained from the first Covid-19 wave could have been due to the virus being new and interesting to the society, which resulted in an increase in searches using Google and other search engines. Furthermore, such methods must account for misspellings and possible other search terms [19] as well the fact that Google might not be the main search engine for different groups of people [19,32–34]. One more disadvantage lies in the data (incidence and death rate) which is used to compare it with the one obtained from Google. Different countries have different testing policies, as well as death reports, thus making it impossible to have a standardised number [19,34]. Moreover, Covid-19 reports in other countries and media coverage everywhere around the world, as well as people's curiosity might have influenced the increase in searches [14,33]. What also should be taken into account, however was not possible given the data, are the demographics (gender, age, education level, literacy) of the searchers [14,34]. One could speculate that older people are not represented in the search volumes, even though they are the most Covid-19 affected group. They, together with children, as well as people living in areas with poor internet connection, cannot be studied with this strategy, i.e. using GT data to make predictions, thus making it implausible for countries with large rural areas [32]. Also, the symptom similarity and prevention methods between Covid-19 and influenza might not allow to differentiate between the two [33,51], potentially showing higher search volumes and influencing the predictions.

5. CONCLUSIONS

The majority of the studies analysed in this paper have reported positive findings regarding prediction and surveillance of Covid-19 spread using data obtained from Google Trends. Many studies showed positive results during the first wave of the pandemic and if these data are incorporated in various complex algorithms, their prediction capabilities could be significantly improved. Furthermore, one could speculate that GT forecasting capabilities would possibly increase in the future when mass media calms down. Further analyses using data obtained during later time periods are needed to further evaluate the forecasting capabilities of GT.

REFERENCES:

1. WHO. Coronavirus [Internet]. [cited 2021 Jul 6]. Available from: https://www.who.int/health-topics/coronavirus#tab=tab_1
2. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. *Lancet* (London, England). 2020 Feb 15;395(10223):470–3.
3. Tang D, Tou J, Wang J, Chen Q, Wang W, Huang J, et al. Prevention and control strategies for emergency, limited-term, and elective operations in pediatric surgery during the epidemic period of COVID-19. *World J Pediatr Surg*. 2020 Mar 26;3(1).
4. Heymann DL, Shindo N. COVID-19: what is next for public health? *Lancet* (London, England). 2020 Feb 22;395(10224):542–5.
5. Cousins HC, Cousins CC, Harris A, Pasquale LR. Regional Infection of COVID-19 Case Rates: Analysis of Search-Engine Query Patterns. *J Med Internet Res*. 2020 Jul 1;22(7).
6. 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 Heal Surveill*. 2016 Jun 1;2(1):e30.
7. Kurian SJ, Bhatti A, Alvi MA, Ting HH, Storlie C, Wilson PM, et al. Correlations Between COVID-19 Cases and Google Trends Data in the United States: A State-by-State Analysis. *Mayo Clin Proc*. 2020 Nov 1;95(11):2370.
8. Effenberger M, Kronbichler A, Shin J, Mayer G, Tilg H, Perco P. Association of the COVID-19 pandemic with Internet Search Volumes: A Google Trends™ Analysis. *Int J Infect Dis*. 2020 Jun 1;95:192.
9. Schneider PP, van Goll CJ, Spreuwenberg P, Hooiveld M, Donker GA, Barnett DJ, Paget J. Using web search queries to monitor influenza-like illness: an exploratory retrospective analysis, Netherlands, 2017/18 influenza season. *Euro Surveill*. 2020 May 28;25(21).
10. Strauss R, Lorenz E, Kristensen K, Eibach D, Torres J, May J, Castro J. Investigating the utility of Google trends for Zika and Chikungunya surveillance in Venezuela. *BMC Public Health*. 2020 Jun 16;20(1).
11. Shin SY, Seo DW, An J, Kwak H, Kim SH, Gwack J, Jo MW. High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea. *Sci Rep*. 2016 Sep 6;6.
12. Ocampo AJ, Chunara R, Brownstein JS. Using search queries for malaria surveillance,

