

RESEARCH

Open Access



Need of care in interpreting Google Trends-based COVID-19 infodemiological study results: potential risk of false-positivity

Kenichiro Sato^{1*}, Tatsuo Mano¹, Atsushi Iwata^{1,2*} and Tatsushi Toda¹

Abstract

Background: Google Trends (GT) is being used as an epidemiological tool to study coronavirus disease (COVID-19) by identifying keywords in search trends that are predictive for the COVID-19 epidemiological burden. However, many of the earlier GT-based studies include potential statistical fallacies by measuring the correlation between non-stationary time sequences without adjusting for multiple comparisons or the confounding of media coverage, leading to concerns about the increased risk of obtaining false-positive results. In this study, we aimed to apply statistically more favorable methods to validate the earlier GT-based COVID-19 study results.

Methods: We extracted the relative GT search volume for keywords associated with COVID-19 symptoms, and evaluated their Granger-causality to weekly COVID-19 positivity in eight English-speaking countries and Japan. In addition, the impact of media coverage on keywords with significant Granger-causality was further evaluated using Japanese regional data.

Results: Our Granger causality-based approach largely decreased (by up to approximately one-third) the number of keywords identified as having a significant temporal relationship with the COVID-19 trend when compared to those identified by Pearson or Spearman's rank correlation-based approach. "Sense of smell" and "loss of smell" were the most reliable GT keywords across all the evaluated countries; however, when adjusted with their media coverage, these keyword trends did not Granger-cause the COVID-19 positivity trends (in Japan).

Conclusions: Our results suggest that some of the search keywords reported as candidate predictive measures in earlier GT-based COVID-19 studies may potentially be unreliable; therefore, caution is necessary when interpreting published GT-based study results.

Keywords: COVID-19, Google Trends, Infodemiology, Vector autoregression model, Granger causality

Background

Google Trends (GT) is a publicly available source of online Google search trafficking data (<https://trends.google.co.jp/trends>), which allows users to visualize changes in time series related to the general public's

online interest in certain keywords. It is used as one of the "infodemiology" tools [1] to study epidemiological trends of certain disease outbreaks such as the Middle East Respiratory Syndrome epidemic and the Ebola outbreak [1]. As for coronavirus disease (COVID-19) that became a worldwide pandemic in early 2020 [2, 3], the potential use of GT to predict COVID-19 cases or deaths has been reported with regard to GT trends and keyword searches of "COVID-19" [4, 5] or any of its symptoms, including chest pain, anosmia, dysgeusia, headache,

*Correspondence: kenisatou-tky@umin.ac.jp; iwata@m.u-tokyo.ac.jp

¹ Department of Neurology, Graduate School of Medicine, University of Tokyo, Tokyo, Japan

² Department of Neurology, Tokyo Metropolitan Geriatric Medical Center Hospital, Tokyo, Japan



shortness of breath, etc. [6–8] within the initial months following the outbreak [5–10].

In many earlier studies analyzing GT trend data as an epidemiological tool, with a few exceptions [11–13], analytical fallacies were of concern. First, Pearson (or Spearman's rank) correlation is often applied to assess the correlation between the time-series trends of COVID-19 cases/deaths and GT trends in symptom keywords without confirming the stationarity of these time series. This is sometimes critically inappropriate in the context of time-series analyses because time-series data often contains unit-root and the correlation between such series often results in high coefficient value and t-statistics [14], and thus it can increase the likelihood of obtaining spurious correlations. Second, the Pearson/Spearman correlation tests were repeated for each of the included symptom keywords (e.g., fever, cough, pneumonia, anosmia, sore throat, headache, etc. [8]) without adequate adjustment for multiple comparisons, which would also increase the risk of false-positive results. Third, because COVID-19 and its symptoms have attracted intensive attention worldwide, the influence of media coverage on GT symptom keywords is inevitable [10, 15, 16], which has hardly been adjusted in a statistically favorable manner.

Based on the above analytical concerns for earlier studies, by using the vector autoregression (VAR) model [11–13], which is designed to deal with time-series data and is robust against weakness as observed in case of using correlation, we aim to identify statistically more reliable symptom keywords for which GT trends may be used as a predictive measure for future COVID-19 positivity trends, and to validate the earlier study results.

Methods

Extracting Google Trends and COVID-19 data

All the following data handling and analyses were performed using R 3.5.2 (R Foundation for Statistical Computing, Vienna, Austria). A statistical level of less than 0.05 is considered significant if not stated otherwise. COVID-19 data and Google Trends (GT) data were separately analyzed in nine different regions: Japan (JP) and eight English-speaking countries, namely, Australia (AU), Canada (CA), Great Britain (GB), Ireland (IE), India (IN), Singapore (SG), United States (US), and South Africa (ZA).

