

ARTICLE OPEN



Internet search patterns reveal clinical course of COVID-19 disease progression and pandemic spread across 32 countries

Tina Lu^{1,2} and Ben Y. Reis^{1,3}

Effective public health response to novel pandemics relies on accurate and timely surveillance of pandemic spread, as well as characterization of the clinical course of the disease in affected individuals. We sought to determine whether Internet search patterns can be useful for tracking COVID-19 spread, and whether these data could also be useful in understanding the clinical progression of the disease in 32 countries across six continents. Temporal correlation analyses were conducted to characterize the relationships between a range of COVID-19 symptom-specific search terms and reported COVID-19 cases and deaths for each country from January 1 through April 20, 2020. Increases in COVID-19 symptom-related searches preceded increases in reported COVID-19 cases and deaths by an average of 18.53 days (95% CI 15.98–21.08) and 22.16 days (20.33–23.99), respectively. Cross-country ensemble averaging was used to derive average temporal profiles for each search term, which were combined to create a search-data-based view of the clinical course of disease progression. Internet search patterns revealed a clear temporal pattern of disease progression for COVID-19: Initial symptoms of fever, dry cough, sore throat and chills were followed by shortness of breath an average of 5.22 days (3.30–7.14) after initial symptom onset, matching the clinical course reported in the medical literature. This study shows that Internet search data can be useful for characterizing the detailed clinical course of a disease. These data are available in real-time at population scale, providing important benefits as a complementary resource for tracking pandemics, especially before widespread laboratory testing is available.

npj Digital Medicine (2021)4:22; <https://doi.org/10.1038/s41746-021-00396-6>

INTRODUCTION

Accurate real-time surveillance of disease spread is essential for effective pandemic response and for the allocation of scarce healthcare resources^{1,2}. During the current COVID-19 pandemic, population-level surveillance has relied primarily on aggregated results from individual laboratory testing³. While laboratory testing remains the primary method for diagnosing individuals, using laboratory testing for population-level surveillance has its limitations—it is difficult to achieve population-level coverage due to delays in development and scaling of de novo laboratory testing methodologies during the crucial early stages of an emergent pandemic^{4–6}. These limitations present an opportunity for exploring additional approaches to population-level surveillance that could serve as complementary information sources during the critical early stages of a pandemic⁷.

Aggregated data on Internet search volumes are freely available in near-real-time and at population-scale in areas with sufficient Internet penetration⁸. These Internet search patterns have been used to track a wide range of health phenomena, including influenza⁹, MERS¹⁰, measles¹¹, abortion¹² and immunization compliance¹³, and are a potential complementary source of information for population-level surveillance of pandemic spread. Previous studies of these data have yielded valuable lessons in their appropriate use, including avoiding non-specific search terms, ensuring that analyses are straightforward and transparent, and avoiding complex models that may increase the risks of overfitting. A number of recent studies have examined whether Internet search data could be useful for tracking the spread of COVID-19 at a population level. These studies have found correlations between specific Internet search terms and reported

COVID-19 cases^{14–29}. A detailed review of these studies is provided in the Discussion section below.

Unlike these recent studies which have focused on using Internet search data for tracking pandemic spread, in this study we examine whether Internet search data can be used for another important purpose: characterizing the clinical course of symptoms in affected individuals, especially during the early stages of an emergent pandemic. Characterizing the expected clinical course of a disease can help clinicians understand where an individual patient is in his or her expected disease course, and what can be expected for that patient in the future. For example, in the case of COVID-19, it is important to know when dyspnea (shortness of breath) is expected to occur for patients, so that appropriate planning of clinical care and resources can take place. Understanding the expected sequence of symptoms at a population level could also help public health officials track the different stages of the onset of a pandemic as it takes hold in a new region. It would therefore be beneficial for pandemic tracking and clinical planning if disease progression of novel pandemics could be studied in a timely manner and at population scales.

A number of recent clinical case studies have analyzed the day-by-day clinical course for COVID-19 patients, as detailed in the Discussion section below. These studies have reported a time lag between the onset of initial symptoms and the appearance of dyspnea (shortness of breath) of around 5 days^{30–34}. These studies relied on a limited number of patients in a hospital setting, and were published weeks and months after the initial spread of the pandemic.

Unlike hospital-based clinical studies, analyses of Internet search data could potentially provide information on disease progression across a much larger population of patients, not only

¹Predictive Medicine Group, Computational Health Informatics Program, Boston Children's Hospital, Boston, MA, USA. ²Harvard University, Cambridge, MA, USA. ³Harvard Medical School, Boston, MA, USA. ✉email: Ben_Reis@harvard.edu

those seen within hospital settings. In this study, we conducted a systematic study across 32 countries on six continents to determine whether Internet search patterns can be useful as a complementary data source for understanding the clinical progression of COVID-19, especially during the critical early stages of the pandemic.

RESULTS

Internet search volume and COVID-19 cases and deaths

We examined a few example countries in which the pandemic peaked at different times. Figure 1 shows search volumes for the terms “fever” and “dry cough”, alongside reported COVID-19 cases and deaths for China, Iran, Italy, United States and India. Even though outbreaks occurred at different times in each country, the temporal relationships between the search terms and reported COVID-19 cases and deaths remained similar across countries.

Expanding this analysis to all 32 countries, Fig. 2a shows the lags between search volumes for the term “fever” and COVID-19-related deaths for each of the 32 countries, along with a histogram showing the distribution of these lags. Figure 2b shows the lags between search volumes for the term “dry cough” and COVID-19-related deaths, along with the distribution of the lags.

Table 1 shows summary statistics over all 32 countries for the lags between each of the ten search terms and COVID-19 cases and deaths, including mean and 95% confidence intervals. Searches for “fever” preceded increases in reported COVID-19 cases and deaths by an average of 18.53 days [95% CI 15.98–21.08] and 22.16 days [95% CI 20.33–23.99], respectively. For “cough” these times were 18.34 days [95% CI 16.05–20.64] and 21.88 days [95% CI 20.19–23.56], respectively. The complete list of lags between each search term and COVID-19 deaths for each individual country is provided in Supplementary Table 1.