- Thailand. *Malar J*. 2013;12(1).
13. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009 Jul;6(7).
 14. Walker A, Hopkins C, Surda P. Use of Google Trends to investigate loss-of-smell-related searches during the COVID-19 outbreak. *Int Forum Allergy Rhinol*. 2020 Jul 1;10(7):839–47.
 15. Mavragani A. Tracking COVID-19 in Europe: Infodemiology Approach. *JMIR public Heal Surveill*. 2020 Apr 1;6(2).
 16. Higgins TS, Wu AW, Sharma D, Illing EA, Rubel K, Ting JY, et al. Correlations of Online Search Engine Trends With Coronavirus Disease (COVID-19) Incidence: Infodemiology Study. *JMIR public Heal Surveill*. 2020 Apr 1;6(2).
 17. Cherry G, Roche J, Chu M, Liu J, Lechner M, Lund VJ, et al. Loss of smell and taste: a new marker of COVID-19? Tracking reduced sense of smell during the coronavirus pandemic using search trends. *Expert Rev Anti Infect Ther*. 2020;18(11):1165–70.
 18. Sharma M, Sharma S. The Rising Number of COVID-19 Cases Reflecting Growing Search Trend and Concern of People: A Google Trend Analysis of Eight Major Countries. *J Med Syst*. 2020 Jul 1;44(7).
 19. Schnoell J, Besser G, Jank BJ, Bartosik TJ, Parzefall T, Riss D, et al. The association between COVID-19 cases and deaths and web-based public inquiries. *Infect Dis (London, England)*. 2021;53(3):176–83.
 20. Jimenez AJ, Estevez-Reboredo RM, Santed MA, Ramos V. COVID-19 Symptom-Related Google Searches and Local COVID-19 Incidence in Spain: Correlational Study. *J Med Internet Res*. 2020 Dec 1;22(12).
 21. Lippi G, Mattiuzzi C, Cervellin G. Google search volume predicts the emergence of COVID-19 outbreaks. *Acta Biomed*. 2020 Sep 11;91(3):1–5.
 22. Strzelecki A, Azevedo A, Albuquerque A. Correlation between the Spread of COVID-19 and the Interest in Personal Protective Measures in Poland and Portugal. *Healthc (Basel, Switzerland)*. 2020 Sep 1;8(3).
 23. Badell-Grau RA, Cuff JP, Kelly BP, Waller-Evans H, Lloyd-Evans E. Investigating the Prevalence of Reactive Online Searching in the COVID-19 Pandemic: Infodemiology Study. *J Med Internet Res*. 2020 Oct 1;22(10).
 24. Brunori P, Resce G. Dipartimento di Economia e Finanza Searching for the peak Google Trends and the Covid-19 outbreak in Italy.
 25. Sulyok M, Ferenci T, Walker M. Google Trends Data and COVID-19 in Europe:

- Correlations and model enhancement are European wide. *Transbound Emerg Dis.* 2021 Jul 1;68(4):2610–5.
26. Pellegrini M, Ferrucci E, Guaraldi F, Bernabei F, Scordia V, Giannaccare G. Emerging application of Google Trends searches on “conjunctivitis” for tracing the course of COVID-19 pandemic. *Eur J Ophthalmol.* 2021;
 27. Cinarka H, Uysal MA, Cifter A, Niksarlioglu EY, Çarkoğlu A. The relationship between Google search interest for pulmonary symptoms and COVID-19 cases using dynamic conditional correlation analysis. *Sci Rep.* 2021 Dec 1;11(1).
 28. Kristensen K, Lorenz E, May J, Strauss R. Exploring the use of web searches for risk communication during COVID-19 in Germany. *Sci Rep.* 2021 Dec 1;11(1).
 29. Prasanth S, Singh U, Kumar A, Tikkiwal VA, Chong PHJ. Forecasting spread of COVID-19 using google trends: A hybrid GWO-deep learning approach. *Chaos Solitons Fractals.* 2021 Jan 1;142:110336.
 30. Niu B, Liang R, Zhang S, Zhang H, Qu X, Su Q, et al. Epidemic analysis of COVID-19 in Italy based on spatiotemporal geographic information and Google Trends. *Transbound Emerg Dis.* 2021 Jul 1;68(4):2384–400.
 31. Peng Y, Li C, Rong Y, Chen X, Chen H. Retrospective analysis of the accuracy of predicting the alert level of COVID-19 in 202 countries using Google Trends and machine learning. *J Glob Health.* 2020 Dec 1;10(2):1–8.
 32. Rabiolo A, Alladio E, Morales E, McNaught AI, Bandello F, Afifi AA, et al. Forecasting the COVID-19 Epidemic by Integrating Symptom Search Behavior Into Predictive Models: Infoveillance Study. *J Med Internet Res.* 2021 Aug 1;23(8).
 33. Peng Y, Li C, Rong Y, Pang CP, Chen X, Chen H. Real-time Prediction of the Daily Incidence of COVID-19 in 215 Countries and Territories Using Machine Learning: Model Development and Validation. *J Med Internet Res.* 2021 Jun 1;23(6).
 34. Szmuda T, Ali S, Hetzger TV, Rosvall P, Słoniewski P. Are online searches for the novel coronavirus (COVID-19) related to media or epidemiology? A cross-sectional study. *Int J Infect Dis.* 2020 Aug 1;97:386–90.
 35. Asseo K, Fierro F, Slavutsky Y, Frasnelli J, Niv MY. Tracking COVID-19 using taste and smell loss Google searches is not a reliable strategy. *Sci Rep.* 2020 Dec 1;10(1).
 36. Muselli M, Cofini V, Desideri G, Necozone S. Coronavirus (Covid-19) pandemic: How may communication strategies influence our behaviours? *Int J disaster risk Reduct IJDRR.* 2021 Feb 1;53.
 37. Rovetta A. Reliability of Google Trends: Analysis of the Limits and Potential of Web

- Infoveillance During COVID-19 Pandemic and for Future Research. *Front Res metrics Anal.* 2021 May 25;6.
38. Sato K, Mano T, Iwata A, Toda T. Need of care in interpreting Google Trends-based COVID-19 infodemiological study results: potential risk of false-positivity. *BMC Med Res Methodol.* 2021 Dec 1;21(1).
 39. Dagher SH, Lamé G, Hubiche T, Ezzedine K, Duong TA. The Influence of Media Coverage and Governmental Policies on Google Queries Related to COVID-19 Cutaneous Symptoms: Infodemiology Study. *JMIR public Heal Surveill.* 2021 Feb 1;7(2).
 40. Panuganti BA, Jafari A, MacDonald B, DeConde AS. Predicting COVID-19 Incidence Using Anosmia and Other COVID-19 Symptomatology: Preliminary Analysis Using Google and Twitter. *Otolaryngol Head Neck Surg.* 2020 Sep 1;163(3):491–7.
 41. Mavragani A, Gkillas K. COVID-19 predictability in the United States using Google Trends time series. *Sci Rep.* 2020 Dec 1;10(1).
 42. 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 Nov 1;18(12):2833–2834.e3.
 43. Rajan A, Sharaf R, Brown RS, Sharaiha RZ, Lebwohl B, Mahadev SH. Association of Search Query Interest in Gastrointestinal Symptoms With COVID-19 Diagnosis in the United States: Infodemiology Study. *JMIR public Heal Surveill.* 2020 Jul 1;6(3).
 44. Hartwell M, Greiner B, Kilburn Z, Ottwell R. Association of Public Interest in Preventive Measures and Increased COVID-19 Cases After the Expiration of Stay-at-Home Orders: A Cross-Sectional Study. *Disaster Med Public Health Prep.* 2020;
 45. Abbas M, Morland TB, Hall ES, El-Manzalawy Y. Associations between Google Search Trends for Symptoms and COVID-19 Confirmed and Death Cases in the United States. *Int J Environ Res Public Health.* 2021 May 1;18(9).
 46. Yousefinaghani S, Dara R, Mubareka S, Sharif S. Prediction of COVID-19 Waves Using Social Media and Google Search: A Case Study of the US and Canada. *Front public Heal.* 2021 Apr 16;9.
 47. Husnayain A, Chuang TW, Fuad A, Su ECY. High variability in model performance of Google relative search volumes in spatially clustered COVID-19 areas of the USA. *Int J Infect Dis.* 2021 Aug 1;109:269–78.
 48. Turk PJ, Tran TP, Rose GA, McWilliams A. A predictive internet-based model for COVID-19 hospitalization census. *Sci Rep.* 2021 Dec 1;11(1).