The 3-year (October 1, 2017–October 25, 2020) time series GT trend data of 'all categories' for keywords of symptoms that may be related to COVID-19 was queried using R package *gtrendsR* [17]. Individual queries were separately conducted for each keyword in all nine regions. Search keywords were defined as listed in Table 1: 54 English keywords were used for search in

Table 1 Included English and Japanese keywords search for Google Trends

English keywords searched in Google Trends		
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
Sneeze	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

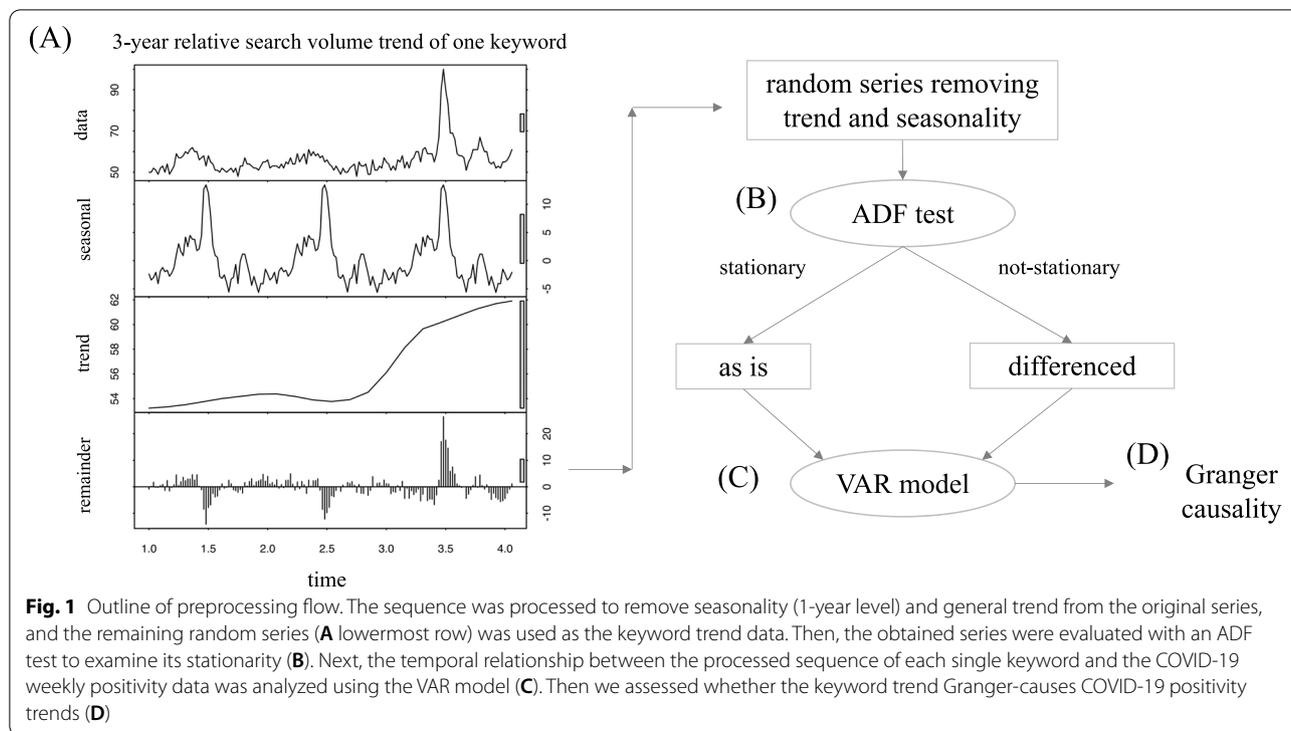
Search keywords were arbitrarily defined: 54 English keywords were used for search in 8 English-speaking country regions, and the corresponding 60 Japanese keywords (as listed in Additional file 1) were used for search in the Japanese region

eight English-speaking country regions, and the corresponding 60 Japanese keywords (as listed in Additional file 1) were used for searches in the Japan region. The obtained data were the weekly relative search volume for each keyword, of which the maximum value during the included period was normalized to 100%. For the timings when the relative search volume was less than 1%, we imputed them as 0%.

For COVID-19 data on serial daily number of positive cases from January 22, 2020, we downloaded data from the web database (<https://data.humdata.org/dataset/novel-coronavirus-2019-ncov-cases>, accessed on October 30, 2020) provided by the United Nations Office for Coordination of Humanitarian Affairs. Since we did not include the number of positive cases from mainland China, we imputed the number of COVID-19 cases before January 22, 2020 as zero (even for 2017–2019). The COVID-19 daily case data were converted to weekly serial data, in reference to the above GT weekly trend data.

Preprocessing and analysis

The keyword weekly trend data were further processed as shown in Fig. 1. Figure 1A (uppermost row) is the 3-year original GT time series for "chest pain" in the United States region. The sequence was processed using R



package *stats* to remove seasonality (1-year level) and the general trend from the original series, and the remaining random series (Fig. 1A, lowermost row) was used as the keyword trend data to analyze [11]. Then, the obtained series were evaluated with an augmented Dickey–Fuller (ADF) test using R package *tseries* [18] to examine whether the sequence was stationary (Fig. 1B). If the series was not considered stationary, the sequence was further differenced so that the differenced series became stationary (as confirmed by the ADF test again).