The average lags between searches and reported cases were shorter than those between searches and reported deaths, as cases are typically diagnosed and reported before deaths. We found that the inter-country variability of the average lags between searches and reported cases was greater (larger confidence intervals) than that for reported deaths, likely because case reporting is more dependent on local testing capacity and rates. We also found that the general term “coronavirus” has the greatest variability in its lags to reported cases and deaths (largest confidence intervals), compared to other symptom-specific terms, as would be expected for a more general search term.

Internet search volume and COVID-19 clinical course of illness

We examined examples from individual countries, this time looking at the progression of symptom-related search terms over time. Figure 3 shows search volumes for a range of symptom-related search terms plotted over time for France (Fig. 3a) and Mexico (Fig. 3b). For each country, searches for “shortness of breath” appear a few days after searches for other symptoms such as “fever”, “cough”, and “dry cough”.

Expanding this analysis to all 32 countries, we calculated ensemble average curves for each search term across all 32 countries. Figure 4a shows the ensemble average search volumes for “fever”, “cough”, “dry cough”, and “shortness of breath”, indexed by searches for “coronavirus symptoms”, alongside reported COVID-19 cases and deaths. Figure 4b shows this same analysis for additional search terms “sore throat”, “runny nose”, “chills”, and “coronavirus test”, also indexed by searches for “coronavirus symptoms”.

The clinical progression that emerges from these data presents the following picture: As the pandemic begins to take hold in a country, people search for “coronavirus symptoms” and “coronavirus test”, followed by initial symptoms “fever”, “cough”, “runny nose”, “sore throat” and “chills”, followed by searches for

“shortness of breath” approximately 5 days after the search for initial symptoms.

Table 2 shows the lags between different definitions for initial symptom onset and searches for “shortness of breath”. The average lag between the searches for “fever” and “shortness of breath” was 5.22 days [95% CI 3.30–7.14]. For “cough” it was 5.16 days [95% CI 3.13–7.18]. These lags, as well as the lags deriving from other symptom onset definitions, are all around 5 days, matching the clinical course of the disease reported in the literature^{30–34}.

The clinical progressions above are based on a 32-country ensemble average. While there is greater variability at the individual country level, the general temporal order of COVID-19 symptom progression remains fairly consistent across individual countries, with few exceptions. For example, only two of the 32 countries had searches for “coronavirus symptoms” peak after searches for “shortness of breath”, only 3 out of 32 countries had searches for “cough” peak after searches for “shortness of breath”, and only 4 out of 32 countries had searches for “fever” peak after searches for “shortness of breath” (Supplementary Table 2).

DISCUSSION

In this systematic analysis of Internet search data from 32 countries across six continents, we found that increases in symptom-related Internet searches preceded increases in reported COVID-19 cases and deaths by approximately 2–3 weeks. Furthermore, we found that by analyzing data from all 32 countries, the temporal relationships between the different symptom-related search terms reflected the clinical progression of the symptoms of COVID-19 reported in the medical literature.

To the best of the authors’ knowledge, this study is the first to conduct a detailed investigation of multiple COVID-19 symptom-related search terms across such a large number of countries and languages. It is also the first to conduct a detailed analysis of the temporal relationships between different symptom-related searches in order to determine whether these data could be useful in understanding the clinical course of illness for a disease. During emergent pandemics, this detailed level of information can support health officials in tracking pandemic spread and planning clinical care and resources.

A number of clinical case studies have analyzed the day-by-day clinical course for COVID-19 patients: A study of 138 patients in Wuhan, China found that common initial symptoms include fever and dry cough, followed by dyspnea a median of 5.0 days later³⁰. A meta-review found that the average length from illness onset to dyspnea was 4.99 days, based on evidence from 179 COVID-19 patients³¹. A study in the US found that dyspnea typically set in between the fourth and eighth day following initial symptoms³². A study of 41 patients in Wuhan, China found that the median duration from illness onset to dyspnea was 8.0 days³³. The CDC’s *Interim Clinical Guidance for Management of Patients with Confirmed Coronavirus Disease (COVID-19)* states “Among patients who developed severe disease, the median time to dyspnea ranged from 5 to 8 days”³⁴.

A number of recent studies have analyzed Internet search data related to the COVID-19 pandemic. Most of these studies examined data from a single country or from a small number of countries: Some studies examined only the search term “coronavirus”^{35–38}. In this study we found that general non-symptom-specific search terms such as “coronavirus” have a greater variability in their relation to reported cases and deaths, likely due to the fact that individuals seeking general information on the pandemic may search for “coronavirus” even if they are not experiencing specific symptoms themselves at that time. Other studies looked at additional search terms such as “handwashing”, “face masks”^{39,40}, “quarantine”, “hand disinfection”¹⁴, “SARS”,