49. Ayyoubzadeh SM, Ayyoubzadeh SM, Zahedi H, Ahmadi M, Niakan Kalhori SR. Predicting COVID-19 Incidence Through Analysis of Google Trends Data in Iran: Data Mining and Deep Learning Pilot Study. *JMIR public Heal Surveill.* 2020 Apr 1;6(2).
50. Venkatesh U, Gandhi PA. Prediction of COVID-19 Outbreaks Using Google Trends in India: A Retrospective Analysis. *Healthc Inform Res.* 2020 Jun 1;26(3):175–84.
51. Satpathy P, Kumar S, Prasad P. Suitability of Google Trends™ for Digital Surveillance During Ongoing COVID-19 Epidemic: A Case Study from India. *Disaster Med Public Health Prep.* 2021;
52. Xie T, Tan T, Li J. An Extensive Search Trends-Based Analysis of Public Attention on Social Media in the Early Outbreak of COVID-19 in China. *Risk Manag Healthc Policy.* 2020;13:1353–64.
53. Husnayain A, Fuad A, Su ECY. Applications of Google Search Trends for risk communication in infectious disease management: A case study of the COVID-19 outbreak in Taiwan. *Int J Infect Dis.* 2020 Jun 1;95:221–3.
54. Lin YH, Liu CH, Chiu YC. Google searches for the keywords of “wash hands” predict the speed of national spread of COVID-19 outbreak among 21 countries. *Brain Behav Immun.* 2020 Jul 1;87:30.
55. Prasanth S, Singh U, Kumar A, Tikkiwal VA, Chong PHJ. Forecasting spread of COVID-19 using google trends: A hybrid GWO-deep learning approach. *Chaos Solitons Fractals.* 2021 Jan 1;142.
56. Individuals using the Internet (% of population) - Low & middle income | Data [Internet]. [cited 2022 Jan 30]. Available from: <https://data.worldbank.org/indicator/IT.NET.USER.ZS?locations=XO>
57. Individuals using the Internet (% of population) - High income | Data [Internet]. [cited 2022 Jan 30]. Available from: <https://data.worldbank.org/indicator/IT.NET.USER.ZS?locations=XD>
58. India has second highest number of Internet users after China: Report - The Economic Times [Internet]. [cited 2022 Jan 30]. Available from: <https://economictimes.indiatimes.com/tech/internet/india-has-second-highest-number-of-internet-users-after-china-report/articleshow/71311705.cms?from=mdr>
59. United States Population (2022) - Worldometer [Internet]. [cited 2022 Jan 30]. Available from: <https://www.worldometers.info/world-population/us-population/>
60. • U.S: Number of internet users 2025 | Statista [Internet]. [cited 2022 Jan 30]. Available from: <https://www.statista.com/statistics/325645/usa-number-of-internet-users/>