Next, the temporal relationship between the processed sequence for each keyword and the COVID-19 weekly positivity data was analyzed with the VAR model [11, 12] (Fig. 1C), using R package *vars* [19]. Since the COVID-19 weekly positivity trend data was actually not stationary by itself, its difference sequence was imputed to the VAR analysis. The adequate lag was determined from the lag order range of 1–8, based on the Akaike’s information criterion which is one of the most frequently used methods [20]. We used the lag range of 1–8 weeks because more than 2 months of lag to predict COVID-19 positivity by the keyword trend might be virtually too long, considering the time span of increase / decrease in the effective reproduction number of COVID-19 as a reference of disease momentum [21, 22]. The following equations (A–B) describe an example of VAR model (of which lag order = 1) used in this study:

$$\begin{aligned}
 \text{A) } Y_{1,t} &= c_1 + (\phi_{11}Y_{1,t-1} + \phi_{12}Y_{2,t-1}) + \varepsilon_{1,t} \\
 \text{B) } Y_{2,t} &= c_2 + (\phi_{21}Y_{1,t-1} + \phi_{22}Y_{2,t-1}) + \varepsilon_{2,t}
 \end{aligned}$$

where Y_1 is the weekly COVID-19 positivity in each country, and Y_2 is the weekly Google trend in relative search volume for one keyword of interest in the same country. Thus, the VAR models are obtained for each of all keywords in each country.

Then, using the obtained VAR model, we assessed whether the keyword’s trend Granger-caused the COVID-19 positivity trends [11, 12] (Fig. 1D). This implied that the change in the keyword trend could have the potential to practically predict the near-future change in the COVID-19 positivity trend. The causality here was merely a statistical one and did not require true causal mechanisms between the two trends. One p -value was obtained for the Granger-causality of one keyword to the COVID-19 trend and the Granger-causality analysis was performed for all the keywords. We adjusted multiple testing using the Benjamini-Hochberg (BH) method [23] within the country-wise groups. The BH method regulates the false discovery rate (FDR), which has a smaller risk of false-positivity than the raw p -value and is more powerful than the most stringent Bonferroni method.

In addition, as a reference, we also calculated the Pearson correlation and Spearman’s rank correlation between the raw GT keyword trends and the

COVID-19 weekly positivity trends, as in the earlier GT-based COVID-19 studies. Correlations' *p*-values were similarly adjusted with the BH method.

Incorporating media coverage trends

We then evaluated the media coverage of the obtained GT keywords with a statistically reliable temporal relationship with the COVID-19 weekly positivity trend. Due to the shortness of available data, we could only analyze the media coverage trend of those keywords in the Japan region. We reviewed *Nikkei Telecom* (<http://telecom.nikkei.co.jp>), a large Japanese database covering newspapers, TV news, Internet news, and general magazines published in Japan, to measure the weekly number of published articles in which the title/abstract/manuscript included the identified Japanese keyword. Specialized magazines were excluded from the reviewed publication review because they might have less exposure to the general population. The obtained time series of the weekly count of articles containing the keyword was used as the media coverage trend in Japan. Then, we again evaluated whether the identified GT keyword trend still Granger-caused the COVID-19 weekly positivity, even when adjusted with the simultaneous media coverage trend of the keyword. This partial Granger-causality analysis was performed using the R package *FIAR* [24].

Ethics

This study was approved by the University of Tokyo Graduate School of Medicine Institutional Ethics Committee (ID: 11,628-(3)). Informed consent was not required because the data were publicly distributed. The study was conducted in accordance with the ethical standards laid out in the Declaration of Helsinki, 1964.

Results

General COVID-19 related trends

During the 3-year period from October 1, 2017 to October 25, 2020, different countries experienced different timings in their COVID weekly positivity trends and the related GT search trends. Figure 2 shows weekly trends of each country (from upper-left to lower-right in alphabetical order by country code). The solid lines show the weekly COVID-19 positivity trends while the dotted lines denote GT search volume trends for the "COVID" keyword in each region (or its corresponding Japanese keyword in Japan). Both trends are plotted in a normalized manner so that the maximum value of each trend within the reviewed period becomes 100%. Briefly, as of late October 2020, for both the COVID-19 weekly positivity trend and the COVID search volume trends, Australia (AU), Japan (JP), and the United States (US) experienced their first and second waves (i.e., large positive peaks), while Canada (CA), Great Britain (GB), and Ireland (IE) are currently experiencing their second wave. Meanwhile, although India (IN) and South Africa (ZA) experienced

