Interest in COVID-19 in Latin America and the Caribbean: an infodemiological study using Google Trends

Interés por la COVID-19 en Latinoamérica y el Caribe: un estudio infodemiológico usando Google Trends

Interesse na COVID-19 na América Latina e Caribe: um estudo infodemiológico com o uso do Google Trends

Carlos Jesús Aragón-Ayala [1]
Julissa Copa-Uscamayta [1]
Luis Herrera [1]
Frank Zela-Coila [1]
Cender Udai Quispe-Juli [2]

Abstract

Infodemiology has been widely used to assess epidemics. In light of the recent pandemic, we use Google Search data to explore online interest about COVID-19 and related topics in 20 countries of Latin America and the Caribbean. Data from Google Trends from December 12, 2019, to April 25, 2020, regarding COVID-19 and other related topics were retrieved and correlated with official data on COVID-19 cases and with national epidemiological indicators. The Latin American and Caribbean countries with the most interest for COVID-19 were Peru (100%) and Panama (98.39%). No correlation was found between this interest and national epidemiological indicators. The global and local response time were 20.2 ± 1.2 days and 16.7 ± 15 days, respectively. The duration of public attention was 64.8 ± 12.5 days. The most popular topics related to COVID-19 were: the country's situation (100 ± 0) and coronavirus symptoms (36.82 ± 16.16). Most countries showed a strong or moderated (r = 0.72) significant correlation between searches related to COVID-19 and daily new cases. In addition, the highest significant lag correlation was found on day 13.35 ± 5.76 (r = 0.79). Interest shown by Latin American and Caribbean countries for COVID-19 was high. The degree of online interest in a country does not clearly reflect the magnitude of their epidemiological indicators. The response time and the lag correlation were greater than in European and Asian countries. Less interest was found for preventive measures. Strong correlation between searches for COVID-19 and daily new cases suggests a predictive utility.

Coronavirus Infections; Epidemiology; Internet

Correspondence
C. J. Aragón-Ayala
Universidad Nacional de San Agustín de Arequipa.
Santa Catalina Nro. 117, Cercado, Arequipa / Arequipa –
04009, Perú.
caragonay@unsa.edu.pe

[1] *Universidad Nacional de San Agustín de Arequipa, Arequipa, Perú.*
[2] *Universidad Peruana Cayetano Heredia, Lima, Perú.*

Introduction

The world is being affected by coronavirus disease 2019 (COVID-19) which is caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). According to the World Health Organization (WHO) the pandemic comes with a rapid and massive diffusion of information, a phenomenon known as infodemic [1]. Frequently, this information is incorrect, causing panic and dangerous behaviors in the population, affecting the effectiveness of public interventions and worsening the situation in low- and middle-income countries. Therefore, infodemiology rises as a tool to combat infodemics [2].

Infodemiology is the science of distribution and determinants of information on the Internet or a population, with a final goal to inform about public health and policies [2,3]. Its main sources of information are social networking sites and search engines, with web platforms Twitter and Google being the most popular [2]. These applications allow a real time analysis [3], facilitating outbreak prediction, especially during pandemics [2]. Further analysis informs about health-related behavior, attitudes and knowledge in the population [3]. This serves public health professionals who need to keep close surveillance on quick changes in demand for information whether it is to calm down the population or to detect outbreaks, using the peaks in Internet searches as an early predictor.

One of the most reliable tools for infodemiology is Google Search, due to its almost universal use [3,4,5]. Google Trends (https://trends.google.com) is a public access tool which analyzes daily searches on Google Search, generating data from a specific time and zone as a Relative Search Volume (RSV) [6]. The use of Google Trends in surveillance of the public interest for an outbreak is promising, because it responds the principal queries and doubts around the population. Google Trends has been used in studies about infectious diseases, mental health, general diseases, public behavior and its seasonality [7,8,9,10,11].

Other studies about COVID-19 and Google Trends searches have been conducted in China [7,12], Taiwan [11], Iran [13], United States, United Kingdom, Canada, Ireland, Australia, New Zealand [14], and Colombia [15]. Aiming to expand related knowledge in Latin American and Caribbean countries, and to improve the previously utilized methods, we analyzed COVID-19 related queries and their relationship with epidemiological indicators such as daily incidence, lethality, and deaths by COVID-19, for each country belonging to Latin American and Caribbean. Additionally, the response time and duration of the public attention.