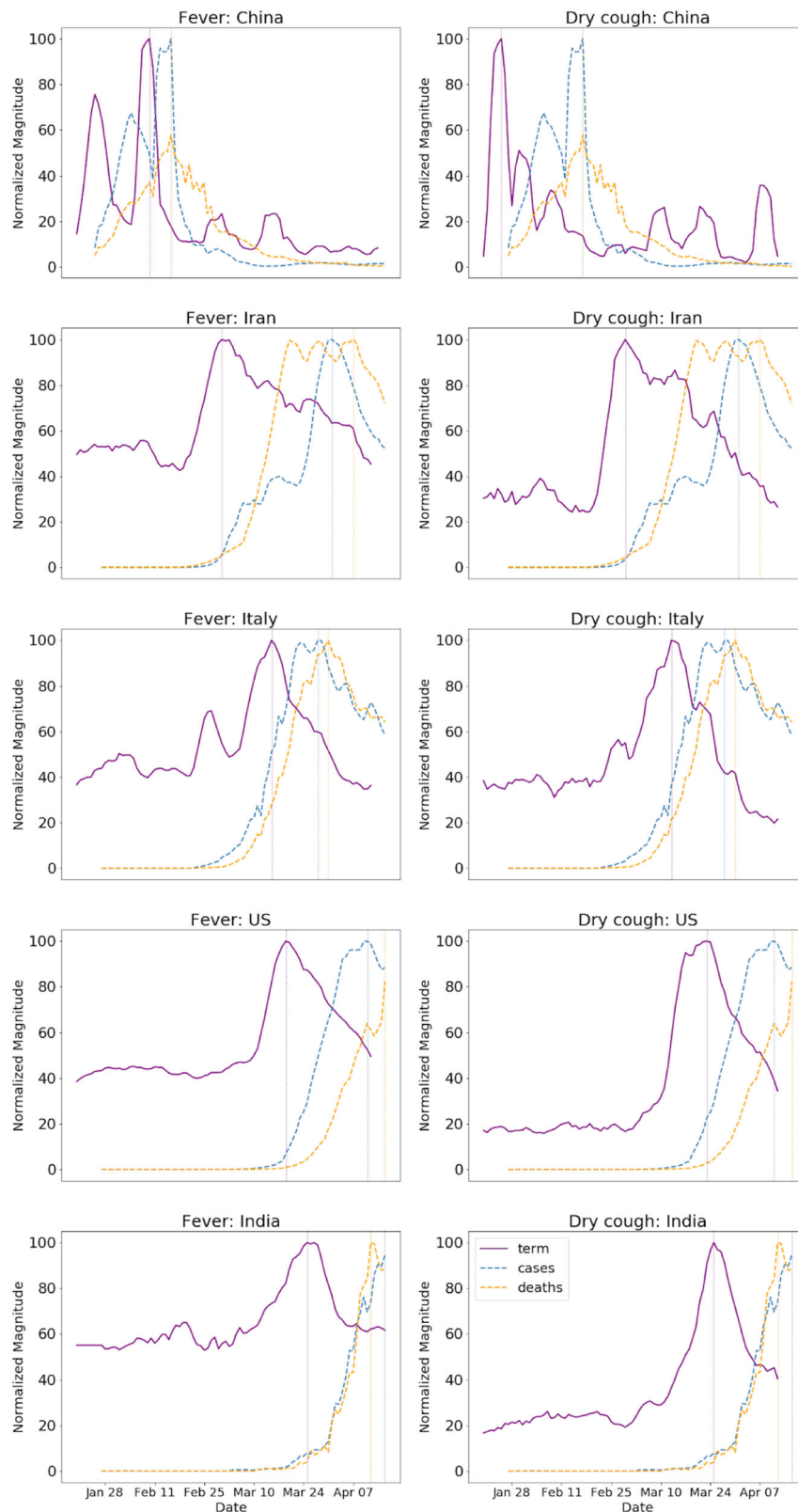


Fig. 1 Search volumes in different countries. Search volumes (purple) for the terms “fever” (left) and “dry cough” (right), alongside reported COVID-19 cases (cyan) and deaths (orange) for China, Iran, Italy, US and India. Even though outbreaks occur at different times in different countries, the relationships between the search terms and reported COVID-19 cases and deaths remain similar across countries. To highlight the temporal relationships between the curves, the magnitude of each curve was independently normalized to fit the vertical dimensions of the plot.