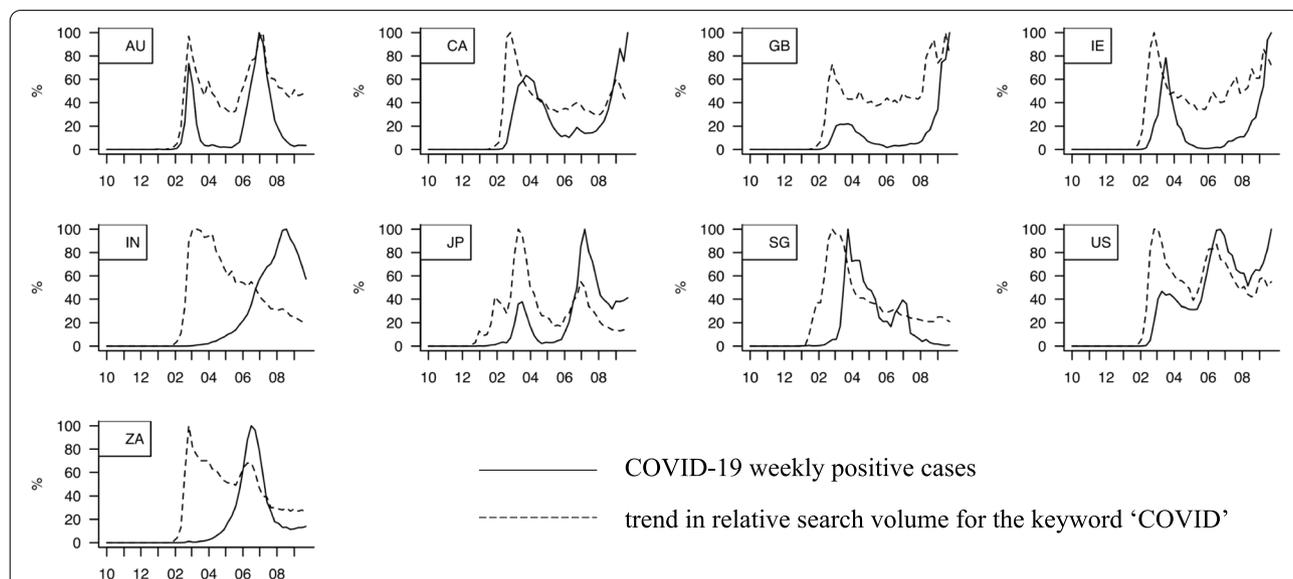


Fig. 2 COVID-19 weekly positivity trends and related GT search volume trends for 'COVID' in each region. Figures show weekly trends of each country (from upper-left to lower-right in alphabetical order of country name), where the solid lines show COVID-19 positivity trends while the dotted lines denote GT search volume trends for 'COVID' word (or its corresponding Japanese word) in each region. Both trends are plotted in a normalized manner so that the maximum value within the period becomes 100%. X-axis in months since October 2019 to September 2020

delayed first waves of weekly COVID-19 positivity compared to other countries, search volume trends showed the first wave surge, the timing of which was similar to that of the other countries.

VAR model in comparison with Pearson / Spearman’s rank correlation

Next, we conducted a VAR model analysis. Table 2 summarizes the number of keywords of which GT trends had significant (*p*-value or FDR<0.05) temporal relationships with the COVID-19 weekly positivity trends, in terms of Granger-causality (by the GT keyword trend

onto the COVID-19 weekly positivity trend; columns A and B), Pearson correlation (columns C and D), or Spearman’s rank correlation (columns E and F). For all the countries, the number of significant keywords was smaller in Granger-causality than in Pearson correlation (columns A vs C, B vs D) or in Spearman’s rank correlation (columns A vs. E, B vs. F), and the influence of multiple test adjustment (BH method) seemed to be larger in terms of Granger-causality (columns A to B) than in correlations (columns C to F). Specifically, the number of significant keywords identified by Granger-causality (with multiple tests adjusted: median 10 words

Table 2 The number of significant keywords which have temporal association with the COVID-19 positivity trend

	Maximum lag in VAR model	Granger causality		Pearson correlation		Spearman’ rank correlation	
		(A) raw <i>p</i> -value <0.05	(B) FDR <0.05	(C) raw <i>p</i> -value <0.05	(D) FDR <0.05	(E) raw <i>p</i> -value <0.05	(F) FDR <0.05
AU	4	21	16	24	20	35	34
	6	21	16	24	20	35	34
	8	21	16	24	20	35	34
CA	4	25	20	29	25	37	37
	6	28	24	29	25	37	37
	8	28	24	29	25	37	37
GB	4	22	18	39	38	42	42
	6	16	8	39	38	42	42
	8	20	10	39	38	42	42
IE	4	17	10	23	18	26	25
	6	19	15	23	18	26	25
	8	20	15	23	18	26	25
IN	4	4	1	49	49	53	53
	6	4	1	49	49	53	53
	8	5	1	49	49	53	53
JP	4	10	5	33	33	44	44
	6	12	5	33	33	44	44
	8	14	5	33	33	44	44
SG	4	10	5	21	18	24	24
	6	11	7	21	18	24	24
	8	11	6	21	18	24	24
US	4	29	26	38	34	41	40
	6	27	27	38	34	41	40
	8	28	25	38	34	41	40
ZA	4	7	3	33	31	37	36
	6	11	3	33	31	37	36
	8	12	4	33	31	37	36

The number of keywords in which GT trend had a significant Granger-caused COVID-19 positivity trends (A, raw *p*-value < 0.05; B, FDR < 0.05), the number of keywords whose GT trend had significant Pearson correlation with the COVID-19 positivity trends (C, raw *p*-value < 0.05; D, FDR < 0.05), and the number of keywords whose GT trend had significant Spearman’s rank correlation with the COVID-19 positivity trends (E, raw *p*-value < 0.05; F, FDR < 0.05). For Granger causality in each region, the lag order of the VAR model is varied in the range of 1–8

In addition, inversely, the number of keywords of which trend is predictable by the COVID-19 positivity trend was very few (not shown in this table): *p*-value / FDR are shown in the ‘Granger (opposite)’ columns in Additional files 3 and 4, where FDR < 0.05 was only 1 in Great Britain, 2 in India, 1 in Japan, and at most 11 in South Africa. Abbreviations: AU Australia, CA Canada, GB Great Britain, IE Ireland, IN India, JP Japan, SG Singapore, US United States, ZA South Africa, FDR False discovery rate, VAR Vector autoregression

(95% confidence interval [CI]: 1.0–25.7 words) (Table 2, column B) significantly decreased ($p < 0.001$ in Welch's t-test) when compared to those identified by unadjusted Pearson correlation (median 33 words, 95% CI: 21–49 words) (Table 2, column C) or Spearman's rank correlation (median 37 words, 95% CI: 24–53 words) (Table 2, column E), especially in countries such as India, Japan, Singapore, and South Africa (outside Europe or North America). The above results are also visualized for each country in Additional file 2: Spearman's rank correlation p -value (x-axis) and Granger-causality FDR (y-axis) of the same keywords, showing biased distribution towards non-significant level in Granger-causality FDR in many of the examined keywords (list of all the p -value and FDR results are provided in Additional files 3 and 4). These results suggest that the current approach with appropriately adjusted Granger-causality analysis yields more stringent and statistically reliable results than the unadjusted correlation tests, depending on the region.