Material and methods

Design and study period

Data from Google Trends [16] and the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (United States) [17,18] were utilized for topics related to COVID-19. The study period was from December 30, 2019 (the day before the first notification to the WHO of disease by COVID-19 in Wuhan, China) to April 25, 2020, for each one of the 20 Latin American and Caribbean countries which were studied (Argentina, Bolivia, Brazil, Chile, Colombia, Costa Rica, Cuba, Ecuador, El Salvador, Guatemala, Honduras, Mexico, Nicaragua, Panama, Paraguay, Peru, Puerto Rico, Dominican Republic, Uruguay, and Venezuela). Data were collected from April 27, 2020 to April 30, 2020.

Google Trends

Google Trends allows access to a variety of queries done with the Google engine. Its results are anonymous, categorized (by related searches) and aggregated. This allows it to reflect the worldwide interest on a particular subject [19,20]. Google Trends normalizes data to ease comparison between terms, as well as normalization on a temporal and geographical level. Normalized data is calculated by dividing the number of searches related to a term by the total of searches done in Google, having previously selected a specific place and time range. Results are scaled from 0 to 100, resulting in a RSV [11,14,19,20,21]. RSV of 100 represents the peak of interest in a particular subject within the chosen area and time, indirectly correcting bias in absolute measurements such as Internet access or population

size, which varies between countries [21]. Search terms on Google Trends are not case sensitive, but results can be affected by the use of accentuation. Combined search terms can be created by merging two or more terms with the use of a plus (+), which works as the Boolean operator "OR". Nonetheless, a limited number of characters (until 100 characters) can be used to form these terms. The comparison between terms (simple or combined) is limited to five.

Search terms

For the general topic of COVID-19 we used the combined term "coronavirus + COVID-19 + SARS-CoV2 + nuevo coronavirus + 2019-nCoV" in Spanish; and: "coronavirus + coronavírus + COVID-19 + SARS-CoV2 + novo coronavirus + novo coronavírus + 2019-nCoV" in Portuguese.

Topics of possible public concern were identified by researchers (C.J.A.A., J.C.U., L.H., F.Z.C.) by searching the most frequently asked questions about COVID-19 as referred by the Centers for Disease Control and Prevention (CDC; United States) and the WHO, as well as identifying them in repetitive commentaries on televised news, newspapers and social networking sites. All researchers conducted short unstructured interviews in a convenience sample of twenty people not involved in health care, to identify their main concerns about COVID-19. All previous procedures were done between April 15, 2020 to April 25, 2020. Then, for each identified topic, the most popular search terms for every country were found using the tools *Related queries – Top* and *Related queries – Rising* both available on the Google Trends page. Then, several combinations of these terms were repeated in Google Trends to obtain the most exhaustive combined search term. One of the explored topics was a promising treatment, in this case we used hydroxychloroquine, having the largest public interest, as well as the largest number of clinical trials [22].

Procedures

The researchers (J.C.U., L.H., F.Z.C.) downloaded data in a CSV (comma-separated values) format from Google Trends, which was then imported into Microsoft Excel 2019 (https://products.office.com/). First, the general term COVID-19 was searched by combining Spanish and Portuguese terms: "coronavirus + COVID-19 + SARS-CoV2 + nuevo coronavirus + 2019-nCoV + novo coronavírus" with "Worldwide" option. Then, a worldwide ranking of mean RSV was obtained per country. Data from Latin American and Caribbean countries was extracted, and each mean RSV was standardized on a scale from 0 to 100, to facilitate the analysis. Also, for every country, the RSV time series were extracted from the general term of COVID-19 in the study period. Secondly, for each country we identify the five topics with the highest mean RSV by conducting several searches in Google Trends. This way, searches were ranked in order of popularity using the mean RSV for every country.

Global and local response time and duration of public attention

Global response time was defined as the time between the day when the WHO was notified of the outbreak in Wuhan, and the start of the RSV related to COVID-19 [14]. The duration of public attention is the time between the beginning of the RSV of the general terms related to COVID-19 and its maximum peak [14]. Local response time was defined as the time between the date in which the first case in each country and the maximum peak of searches.