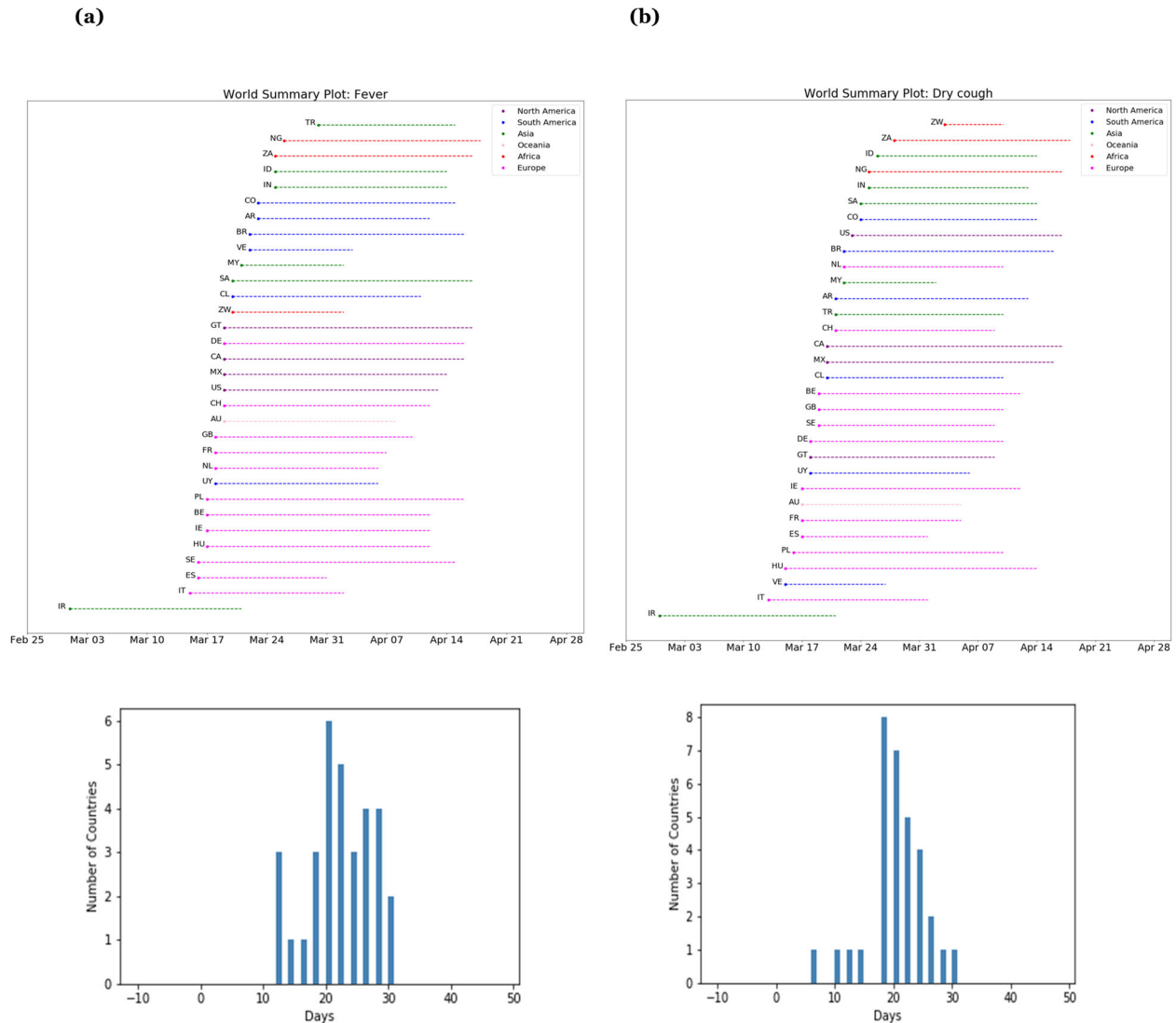


Fig. 2 Search lags in different countries. **a** Lags between searches for “fever” and reported COVID-19 deaths across 32 countries, with a histogram showing the distribution of these lags. Each country is labeled with its ISO Alpha-2 country code. **b** The same plots shown for searches for “dry cough”.

“MERS”¹⁵, “antiseptic”, and “sanitizer”¹⁶, but did not include specific symptom-related search terms.

Another group of studies did examine symptom-related search terms, including: “COVID pneumonia” and “COVID heart” in the US¹⁷; “COVID” and “pneumonia” in China¹⁸; “worry”, “hysteria” and other mental health symptoms in the US¹⁹; “diarrhea”, “nausea”, “vomiting”, “abdominal pain”, “fever” and “cough” in the US²⁰; “can’t smell”, “ear pain”, “sinus pain”, “voice pain”, “ears ringing”. Among other terms in the US²¹; “ageusia”, “abdominal pain”, “loss of appetite”, “anorexia”, “diarrhea”, and “vomiting” in the US²²; “chest pain”, “myocardial infarction”, “cough”, and “fever” in four countries²³; “Loss of sense of smell”, “Sense of smell”, “Loss of sense of taste”, and “Sense of taste” across multiple regions in five countries²⁴; and “smell”, “loss of smell”, “anosmia”, “hyposmia”, “olfaction”, “taste”, “loss of taste”, and “dysgeusia” in eight countries²⁵. While these studies investigated a range of important questions, including the effects of COVID-19 on mental health¹⁹, gastrointestinal health^{20,22}, otolaryngological health²¹, and coronary-related conditions²³, they did not study whether the temporal relationships between these search terms could be useful in understanding the clinical progression of the disease.

Some previous studies have attempted to analyze information on temporal progression in online data sources. For example, Liu et al. analyzed Twitter data to resolve different stages of behavior relating to alcohol consumption (alcohol seeking, alcohol consuming, post-consumption reflection)²⁶. Wu et al. studied seasonality in global public interest in psoriasis by analyzing seasonality in Internet search trends²⁷. To the best of the authors’ knowledge, no previous studies have attempted to reconstruct the clinical progression of a disease based on Internet search data.

The use of Internet search data is subject to a number of important limitations^{9,28,29}. Internet infrastructure and digital access levels differ across countries and communities. Even though digital access rates are generally increasing worldwide, many developing countries currently lack sufficient search volumes to support search-based tracking. Search data may be subject to demographic, socio-economic, geographic, or other biases inherent in the local digital divide^{8,41,42}. In each country, the population of individuals who perform Internet searches may have different characteristics than those who do not, and the results inferred from Internet searching behaviors may not generalize to other populations.

Table 1. Mean lags between search volumes for specific search terms and reported COVID-19 cases and deaths across 32 countries.

Search term	Days to COVID-19 cases Mean [95% CI]	Days to COVID-19 deaths Mean [95% CI]
Coronavirus symptoms	21.97 [19.59–24.35]	24.50 [22.79–26.21]
Coronavirus test	19.44 [17.12–21.76]	22.66 [20.93–24.38]
Fever	18.53 [15.98–21.08]	22.16 [20.33–23.99]
Cough	18.34 [16.05–20.64]	21.88 [20.19–23.56]
Coronavirus	17.53 [14.33–20.73]	21.12 [18.63–23.62]
Runny nose	17.50 [14.71–20.29]	21.06 [18.57–23.55]
Dry cough	17.44 [14.69–20.19]	20.81 [19.09–22.52]
Sore throat	16.75 [14.22–19.28]	20.84 [18.71–22.98]
Chills	16.75 [14.12–19.38]	19.38 [17.49–21.26]
Shortness of Breath	13.81 [11.23–16.39]	18.00 [15.77–20.23]

Mean lags are reported in days, along with 95% confidence intervals. Search terms are sorted chronologically, with those that peaked earliest appearing at the top of the table.

Furthermore, the motivation behind each individual user's search is not known: Changes in search volumes for symptom-related terms such as "fever" can result not only from increases in COVID-19 cases, but also from general curiosity about the pandemic, the occurrence of other diseases (e.g., influenza, Lassa fever⁴³), news coverage, or other factors. Another important limitation of this study is that there is no definitive gold standard with which to compare the data, as individualized laboratory testing was very limited in the early phases of the pandemic. In addition, while we attempted to work with native speakers to help us with the translations, ultimately there were a few countries for which we had to rely on automated translations, which may not have yielded the optimal search terms.

In this study, we used specific symptom-related search terms and examined data from a geographically diverse set of 32 countries across six continents and multiple languages. In recognition of the inherent variability of the data at the individual country level, we performed a global analysis combining data from all 32 countries. Despite the inherent variability of country-specific search data, our results show that the temporal relationships between the symptom-specific Internet search terms and COVID-19-related cases and deaths remained generally consistent between countries.

The goal of this study is to examine whether Internet search patterns can be useful as a complementary data source for understanding the clinical course of a disease in the early stages of a novel pandemic, when the clinical course of the disease is not yet fully characterized. In such an operational real-time setting, when a novel pandemic emerges, public health authorities could analyze Internet search patterns relating to the reported symptoms of the pandemic in affected countries. Since not all symptoms of an emerging disease may be known in the early stages of a novel pandemic, a systematic scan over a range of possible symptoms could help identify candidate symptoms whose search volumes are trending upwards. If multiple countries display a similar progression in search trends for the same symptoms, this progression of symptoms can help to inform public health officials' understanding of the clinical course of the disease.