The detailed results of the keywords that had significant Granger-causality (FDR < 0.05) to the weekly COVID-19 positivity trends are shown in Table 3 in decreasing order of identified frequency across the nine countries. Only keyword trends that had significant Granger-causality in four or more countries (out of the nine countries) are listed. The check mark indicates that the keyword (in row) had significant Granger-causality in that country (in column). The lag order of the VAR model of each keyword is determined from the range of 1–4. The anosmia-related keyword “loss of smell” (or its corresponding Japanese keywords (Additional file 1)) was identified in

all nine countries, and the keyword “sense of smell” (or its corresponding Japanese keyword) was identified in five out of the nine countries.

Figure 3 visualizes GT search volume trends for “loss of smell” (or its corresponding Japanese word) for each country (dotted lines), which showed a clear temporal relationship with weekly COVID-19 positivity trends (solid lines). Other identified symptom keywords were as follows: “cough” (5/9 countries), “loss of taste” (5/9), “runny nose” (5/9), “stuffy nose” (5/9), “sore throat” (5/9), “sore” (5/9), “shortness of breath” (5/9), “diarrhea” (4/9), “headache” (4/9), and “pneumonia” (4/9). These are well-known symptoms of COVID-19 [2, 20] and partly overlap with the GT keywords reported to have significant associations with weekly COVID-19 case trends [6–8].

Media coverage of keywords

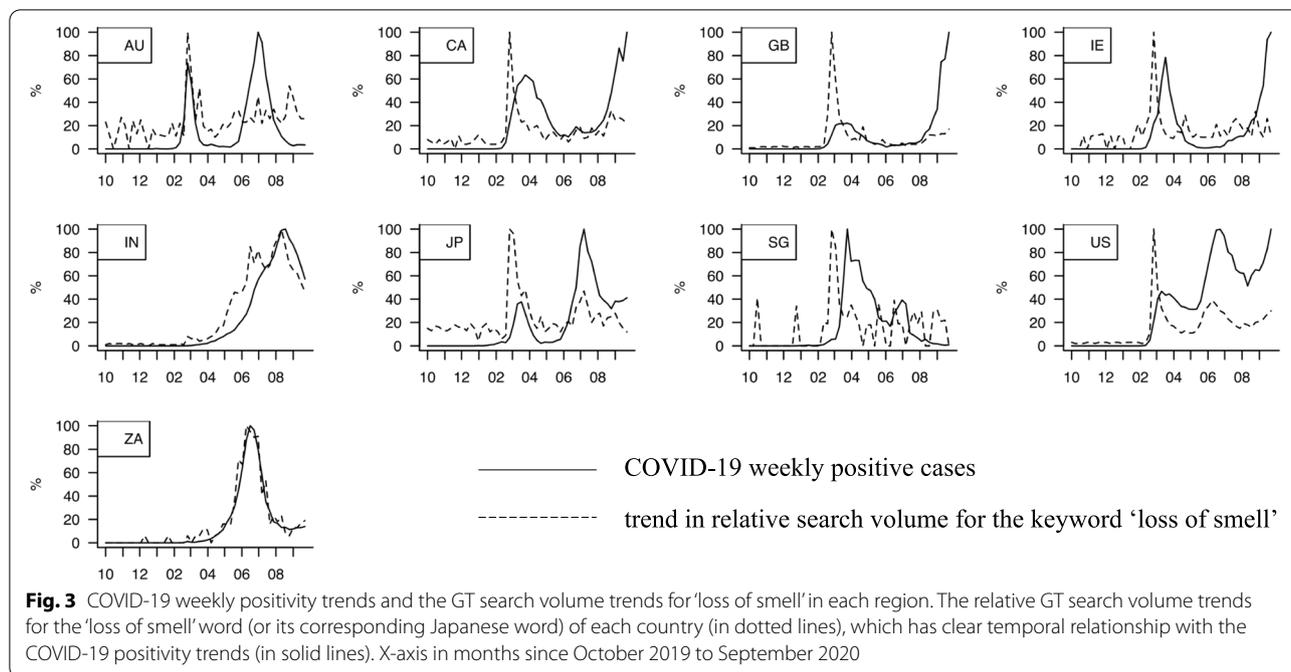
The Japanese keywords corresponding to “loss of smell” and “sense of smell” were the only significant ones in Japan (Table 3, filled cells) and were also the most frequently identified keywords across the different countries, so we selected them to further assess the effect of media coverage trends on these keywords in the Japanese data. Figure 4 presents the temporal relationship between the weekly COVID-19 positivity trend (solid lines), the GT trend of the Japanese keywords (dotted lines) corresponding to (A) “loss of smell” or (B) “sense of smell,” and their media coverage trends (dashed lines). Apparently, in both keywords (A and B), the GT keyword trends were very similar to the trends in their media coverage. Notably, both the Granger-causality of the keywords “loss of