Epidemiological data of COVID-19

The CSSE database is publicly available on GitHub (https://github.com/) [17,18], which allows a real-time record of confirmed cases, deaths, and recoveries of COVID-19 from every country to be maintained, using DXY, a platform directed by the Chinese medical community (https://portal.dxy.cn/), as its main data source. Data is updated in a semi-automatic manner, monitoring and requesting information from government websites, as well as the review of social network sites and local news. In addition, it shows proper alignment with the daily reports of the WHO and the Chinese Center for Disease Control and Prevention [18].

New cases and daily deaths were calculated with the accumulated data. In this way, the epidemiological indicators analyzed in this study were: total confirmed cases, mean new daily cases, mean incidence (per 100 inhabitants), total deaths, mortality (per 100,000 inhabitants), lethality rate ([cumulative deaths/cumulative confirmed cases]*100) and the percentage of total confirmed cases by total population. United Nations total population data by country were used [23].

Statistical analysis

The RSV values reported as "< 1" were replaced by 0.5 to facilitate analysis, as performed in a previous study [24]. The arbitrary value used in this replacement would not significantly alter the analysis as long as it is between 0 and 1. Analysis and graphs were done in Stata 14 (https://www.stata.com). A linear time series analysis of the RSV related to COVID-19 is presented, as well as new cases, daily deaths, and main concerns. Furthermore, scatterplot graphs were made utilizing the mean RSV by country and its epidemiological indicators.

Spearman's correlation coefficient was calculated (due to the non-normality of the data), considering weak ($r \leq 0.35$), moderate (0.36-0.67) and strong (0.68-1.00) [25] and considering $p < 0.05$, significant. Correlation was calculated between mean RSV and COVID-19 per country and its epidemiological indicator. Spearman's analysis of lag correlation ($p < 0.05$) was performed between the number of daily new cases and the daily RSV on COVID-19 per country, to determine if an increase in the RSV is correlated with a subsequent increase in new cases of COVID-19, as previous studies showed [11,26].

Ethical considerations

This study was guided by the principles of scientific and ethical integrity. Its approval by an ethics committee was not necessary because the data used is anonymous and freely accessible.

Results

Interest in COVID-19 in Latin American and Caribbean countries

The highest public interest (expressed as mean RSV) in COVID-19 between December 30, 2019 to April 25, 2020 was shown by Peru (100%), followed by Panama (98.39%), Colombia (83.65%), Paraguay (82.65%), Uruguay (80.65%), Argentina (80.65%), Bolivia (77.42%), Ecuador (75.81%), Costa Rica (72.58%), Chile (70.97%), Guatemala (70.97%), Venezuela (66.13%), Mexico (66.13%), Dominican Republic (58.06%), Brazil (46.77%), Cuba (1.61%), El Salvador (1.61%), Honduras (1.61%), Nicaragua (1.61%) and Puerto Rico (1.61%). The 90% of South American countries and 50% of Central American countries (Mexico, Dominican Republic, Guatemala, Costa Rica, and Panama) present over 50% mean RSV.