61. Digital economy and society statistics - households and individuals - Statistics Explained [Internet]. [cited 2022 Jan 30]. Available from: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Digital_economy_and_society_statistics_-_households_and_individuals
62. Schuster B, Tizek L, Schielein MC, Ziehfrend S, Rothe K, Spinner CD, et al. Retracing the COVID-19 Pandemic in Germany from a Public Perspective using Google Search Queries Related to “coronavirus.” *Gesundheitswesen*. 2021 May 1;83(5):E9–14.
63. Buneviciene I, Bunevicius R, Bagdonas S, Bunevicius A. COVID-19 media fatigue: predictors of decreasing interest and avoidance of COVID-19–related news. *Public Health*. 2021 Jul 1;196:124–8.
64. Husnayain A, Shim E, Fuad A, Su ECY. Understanding the Community Risk Perceptions of the COVID-19 Outbreak in South Korea: Infodemiology Study. *J Med Internet Res*. 2020 Sep 29;22(9).
65. Bento AI, Nguyen T, Wing C, Lozano-Rojas F, Ahn YY, Simon K. Evidence from internet search data shows information-seeking responses to news of local COVID-19 cases. *Proc Natl Acad Sci U S A*. 2020 May 26;117(21).
66. Khan T, Khan R, Fatima S, Sunesara S, Masud F, Ratnani I. TP011 Use Of Health Services And Medical Education Research To Evaluate The Effect Of Covid-19 / Thematic Poster Session COVID-19 News Fatigue: Are People Tired of Hearing About the Pandemic?

APPENDICES

Appendix A

The screening library of 463 search queries was obtained using the GT “Related Queries” function on an initial bank of 23 coronavirus-related terms. Seed Terms: 'am i sick', 'cdc', 'corona virus near me', 'coronavirus doctor', 'coronavirus help', 'coronavirus hospital', 'coronavirus symptoms', 'coronavirus testing', 'coronavirus treatment', 'coronavirus vs flu symptoms', 'cough', 'covid 19 symptoms', 'covid 19 testing', 'do i have coronavirus', 'doctor near me', 'insurance coronavirus', 'sick', 'sore throat', 'symptoms coronavirus 2020', 'testing'.

Appendix B

The authors used 422 time-series datasets extracted from the Google COVID-19 Search Trends Symptoms dataset, which was released by Google on 2 September 2020 and is available at: <https://github.com/google-research/open-covid-19-data/>. The study authors accessed the dataset 6 November 2020.

Appendix C

Keywords used in the correlation analysis: COVID-19, Novel coronavirus, Coronavirus, SARS, COVID-19 data, COVID-19 Italy, SARS-CoV-2, COVID-19 symptoms, Influenza, Pneumonia, Fever, Cough, Sore throat, Chest distress, Difficulty breathing, Fatigue, Fell sick and vomit, Diarrhoea, Muscle ache, Mask, Disinfect, Isolation, Protective suit, Goggle, Thermometer, Disposable gloves, Medical supplies, COVID-19 vaccine, N95, Confirmed cases, New cases, Suspected case, Infection, Epidemic, Fatality rate, ISS, WHO, Incubation period, Hospital, Nurse.

Appendix D

202 countries: Argentina, Australia, Austria, Belgium, Brazil, Finland, France, Germany, India, Indonesia, Iran (Islamic Republic of), Ireland, Italy, Peru, Poland, Puerto Rico, South Africa, Spain, Switzerland, United States of America, Aruba, Central African Republic, French Polynesia, Ghana, Venezuela (Bolivarian Republic of), Canada, Chad, Colombia, Costa Rica, Coted Ivoire, Greece, Guadeloupe, Iceland, Kuwait, Morocco, Netherlands, Panama, Republic of Moldova, Rwanda, The United Kingdom, United States Virgin Islands, Uruguay, Uzbekistan, Albania, Anguilla, Antigua and Barbuda, Bahrain, Bolivia (Plurinational State of), Botswana, Bulgaria, Burundi, Cameroon, Chile, Croatia, Cuba, Cyprus, Dominican Republic, Eritrea, Eswatini, Falkland Islands (Malvinas), French Guiana, Gambia, Grenada, Honduras, Kazakhstan, Kenya, Kyrgyzstan, Lebanon, Luxembourg, Malta, Martinique, Mauritania, Montenegro, Nigeria, Norway, Oman, Portugal, Qatar, Reunion, Saint Kitts and Nevis, Saint Martin, Serbia, Sierra Leone, Slovakia, Slovenia, Ukraine, Algeria, Armenia, Azerbaijan,