Table 3 Top frequent keywords significantly Granger-caused the COVID-19 positivity trends

Keywords	Total frequency	Regions								
		AU	CA	GB	IE	IN	JP	SG	US	ZA
Loss of smell	9	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sense of smell	5	-	✓	✓	-	-	✓	-	✓	✓
Loss of taste	5	✓	-	✓	✓	-	-	-	✓	✓
Cough	5	✓	✓	✓	✓	-	-	-	✓	-
Runny nose	5	✓	✓	✓	✓	-	-	-	✓	-
Shortness of breath	5	✓	✓	✓	-	-	-	✓	✓	-
Sore	5	✓	✓	-	✓	-	-	✓	✓	-
Sore throat	5	✓	✓	✓	✓	-	-	-	✓	-
Stuffy nose	5	✓	-	✓	✓	-	-	✓	✓	-
Diarrhea	4	-	✓	✓	✓	-	-	-	✓	-
Headache	4	✓	✓	✓	-	-	-	-	✓	-
Pneumonia	4	✓	✓	✓	-	-	-	-	✓	-

Detailed results of the keywords that have significant Granger-causality to the COVID-19 positivity trends, in the order of frequency across all the countries. Only keywords that were significant in 4 or more countries (out of 9) are shown. The check mark indicates that the keyword (in row) had significant Granger-causality in that country (in column)

Abbreviations: AU Australia, CA Canada, GB Great Britain, IE Ireland, IN India, JP Japan, SG Singapore, US United States, ZA South Africa, FDR False discovery rate, VAR Vector autoregression



smell” and “sense of smell” to the weekly COVID-19 positivity trend became non-significant when adjusted with their media coverage by partial Granger-causality analysis ($p=0.257$ and $p=0.384$, respectively). These results suggest a relationship between weekly COVID-19 positivity trends and that the GT trends of anosmia-related keywords are highly confounded by their media coverage.

Discussion

In summary, based on the potential analytical fallacies that are of concern in earlier GT studies, our current study aimed to identify symptom keywords in GT trends that could be used as a predictive measure for future weekly COVID-19 positivity trends by applying more statistically favorable methods. However, the current analysis showed that the number of search keywords that are truly associated with weekly COVID-19 positivity trends may be smaller than reported in earlier studies using a simple Pearson/Spearman correlation, of which the degree depends on the region. In addition, even the GT trends of most reliable anosmia-related keywords were actually a strong reflection of its media coverage (at least in Japan). These results suggest that many of the search keywords reported as candidate predictive measures in earlier GT studies may actually turn out to be false-positive. In other words, the potential candidate keywords listed in the earlier GT-based COVID-19 infodemiological studies are not always reliably usable as true predictive measures. We need to be careful when interpreting published study results as the utility of Google Trends for

studying COVID-19 epidemiology may be more limited than previously expected.

The major strength of our study is its statistically favorable approach with a longer period of included observations. For example, our results evaluating the trend in media coverage of the “loss of smell” keyword is partly consistent with a few of the earlier studies [8, 10]. However, in previous studies, the potential effect of media coverage was not evaluated in a statistically favorable manner, and the association between GT trends and weekly COVID positivity trends had been evaluated in an inappropriate way (i.e., Pearson correlation). Moreover, earlier GT studies did not always examine many symptom keywords related to COVID-19 comprehensively as in our study, so that selection bias cannot be excluded. In contrast, our approach of narrowing down the candidate keywords to adjust for their media coverage was data-driven with a smaller risk of bias in keyword selection. In addition, because our study included a longer period of data (up until October 2020) than most of the earlier GT-based COVID-19 studies, which only included serial data within the first wave (e.g., up until July 2020 in the United States and Japan), lessons based on our results may have higher applicability to the second or later waves of weekly COVID-19 positivity trends.

Our study has some limitations. For example, in the VAR model, the effect of each variable is assumed to be fixed throughout the reviewed period, which may not always be true because the public interest and attitude toward COVID-19 could vary over time [25]. This can