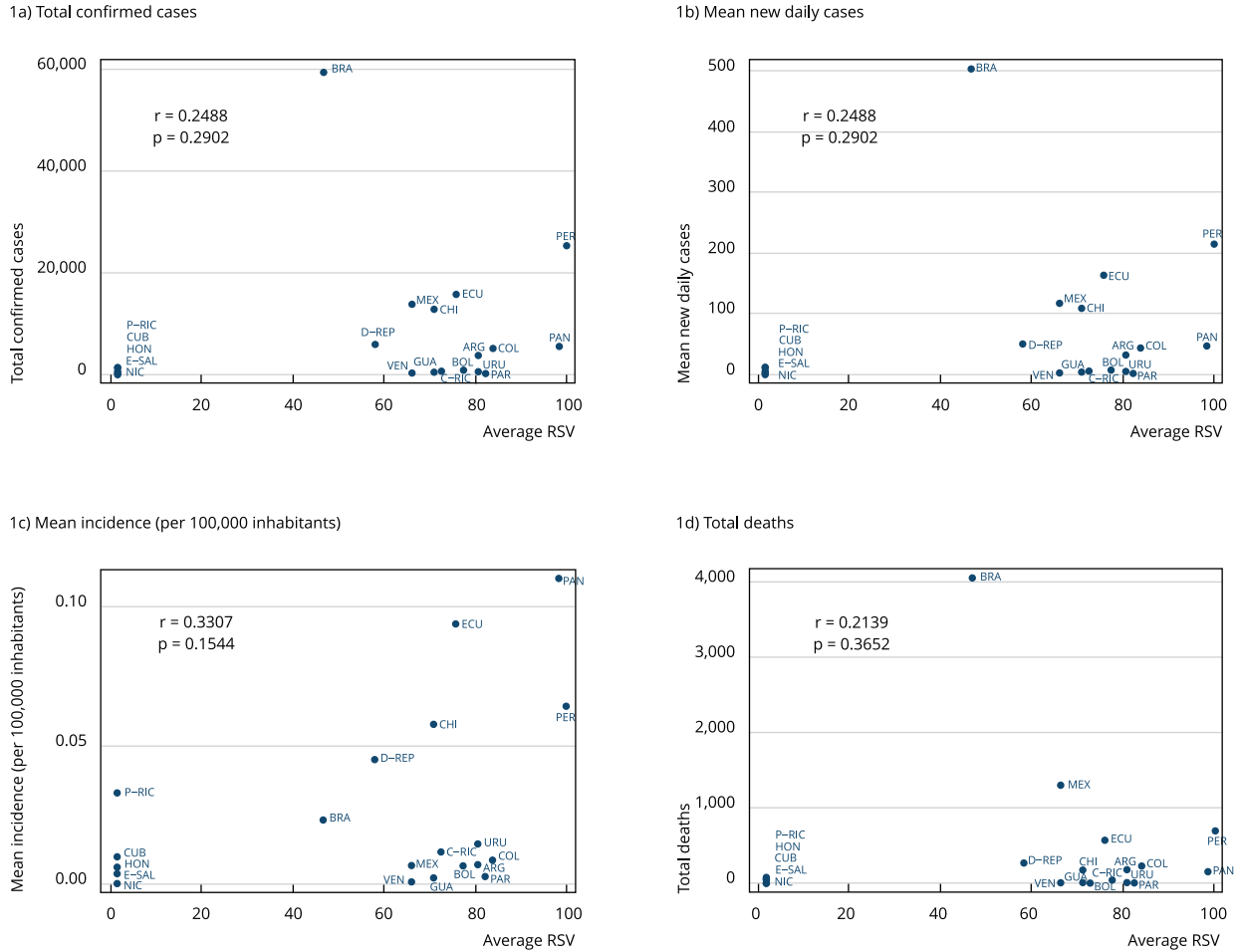### Relationship between epidemiological indicators and interest about COVID-19

No statistically significant correlation was found between the mean RSV of each country and epidemiological indicators such as: percentage of the total population that presented COVID-19, mean incidence per 100,000 inhabitants, observed lethality rate, total confirmed cases, mean daily new cases, accumulated deaths, and deaths per 100,000 inhabitants (Figure 1).

Trends of interest for COVID-19 and its relationship with daily new cases and deaths

In most countries, the presence of three main search peaks with variable magnitude were observed approximately on January 30, 2020 (± 18 days), February 26, 2020 (± 8 days) and March 19, 2020 (± 10 days), respectively. The mean magnitude of these peaks is around 18%, 25% and 100% for the first, second, and third peak respectively. Furthermore, a general decrease in searches can be observed in most countries. For more details, see Figure 2.

Figure 1

Relative Search Volume (RSV) for COVID-19 and epidemiological indicators in Latin American and Caribbean countries. Data from December 30, 2019 to April 25, 2020.

1a) Total confirmed cases

1b) Mean new daily cases

1c) Mean incidence (per 100,000 inhabitants)

1d) Total deaths

(continues)

Global and local response time, and duration of public attention

The mean global and local response time were 20.2 ± 1.2 days and 16.7 ± 15 days, respectively. The duration of public attention was 64.8 ± 12.5 days. The local response time was close to zero in Cuba, El Salvador, Guatemala, Nicaragua, Panama, Puerto Rico, Uruguay and Venezuela.