Barbados, Belarus, Burkina Faso, Cayman Islands, Curacao, Czechia, Denmark, Ecuador, Egypt, El Salvador, Equatorial Guinea, Estonia, Guatemala, Guinea, Guinea-Bissau, Hungary, Jordan, Kosovo, Latvia, Lithuania, Malawi, Mali, Mauritius, Mexico, New Caledonia, Niger, Paraguay, Saint Barthelemy, Saint Vincent and the Grenadines, Senegal, Sint Maarten, Sweden, Togo, Tunisia, Turkey, Saudi Arabia, Bahamas, Bangladesh, Benin, Bermuda, Bhutan, Djibouti, Gibraltar, Guernsey, Guyana, Haiti, Jersey, Liberia, Libya, Madagascar, Montserrat, Mozambique, North Macedonia, Philippines, Romania, Russian Federation, Seychelles, Somalia, South Sudan, Turks and Caicos Islands, Uganda, United Republic of Tanzania, Afghanistan, Pakistan, Sudan, Bosnia and Herzegovina, Fiji, Georgia, Greenland, Guam, Isle of Man, Jamaica, Japan, Liechtenstein, Myanmar, Nepal, Papua New Guinea, San Marino, Singapore, Sri Lanka, Suriname, United Arab Emirates, Zambia, Zimbabwe, Andorra, Belize, Ethiopia, Gabon, Malaysia, Mayotte, Mongolia, Nicaragua, Saint Lucia, Sao Tome and Principe, Timor-Leste, Yemen, Angola, British Virgin Islands, Dominica, New Zealand, Thailand, Trinidad and Tobago, Brunei Darussalam, Cambodia, Faroe Islands, Iraq, Israel, Maldives, Northern Mariana Islands (Commonwealth of the), Syrian Arab Republic, China, Vietnam, Laos.

Appendix E

215 countries and territories: Afghanistan, Albania, Algeria, Andorra, Anguilla, Argentina, Armenia, Australia, Austria, Azerbaijan, Bahrain, Bangladesh, Belarus, Belgium, Bolivia (Plurinational State of), Bonaire, Sint Eustatius and Saba, Bosnia and Herzegovina, Brazil, British Virgin Islands, Brunei Darussalam, Bulgaria, Burkina Faso, Cabo Verde, Canada, Chad, Chile, China, Colombia, Congo, Costa Rica, Coted Ivoire, Croatia, Cuba, Cyprus, Czechia, Democratic Republic of the Congo, Denmark, Djibouti, Dominica, Dominican Republic, Ecuador, Egypt, El Salvador, Estonia, Eswatini, Ethiopia, Falkland Islands (Malvinas), Fiji, Finland, France, French Guiana, Georgia, Germany, Ghana, Greece, Greenland, Grenada, Guam, Guatemala, Guinea, Haiti, Honduras, Hungary, Iceland, India, Indonesia, Iran (Islamic Republic of), Iraq, Ireland, Italy, Jamaica, Japan, Jordan, Kazakhstan, Kenya, Kosovo, Kuwait, Laos, Latvia, Liberia, Lithuania, Luxembourg, Madagascar, Malawi, Malaysia, Mali, Mauritania, Mayotte, Mexico, Montenegro, Montserrat, Morocco, Mozambique, Myanmar, Namibia, Nepal, Netherlands, New Zealand, Niger, Nigeria, North Macedonia, Northern Mariana Islands (Commonwealth of the), Norway, occupied Palestinian territory, Oman, Pakistan, Panama, Paraguay, Peru, Philippines, Poland, Portugal, Puerto Rico, Qatar, Republic of Korea, Republic of Moldova, Reunion, Romania, Russian Federation, Saint Kitts and Nevis, Saint Lucia, Saint Pierre and Miquelon, San Marino, Saudi Arabia, Senegal,