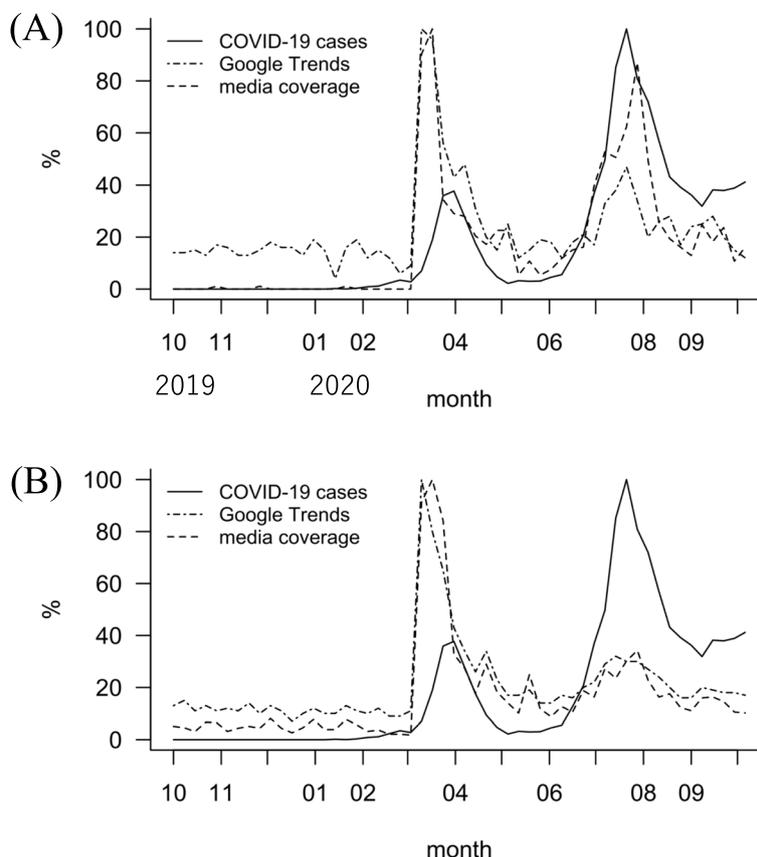


Fig. 4 COVID-19 weekly positivity trends, Google Trends relative search volume trends for ‘loss of smell’ and ‘sense of smell’, and their media coverage trends in Japan. Temporal relationship between the COVID-19 positivity trend (in solid lines), the GT trend of the Japanese keywords (in dotted lines) corresponding to **A** ‘loss of smell’ or **B** ‘sense of smell’, and their media coverage trends (in dashed lines). Apparently, in both keywords (**A** and **B**), the GT keyword trends were very similar to the trends in their media coverage

be suspected by the decreased peak of GT trend for the “COVID” keyword in the second wave (Fig. 2, in Australia, Japan, and the United States). In that sense, the VAR-model used in this study may not always be statistically robust to identify the true predictor of symptom search keywords, although it is still more favorable than using mere Pearson or Spearman’ correlation so far. In future studies, state space modeling [26] to incorporate potentially time-varying effects may be useful to overcome the potential weakness of the VAR model, especially when the included period becomes so long. In addition, the keywords’ media coverage was adjusted only in Japanese regional data, which makes the obtained results slightly less generalizable to other countries. When investigating other countries, other types of database such as *Nexis Uni database* (<https://www.lexisnexis.com/en-us/professional/academic/nexis-uni.page>) may be suitable to analyze in order to confirm the reproducibility of our results. Also, the *Nikkei telecom* we used for media review would not cover all potentially influencing

media such as TV talk shows, or social media (e.g., *Twitter* [9] or *Instagram* [16]).

To conclude, our current results using a more statistically favorable approach suggest that many of the search keywords identified as candidate predictive measures in earlier GT studies have the potential risk of false positives, and that we need to be careful in interpreting the earlier GT-based COVID-19 study results.

Abbreviations

COVID-19: Coronavirus disease; GT: Google Trends; VAR: Vector autoregression.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12874-021-01338-2>.

Additional file 1. English-Japanese corresponding table for symptom keywords searched. The corresponding English-Japanese table for the searched symptom keywords: 54 English keywords (left-sided 3 columns) were used for search in 8 English-speaking country regions and the

corresponding 60 Japanese keywords (right-sided 3 columns) were used for search in the Japan region. Words listed in the same row roughly belong to the similar symptom category.

Additional file 2. Difference in the distribution between Spearman's rank correlation p -value and Granger causality FDR. In each country from upper left to lower right in an alphabetical order, each dot plots the Spearman's rank correlation p -value and the Granger-causality FDR value (each of lag 4, 6, and 8: differently colored) of the same search keyword. Vertical and horizontal dotted lines show value of 0.05 for the reference of significance. For many of the examined keywords, the p -value/FDR clearly became non-significant level when using Granger-causality analysis (in y-axis) instead of Spearman's rank correlation test (in x-axis). Abbreviations: AU, Australia; CA, Canada; GB, Great Britain; IE, Ireland; IN, India; JP, Japan; SG, Singapore; US, United States; ZA, South Africa; FDR, false discovery rate.

Additional file 3. P -value/FDR results of all keywords in each country (excluding Japan). P -value/FDR results of all English symptom keywords as listed in the left-most row, for each country (excluding Japan) and each statistical condition (lag order 4-8, Spearman's rank correlation or granger-causality, and p -value or FDR). Abbreviations: AU, Australia; CA, Canada; GB, Great Britain; IE, Ireland; IN, India; JP, Japan; SG, Singapore; US, United States; ZA, South Africa; FDR, false discovery rate.

Additional file 4. P -value/FDR results of all keywords in Japan. P -value/FDR results of all Japanese symptom keywords in Japan as listed in the left-most row, for each statistical condition (lag order 4-8, Spearman's rank correlation or granger-causality, and p -value or FDR). Abbreviations: AU, Australia; CA, Canada; GB, Great Britain; IE, Ireland; IN, India; JP, Japan; SG, Singapore; US, United States; ZA, South Africa; FDR, false discovery rate.

Acknowledgements

Not applicable.

Authors' contributions

KS made a study concept and design, acquired and analyzed the data, interpreted the results, and drafted the manuscript. TM, AI, and TT have contributed to the interpretation of the results and revision of the manuscript. The author(s) read and approved the final manuscript.

Authors' information

Not applicable.

Funding

This work was supported by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Numbers 20J11009 (K.S) and 20H03587 (A.I), and also supported by AMED under Grant Number 16dk0207028h0001, 20dk0207048h0002, 21dk0207057h0001.