Interest in topics related to COVID-19 per country

The most interesting topics related to COVID-19 were the country's situation regarding coronaviruses (100.00 ± 0.00), coronavirus symptoms (36.82 ± 16.16), mask (23.55 ± 16.43), coronavirus definition (18.71 ± 10.60), quarantine and social isolation (10.69 ± 6.92). Additionally, the mean interest on other subjects was lower: vaccine (5.72 ± 2.35), hand sanitizer (5.60 ± 6.85), coronavirus prevention (5.32 ± 3.85), hand washing (3.46 ± 4.38), medical treatment (3.40 ± 1.79), coronavirus transmission

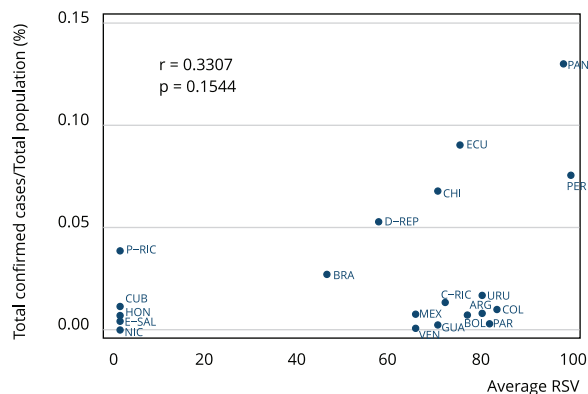**Figure 1 (continued)**

1e) Mortality (per 100,000 inhabitants)
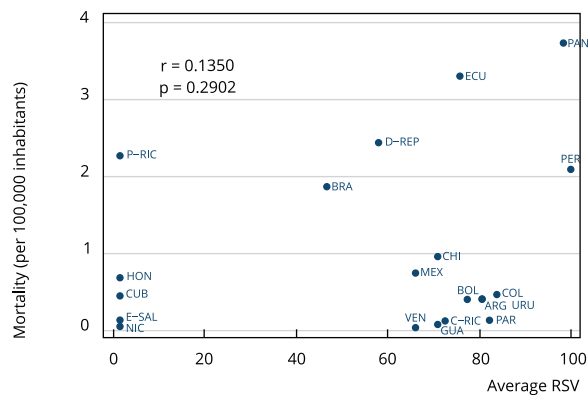
1f) Lethality rate
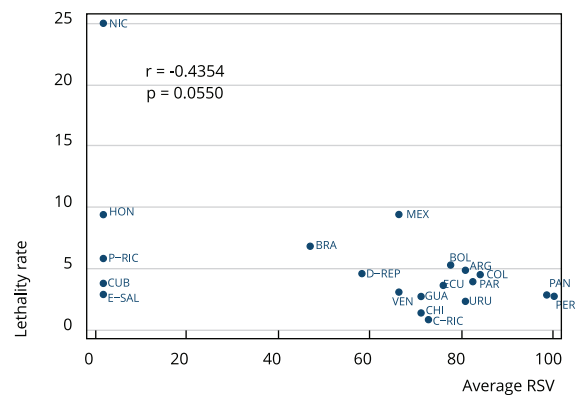
1g) Total confirmed cases/Total population (%)

p: p-value; r: Spearman's rank correlation coefficient.

Countries: ARG: Argentina; BOL: Bolivia; BRA: Brazil; CHI: Chile; COL: Colombia; C-RIC: Costa Rica; CUB: Cuba; D-REP: Dominican Republic; ECU: Ecuador; E-SAL: El Savador; GUA: Guatemala; HON: Honduras; MEX: Mexico; NIC: Nicaragua; PAN: Panamá; PAR: Paraguay; PER: Peru; P-RIC: Puerto Rico; URU: Uruguay; VEN: Venezuela.

(3.04 ± 1.53), coronavirus tests (1.90 ± 1.23), coronavirus origin (1.79 ± 1.14), attention to coronavirus (1.64 ± 0.82), risk population (0.86 ± 0.89), cleaning and disinfection against coronavirus (0.51 ± 0.31), promising treatment (hydroxychloroquine) (0.41 ± 0.35), home or natural remedies (0.28 ± 0.31), mental health (0.23 ± 0.19). Table 1 presents these concerns for each country.