Serbia, Sierra Leone, Singapore, Slovakia, Slovenia, Somalia, South Africa, Spain, Sri Lanka, Suriname, Sweden, Switzerland, Tajikistan, The United Kingdom, Togo, Tunisia, Turkey, Uganda, Ukraine, United Arab Emirates, United States of America, Uruguay, Uzbekistan, Venezuela (Bolivarian Republic of), Yemen, Antigua and Barbuda, Bahamas, Bermuda, Burundi, Cambodia, Cameroon, Cayman Islands, Central African Republic, Comoros, Equatorial Guinea, Eritrea, Gabon, Gibraltar, Guernsey, Guinea-Bissau, Isle of Man, Israel, Jersey, Lebanon, Lesotho, Liechtenstein, Maldives, Mauritius, Monaco, Mongolia, New Caledonia, Nicaragua, Saint Vincent and the Grenadines, Sao Tome and Principe, Seychelles, South Sudan, Sudan, Syrian Arab Republic, Thailand, Timor-Leste, United Republic of Tanzania, United States Virgin Islands, Zambia, Angola, Barbados, Benin, Bhutan, Curacao, Libya, Malta, Rwanda, Trinidad and Tobago, Viet Nam, Zimbabwe, Belize, Botswana, Gambia, Guadeloupe, Martinique, Papua New Guinea, Saint Barthelemy, Faroe Islands, French Polynesia, Guyana, Kyrgyzstan, Sint Maarten, Turks and Caicos Islands, Aruba, Holy See, Saint Martin.

Appendix F

40 countries: Albania, Andorra, Austria, Belarus, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Holy See, Hungary, Iceland, Ireland, Italy, Latvia, Liechtenstein, Lithuania, Luxembourg, Malta, Moldova, Monaco, Montenegro, Netherlands, North Macedonia, Norway, Poland, Portugal, Romania, Russia, San Marino, Serbia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Ukraine, United Kingdom. Holy See, Liechtenstein, Monaco and San Marino were excluded due to being small and therefore having insufficient data from Google Trends.

Appendix G

88 search terms: Coronavirus, Corona, Covid 19, Covid, SARS-CoV-2, SARS Novel coronavirus, Novel corona, Virus, Infection, Disease, Social distancing, Hand wash, Hand rub, Mask, Facemask, Sanitizer, Soap, Fever, Cough, Cold, Breathlessness, Fatigue, Rhinorrhoea, Nasal congestion, Sneeze, Myalgia, Sore throat, Diarrhoea, Anorexia, Chest pain, Headache, Nausea, Ageusia, Abdominal pain, Dizziness, Vomiting, Eye pain, Anosmia, Doctor, Nurse, Hospital, Clinic, Medicine, Check-up, OPD, Treatment, Testing, Lockdown, Quarantine, Isolation, Bhilwada model, Curfew, Diya, Thali, Warrior, Shop, Market, Open, Ticket, Rail, Bus, Modi, PM cares, 20 lakh, Kerala, Mumbai, *Khansi* (Cough), *Bukhar* (Fever), *Dawa* (Medicine), *Dawai* (Medicine), *Kharash* (Sore throat), *Sans* (Breathlessness), *Sardi* (Cold), *Jukam* (Cold), *Dama* (Asthma), coronavirus, Corona, Covid 19, Covid, Crona,

Lockdown, Social distancing, *hath dhona* (Hand wash), *Mukhauta* (Mask), *Sabun* (Soap), *Deepak* (Lamp), *Thali* (Plate).

Appendix H

54 English keywords: Malaise, fatigue, tired, Anorexia, diarrhea, constipation, Abdominal pain, stomach ache, nausea, Chest pain, dyspnea, vomiting, Shortness of breath, short of breath, pneumonia, Cough, sputum, rhinitis, Runny nose, nasal discharge, stuffy nose, Sneezes, sore throat, throat pain, Fever, chills, cold, Sense of smell, loss of smell, anosmia, Sense of taste, loss of taste, dysgeusia, Hair loss, loss of hair, bald, Myalgia, muscle pain, body aches, Arthralgia, joint pain, pain, Eye pain, sore, congestion, Headache, memory loss, confusion, Vertigo, dizziness, dizzy, Insomnia, anxiety, numbness. Corresponding Japanese keywords can be found at:

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8286439/bin/12874_2021_1338_MOESM1_ESM.docx.