Availability of data and materials

The data used in this study can be retrieved from Google Trends (<https://trends.google.co.jp/trends>).

Declarations

Ethics approval and consent to participate

This study was approved by the University of Tokyo Graduate School of Medicine Institutional Ethics Committee (ID: 11628-3)). Informed consent was not required because the data were publicly distributed.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 20 January 2021 Accepted: 19 June 2021

Published online: 18 July 2021

References

- Mavragani A, Ochoa G, Tsagarakis KP. Assessing the methods, tools, and statistical approaches in Google Trends research: systematic review. *J Med Internet Res*. 2018;20(11):e270.
- Tenforde MW, Kim SS, Lindsell CJ, Billig Rose E, Shapiro NI, Files DC, et al. Symptom duration and risk factors for delayed return to usual health among outpatients with COVID-19 in a multistate health care systems network - United States, March-June 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(30):993–8.
- Timeline: WHO's COVID-19 response. <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/interactive-timeline>. Accessed 6 Apr 2021.
- Ayyoubzadeh SM, Ayyoubzadeh SM, Zahedi H, Ahmadi M, Kalhori SRN. Predicting COVID-19 incidence through analysis of Google Trends data in Iran: data mining and deep learning pilot study. *JMIR Public Health Surveill*. 2020;6(2):e18828.
- Mavragani A. Tracking COVID-19 in Europe: infodemiology approach. *JMIR Public Health Surveill*. 2020;6(2):e18941.
- Cherry G, Rocke 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;16:1–6.
- Ciofani JL, Han D, Allahwala UK, Asrress KN, Bhindi R. Internet search volume for chest pain during the COVID-19 pandemic. *Am Heart J*. 2020;S0002–8703(20):30258–61.
- Higgins TS, Wu AW, Sharma D, Illing EA, Rubel K, Ting JY, Snot Force Alliance. Correlations of online search engine trends with coronavirus disease (COVID-19) incidence: infodemiology study. *JMIR Public Health Surveill*. 2020;6(2):e19702.
- 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;163(3):491–7.
- Sousa-Pinto B, Anto A, Czarlewski W, Anto JM, Fonseca JA, Bousquet J. Assessment of the impact of media coverage on COVID-19-related Google Trends data: infodemiology study. *J Med Internet Res*. 2020;22(8):e19611.
- Chiu APY, Lin Q, He D. News trends and web search query of HIV/AIDS in Hong Kong. *PLoS One*. 2017;12(9):e0185004.
- Crowson MG, Witsell D, Eskander A. Using Google Trends to predict pediatric respiratory syncytial virus encounters at a major health care system. *J Med Syst*. 2020;44(3):57.
- Syamsuddin M, Fakhruddin M, Sahetapy-Engel JTM, Soewono E. Causality analysis of Google Trends and dengue incidence in Bandung, Indonesia with linkage of digital data modeling: longitudinal observational study. *J Med Internet Res*. 2020;22(7):e17633.
- Rehman AU, Malik MI. The modified R a robust measure of association for time series. In: MPRA paper 60025. Germany; University Library of Munich; 2014.
- Cervellin G, Comelli I, Lippi G. Is Google Trends a reliable tool for digital epidemiology? Insights from different clinical settings. *J Epidemiol Glob Health*. 2017;7(3):185–9.
- Rovetta A, Bhagavathula AS. Global infodemiology of COVID-19: analysis of Google web searches and Instagram hashtags. *J Med Internet Res*. 2020;22(8):e20673.
- Massicotte P, Eddelbuettel D. gtrendsR: perform and display Google Trends queries. R package version 1.4.2. 2018. <https://CRAN.R-project.org/package=gtrendsR>.
- Trapletti A, Hornik K. tseries: time series analysis and computational finance. R package version 0.10–47. 2019.
- Pfaff B. VAR, SVAR and SVEC models: implementation within R package vars. *J Stat Softw*. 2008;27(4):1–32. <http://www.jstatsoft.org/v27/i04/>.
- Liew VK-S. Which lag length selection criteria should we employ? *Econ Bull*. 2004;3(33):1–9.
- Expert meeting on the novel coronavirus disease control analysis of the response to the novel coronavirus (COVID-19) and recommendations (Excerpt), in March 19, 2020. <https://www.mhlw.go.jp/content/10900000/000611515.pdf>. Accessed 6 Apr 2021.
- Johnson KD, Beiglböck M, Eder M, Grass A, Hermisson J, Pammer G, Polechová J, Toneian D, Wölfl B. Disease momentum: estimating the

- reproduction number in the presence of superspreading. *Infect Dis Model.* 2021. <https://doi.org/10.1016/j.idm.2021.03.006>.
23. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res.* 2001;125(1–2):279–84.
 24. Roelstraete B, Rosseel Y. FIAR: an R package for analyzing functional integration in the brain. *J Stat Softw.* 2011;44(13):1–32. <http://www.jstatsoft.org/v44/i13/>.
 25. Husain I, Briggs B, Lefebvre C, Cline DM, Stopyra JP, O'Brien MC, et al. Fluctuation of public interest in COVID-19 in the United States: retrospective analysis of Google Trends search data. *JMIR Public Health Surveill.* 2020;6(3):e19969.
 26. Kobayashi G, Sugawara S, Tamae H, Ozu T. Predicting intervention effect for COVID-19 in Japan: state space modeling approach. *Biosci Trends.* 2020;14(3):174–81.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