Trends of the 5 main topics related to COVID-19 by countries

In most countries (17/20) a first peak of searches was observed near January 30, 2020 for topics about the country's situation, symptoms, definition, while searches for prevention, face masks, quarantine, hand-washing and gel alcohol remain constant. However, since the last week of February, a first irregular increase in most searches (highest peak for the country's situation and symptoms) was found and a later decrease in other searches, with some exceptions (Figure 3).
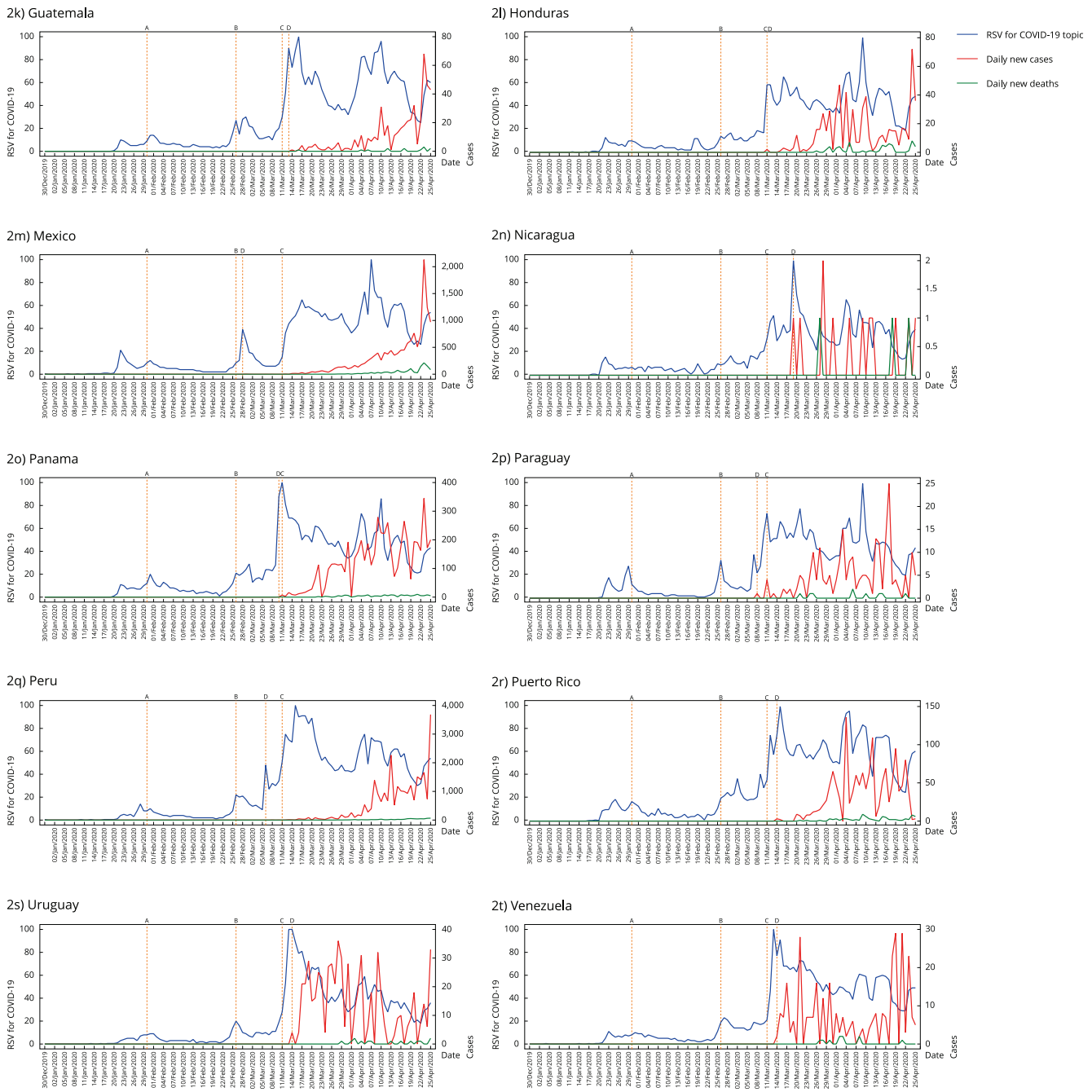
Figure 2