An understanding of the clinical progression of symptoms in the early stages of a pandemic could help hospitals and healthcare workers deliver better care to patients. For example, clinicians would be better able to anticipate future symptoms that a patient might experience, as well as when such symptoms are

expected to occur. This would also help with capacity and resource planning for hospital beds and necessary equipment such as ventilators.

The results of this study show that Internet search trends are able to reflect the clinical course of symptoms during the early stages of a pandemic. However, the relationships between search terms and illness may not hold during later waves of the pandemic, as the information-seeking behaviors of members of the public may change as they become better educated about the pandemic over time. The study of these patterns during later phases of a pandemic are a topic for future research.

Future work also includes applying these approaches to other diseases, as well as combining Internet search data with additional data sources such as local testing rates, public health measures, news reports, climatological and air quality variables, among others.

The ability of search data to not only help track pandemic spread, but also to reveal the clinical course of symptoms in emergent pandemics is significant. Given the limitations of laboratory testing for population-level surveillance, search data can be a valuable complementary source for population-scale tracking of pandemics in real time, particularly during the early stages of a pandemic when local testing is not yet available at scale, and can help guide public health response.

METHODS

Data acquisition

We selected a diverse set of 32 countries from six continents (Table 3), in which sufficient search data volumes were available for the search terms of interest. We obtained data on reported COVID-19 cases and deaths for each of these countries from a publicly available dataset maintained by the Center for Systems Science and Engineering at Johns Hopkins University⁴⁴.

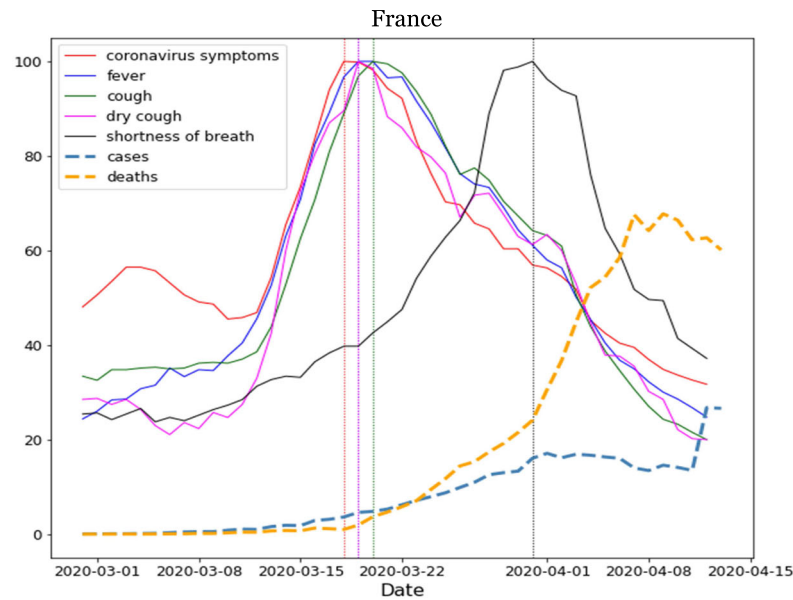
We collected daily relative search volume (RSV) data on a per-country basis for the period of January 1, 2020 through April 20, 2020, from Google Trends (Website: <https://trends.google.com/trends/explore>) using the pytrends API⁴⁵. We used pytrends' "Interest Over Time" API method, which returns aggregated and normalized RSV data from Google Trends' "Interest Over Time" section. Google Trends normalizes the data by dividing each data point by the total searches in the geography and time range it represents in order to allow for comparisons of relative search-term popularity. These numbers are then scaled on a range of 0 to 100. Since Google Trends normalizes the search data and provides only the relative popularity of a search term over time, it was not necessary to include additional control search-terms in the study. As Google has limited availability in China, we also accessed search trend data from the search engine Weibo (Weibo Search Trends: <https://data.weibo.com/index>).

We accessed data for the following common symptoms of COVID-19: "fever", "cough", "dry cough", "chills", "sore throat", "runny nose", and "shortness of breath", as well as the more general terms "coronavirus", "coronavirus symptoms" and "coronavirus test". Each search term was searched as an "exact phrase". We also considered other less common symptoms such as loss of smell and loss of taste, but the search volumes on those terms were too sparse for many countries. As daily search data are inherently noisy, all search data were smoothed with a 7-day moving average²⁴.

For translations of the English search terms into local languages, we recruited native speakers to translate each term into the following languages: Arabic, Mandarin Chinese, Dutch, French, German, Italian, Persian, Polish, Portuguese, and Spanish. These languages covered almost all of the countries included in the study. For languages in which we were unable to recruit native speakers for translations, we used Google Translate (<https://translate.google.com/>) and confirmed that sufficient data were available for these translated search terms on Google Trends (Website: <https://trends.google.com/trends/explore>). A complete table of search terms and translations for each country is provided in Supplementary Table 2.

While both "COVID" and "coronavirus" are commonly used in the United States, detailed consultations with native speakers in different countries revealed that in the majority of countries surveyed, members of the public did not commonly use the terms "COVID" or "COVID-19", but rather only terms relating to "coronavirus". Therefore, for reasons of standardization in

(a)



(b)

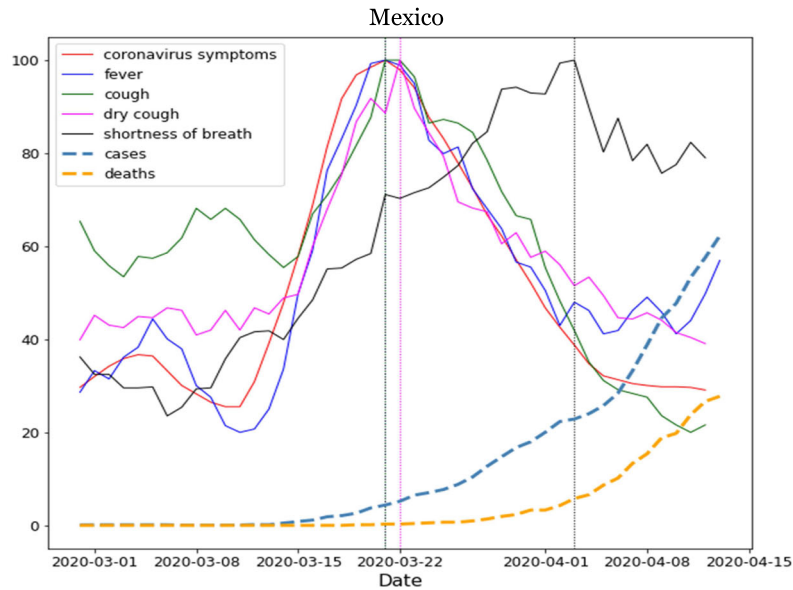


Fig. 3 Clinical course of symptoms in individual countries. Clinical Course of COVID-19 symptoms in (a) France and (b) Mexico. Search volumes for the terms “coronavirus symptoms,” “fever,” “cough,” “dry cough,” and “shortness of breath” (black) are shown alongside COVID-19 cases (dashed cyan line) and deaths (dashed orange line). Initial symptoms appear clustered together in time, with searches for shortness of breath appearing a few days later. To highlight the temporal relationships between curves, the magnitude of each curve was independently normalized to fit the vertical dimensions of the plot.

this international study, we used terms relating to the word “coronavirus” in all countries.

Data analysis: pandemic spread

We conducted temporal correlation studies to study the relationships between Internet search data and reported COVID-19 cases and deaths. For each country and search term, we calculated the Pearson correlation coefficient between the time series of search volumes for that search term and the time series of COVID-19 cases. We then shifted the search term data by a variable lag, and identified the lag that yielded the highest correlation. We computed the mean of these optimal lags for each search term across all countries. We then repeated these analyses substituting reported COVID-19 deaths for reported COVID-19 cases.

Data analysis: clinical course of illness

We investigated whether Internet search data could be used to characterize the clinical course of COVID-19 symptoms over time. “Coronavirus symptoms” was selected as the index search term since it peaked first relative to the other search terms in 22 out of 32 countries. For each country, the date of peak search volume for the index search term was defined as the index date, and the dates for all other search data for that country were defined in relation to that date (Day 0, Day 1, Day 2, etc.). With the data from all countries aligned by the index date, cross-country ensemble average curves were calculated for each search term. (e.g., Day 1 values for “fever” searches in each of the 32 countries were averaged together to calculate the Day 1 value of the ensemble average “fever” curve. The same for Day 2, etc.) The ensemble average curves for all search terms were then overlaid on one plot, providing a search-data-based view of the clinical course of illness.