Relative Search Volume (RSV) for COVID-19, daily number of new cases and daily number of deaths by COVID-19 in Latin American and Caribbean countries. Data from December 12, 2019 to April 25, 2020.
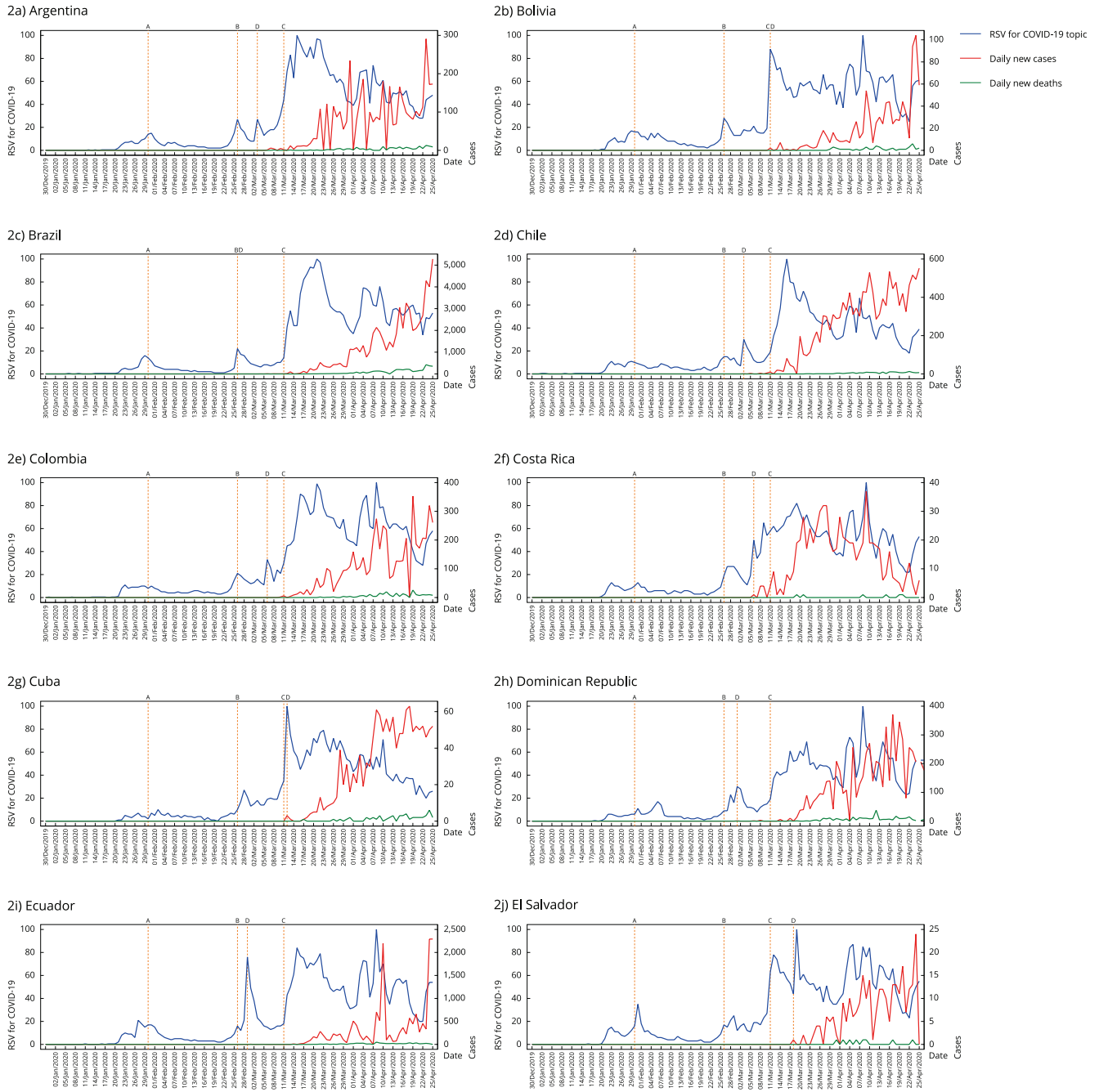
(continues)

Figure 2 (continued)

A: declaration of a Public Health Emergency of International Concern (PHEIC) by the World Health Organization (WHO); B: first case of COVID-19 in Latin America (Brazil); C: declaration of a pandemic by WHO; D: first confirmed case in each country.