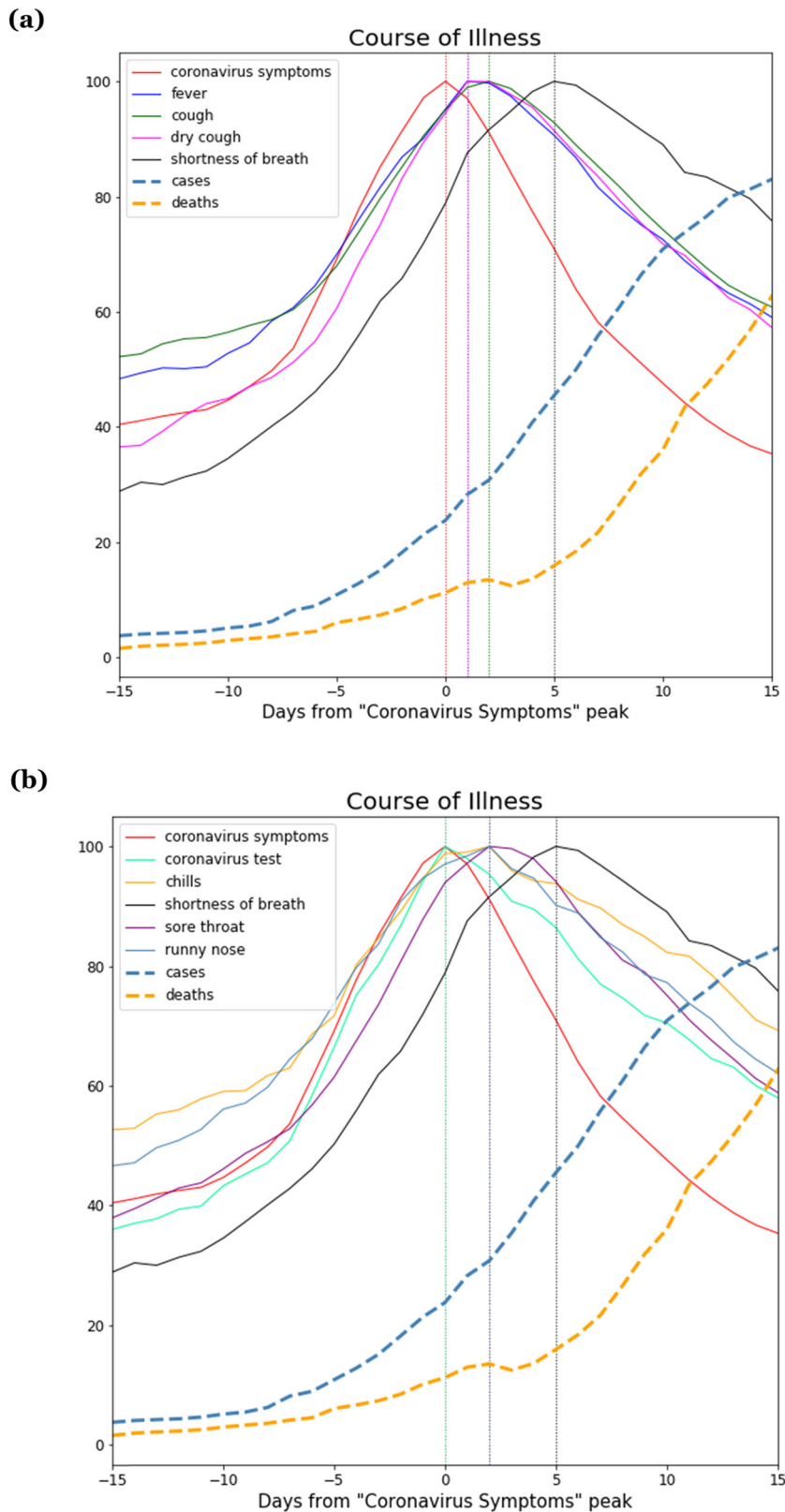


Fig. 4 Average clinical course of symptoms across 32 countries. Ensemble average clinical course of COVID-19 symptoms across 32 countries, as seen through Internet search volumes of search terms related to COVID-19 symptoms. **a** Search volumes for the terms “fever”, “cough”, “dry cough”, “shortness of breath” (black), indexed by searches for “coronavirus symptoms”, shown alongside COVID-19 cases (dashed line cyan) and deaths (dashed line orange). **b** Search volumes for the terms “sore throat”, “runny nose”, “chills” and “coronavirus test”, alongside “shortness of breath” (black), indexed by searches for “coronavirus symptoms”, shown alongside COVID-19 cases (dashed line cyan) and deaths (dashed line orange).

Table 2. Different search-term-based definitions for symptom onset were examined by looking at different combinations of early-peaking search terms.

Symptom onset Search-Term-Based Definition	Days to "Shortness of Breath" Mean [95% CI]
Fever	5.22 [3.30–7.14]
Cough	5.16 [3.13–7.18]
Average of "Fever", "Cough"	5.19 [3.28–7.09]
Average of "Fever", "Cough", "Coronavirus symptoms"	5.81 [3.92–7.70]
Average of "Fever", "Cough", "Coronavirus symptoms", "Coronavirus test"	5.71 [3.85–7.57]

The table shows average lag in days from various search-based definitions of symptom onset to searches for "Shortness of Breath" across 32 countries.

Table 3. Countries included in the study, categorized by continent.

Continent	Countries
Africa	Nigeria, South Africa, Zimbabwe
Asia	India, Iran, Malaysia, Saudi Arabia, Turkey
Oceania	Australia, Indonesia
North America	Canada, Guatemala, Mexico, United States
South America	Argentina, Brazil, Chile, Colombia, Uruguay, Venezuela
Europe	Belgium, France, Germany, Hungary, Ireland, Italy, Netherlands, Poland, Spain, Sweden, Switzerland, United Kingdom

Since studies in the medical literature report the number of days between initial symptom onset and shortness of breath (dyspnea)^{30–34}, we examined a range of possible search-term-based definitions for initial symptom onset, based on various combinations of the earliest-peaking search terms "fever", "cough", "coronavirus symptoms", and "coronavirus test" (Table 2).

Reporting summary

Further information on experimental design is available in the Nature Research Reporting Summary linked to this paper.

DATA AVAILABILITY

All data used in this study are publicly available through the sources referenced in the "Methods" section. The aggregated datasets analyzed in this study are available from the corresponding author on reasonable request.

CODE AVAILABILITY

The pytrends API is publicly available at <https://github.com/GeneralMills/pytrends>.

Received: 16 September 2020; Accepted: 13 January 2021;

Published online: 11 February 2021

REFERENCES

1. Fineberg, H. V. Pandemic preparedness and response—lessons from the H1N1 influenza of 2009. *N. Engl. J. Med.* **370**, 1335–1342 (2014).
2. Balter, S. et al. Pandemic (H1N1) 2009 surveillance for severe illness and response, New York, New York, USA, April–July 2009. *Emerg. Infect. Dis.* **16**, 1259–1264 (2010).

3. Peto, J. Covid-19 mass testing facilities could end the epidemic rapidly. *BMJ* **368**, m1163 (2020).
4. Godlee, F. The burning building. *BMJ* **368**, m1101 (2020).
5. Sheridan, C. Fast, portable tests come online to curb coronavirus pandemic. *Nat. Biotechnol.* <https://doi.org/10.1038/d41587-020-00010-2>. (2020).
6. Xu, M. et al. COVID-19 diagnostic testing: technology perspective. *Clin. Transl. Med.* **10**, e158 (2020).
7. Fagherazzi, G., Goetzing, C., Rashid, M. A., Aguayo, G. A. & Huiart, L. Digital health strategies to fight COVID-19 worldwide: challenges, recommendations, and a call for papers. *J. Med. Internet Res.* **22**, e19284 (2020).
8. Alicino, C. et al. Assessing Ebola-related web search behaviour: insights and implications from an analytical study of Google Trends-based query volumes. *Infect. Dis. Poverty* **4**, 54 (2015).
9. Kandula, S. & Shaman, J. Reappraising the utility of Google Flu Trends. *PLoS Comput. Biol.* **15**, e1007258 (2019).
10. Shin, S.-Y. et al. High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea. *Sci. Rep.* **6**, 32920 (2016).
11. Santangelo, O. E. et al. Digital epidemiology: assessment of measles infection through Google Trends mechanism in Italy. *Ann. Ig.* **31**, 385–391 (2019).
12. Reis, B. Y. & Brownstein, J. S. Measuring the impact of health policies using Internet search patterns: the case of abortion. *BMC Public Health* **10**, 514 (2010).
13. Barak-Corren, Y. & Reis, B. Y. Internet activity as a proxy for vaccination compliance. *Vaccine* **33**, 2395–2398 (2015). vol.
14. Jarynowski, A., Wojta-Kempa, M. & Belik, V. Perception of emergent epidemic of COVID-2019/SARS CoV-2 on the Polish Internet. <https://doi.org/10.1101/2020.03.29.20046789> (2020).
15. Strzelecki, A. & Rizun, M. Infodemiological study using Google Trends on Coronavirus epidemic in Wuhan. *China Int. J. Online Biomed. Eng.* **16**, 139 (2020).
16. Ayyoubzadeh, S. M., Ayyoubzadeh, S. M., Zahedi, H., Ahmadi, M. & R Niakan Kalhori, S. Predicting COVID-19 incidence through analysis of Google Trends data in Iran: data mining and deep learning pilot study. *JMIR Public Health Surveill.* **6**, e18828 (2020).
17. Yuan, X. et al. Trends and prediction in daily incidence and deaths of COVID-19 in the United States: a search-interest based model. *medRxiv* <https://doi.org/10.1101/2020.04.15.20064485>. (2020).
18. Li, C. et al. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from Internet searches and social media data, China, 2020. *Euro Surveill.* **25**, 2000199 (2020).
19. Hoerger, M. et al. Impact of the COVID-19 pandemic on mental health: real-time surveillance using Google Trends. *Psychol. Trauma* **12**, 567–568 (2020).
20. Rajan, A. et al. Association of search query interest in gastrointestinal symptoms With COVID-19 diagnosis in the United States: infodemiology study. *JMIR Public Health Surveill.* **6**, e19354 (2020).
21. Pier, M. M., Pasick, L. J., Benito, D. A., Alnouri, G. & Sataloff, R. T. Otolaryngology-related google search trends during the COVID-19 pandemic. *Am. J. Otolaryngol.* **41**, 102615 (2020).
22. Ahmad, I., Flanagan, R. & Staller, K. Increased internet search interest for GI symptoms may predict COVID-19 cases in U.S. hotspots. *Clin. Gastroenterol. Hepatol.* <https://doi.org/10.1016/j.cgh.2020.06.058>. (2020).
23. Senecal, C., Gulati, R. & Lerman, A. Google trends insights into reduced acute coronary syndrome admissions during the COVID-19 pandemic: infodemiology study. *JMIR Cardio.* **4**, e20426 (2020).
24. Cherry, G. 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.* **18**, 1–6 (2020).
25. Walker, A., Hopkins, C. & Surda, P. The use of Google Trends to investigate the loss of smell related searches during COVID-19 outbreak. *Int. Forum Allergy Rhinol.* <https://doi.org/10.1002/alar.22580>. (2020).
26. Liu, J., Weitzman, E. R. & Chunara, R. Assessing behavior stage progression from social media data. *2017 ACM Conference on Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. Association for Computing Machinery, New York, NY, United States (2017).
27. Wu, Q. et al. Seasonality and global public interest in psoriasis: an infodemiology study. *Postgrad. Med. J.* **96**, 139–143 (2020).
28. Arora, V. S., McKee, M. & Stuckler, D. Google trends: opportunities and limitations in health and health policy research. *Health Policy* **123**, 338–341 (2019).
29. Lazer, D., Kennedy, R., King, G. & Vespignani, A. Big data. the parable of Google Flu: traps in big data analysis. *Science* **343**, 1203–1205 (2014).
30. Wang, D. et al. Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus–infected pneumonia in Wuhan, China. *JAMA* **323**, 1061 (2020).
31. Tan, S. C. Clinical and epidemiological characteristics of Coronavirus Disease 2019 (COVID-19) patients. *medRxiv* <https://doi.org/10.1101/2020.04.02.20050989>. (2019).
32. Cohen, P. A., Hall, L. E., John, J. N. & Rapoport, A. B. The early natural history of SARS-CoV-2 infection: clinical observations from an urban, ambulatory COVID-19 clinic. *Mayo Clin. Proc.* **95**, 1124–1126 (2020).

33. Huang, C. et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020).
34. Centers for Disease Control. Interim Clinical Guidance for Management of Patients with Confirmed Coronavirus Disease (COVID-19). <https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html> (2020).
35. Strzelecki, A. The second worldwide wave of interest in coronavirus since the COVID-19 outbreaks in South Korea, Italy and Iran: a Google Trends study. *Brain Behav. Immun.* <https://doi.org/10.1016/j.bbi.2020.04.042> (2020).
36. Mavragani, A. Tracking COVID-19 in Europe: infodemiology approach. *JMIR Public Health Surveill.* **6**, e18941 (2020).
37. Effenberger, M. et al. Association of the COVID-19 pandemic with internet search volumes: a Google TrendsTM analysis. *Int. J. Infect. Dis.* <https://doi.org/10.1016/j.ijid.2020.04.033> (2020).
38. Benjamin Lucas Brendan Elliot. Online Information Search During COVID-19. Preprint at arXiv <https://arxiv.org/pdf/2004.07183.pdf> (2020).
39. Husnayain, A., Fuad, A. & Su, E. C.-Y. Applications of google search trends for risk communication in infectious disease management: a case study of COVID-19 outbreak in Taiwan. *Int. J. Infect. Dis.* <https://doi.org/10.1016/j.ijid.2020.03.021> (2020).
40. Lin, Y.-H., Liu, C.-H. & Chiu, Y.-C. Google searches for the keywords of 'wash hands' predict the speed of national spread of COVID-19 outbreak among 21 countries. *Brain Behav. Immun.* <https://doi.org/10.1016/j.bbi.2020.04.020> (2020).
41. Din, H. N., McDaniels-Davidson, C., Nodora, J. & Madanat, H. Profiles of a health information-seeking population and the current digital divide: cross-sectional analysis of the 2015-2016 California health interview survey. *J. Med. Internet Res.* **21**, e11931 (2019).
42. McCloud, R. F., Okechukwu, C. A., Sorensen, G. & Viswanath, K. Beyond access: barriers to internet health information seeking among the urban poor. *J. Am. Med. Inform. Assoc.* **23**, 1053–1059 (2016).
43. Wogu, J. O. et al. Mass media reportage of Lassa fever in Nigeria: a viewpoint. *J. Int. Med. Res.* **48**, 300060518821552 (2020).
44. CSSEGISandData. CSSEGISandData/COVID-19. *GitHub* <https://github.com/CSSEGISandData/COVID-19> (2020).
45. GeneralMills. GeneralMills/pytrends. *GitHub* <https://github.com/GeneralMills/pytrends> (2020).

ACKNOWLEDGEMENTS

This work was funded by NIH grant R01HG009129.

AUTHOR CONTRIBUTIONS

B.Y.R. and T.L. conceived of the study, supervised its conduct, and oversaw data collection. B.Y.R. and T.L. designed the study, conducted statistical analysis, drafted the manuscript and formulated the implications of the results. All authors contributed substantially to the revision of the manuscript and approved the final manuscript as submitted.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41746-021-00396-6>.

Correspondence and requests for materials should be addressed to B.Y.R.

Reprints and permission information is available at <http://www.nature.com/reprints>

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021