Table 1

Mean Relative Search Volume (RSV) for topics related to COVID-19 in Latin American and Caribbean countries.

Topics	RSV									
	ARG	**BOL**	**BRA**	**CHI**	**COL**	**C-RIC**	**CUB**	**ECU**	**E-SAL**	**GUA**
Coronavirus definition	21	21	13	7	10	9	1	14	20	16
Symptoms	30	33	43	28	19	38	6	37	47	24
Country's situation about COVID-19	100	100	100	100	100	100	100	100	100	100
Coronavirus transmission	6	5	1	3	2	2	0	3	3	2
Coronavirus prevention	6	6	8	2	2	2	1	3	4	4
Mask	33	21	67	2	7	15	1	24	33	24
Hand washing	3	4	1	21	2	3	0	3	3	1
Hand sanitizer	16	7	27	2	0	9	0	1	13	5
Vaccine	8	7	8	4	4	4	3	5	7	3
Quarantine, and social isolation	25	24	3	9	11	4	1	7	20	6
Medical treatment	3	4	4	2	1	1	3	2	4	3
Home or natural remedies	0	0	0	0	0	0	0	0	0	0
Cleaning and disinfection against coronavirus	1	1	0	0	0	0	0	1	1	0
Attention to coronavirus	3	1	1	1	2	1	0	1	2	2
Mental health	0	0	0	0	0	0	0	0	0	0
High-risk population	4	1	2	1	0	1	0	1	0	0
Coronavirus origin	2	4	1	1	1	1	0	1	2	2
Coronavirus tests	1	2	3	3	1	2	0	2	1	1
Promising treatment (hydroxychloroquine)	0	1	1	0	0	0	0	1	0	0
Topics	**HON**	**MEX**	**NIC**	**PAN**	**PAR**	**PER**	**P-RIC**	**D-REP**	**URU**	**VEN**
Coronavirus definition	28	21	31	13	25	11	22	39	8	42
Symptoms	44	42	56	36	40	21	55	82	23	34
Country's situation about COVID-19	100	100	100	100	100	100	100	100	100	100
Coronavirus transmission	3	3	6	4	5	2	3	3	2	4
Coronavirus prevention	11	4	11	4	6	4	3	7	2	16
Mask	30	42	33	21	9	13	38	41	7	11
Hand washing	2	4	4	2	5	4	3	2	1	1
Hand sanitizer	2	1	7	1	10	2	0	1	7	1
Vaccine	8	4	11	7	5	3	5	10	5	4
Quarantine, and social isolation	8	8	3	14	5	14	11	15	7	18
Medical treatment	6	2	7	3	3	1	6	7	2	4
Home or natural remedies	0	0	0	0	1	0	1	1	0	0
Cleaning and disinfection against coronavirus	0	0	1	0	1	0	1	1	1	0
Attention to coronavirus	1	1	2	2	1	1	4	3	3	1
Mental health	0	0	0	0	0	0	1	0	0	0
High-risk population	0	0	1	0	1	1	1	1	1	1
Coronavirus origin	3	1	3	2	2	1	2	2	1	4
Coronavirus tests	2	2	2	2	1	2	5	4	1	1
Promising treatment (hydroxychloroquine)	1	0	1	0	0	0	0	1	0	0

Countries: ARG: Argentina; BOL: Bolivia; BRA: Brazil; CHI: Chile; COL: Colombia; C-RIC: Costa Rica; CUB: Cuba; D-REP: Dominican Republic; ECU: Ecuador; E-SAL: El Savador; GUA: Guatemala; HON: Honduras; MEX: Mexico; NIC: Nicaragua; PAN: Panamá; PAR: Paraguay; PER: Peru; P-RIC: Puerto Rico; URU: Uruguay; VEN: Venezuela.
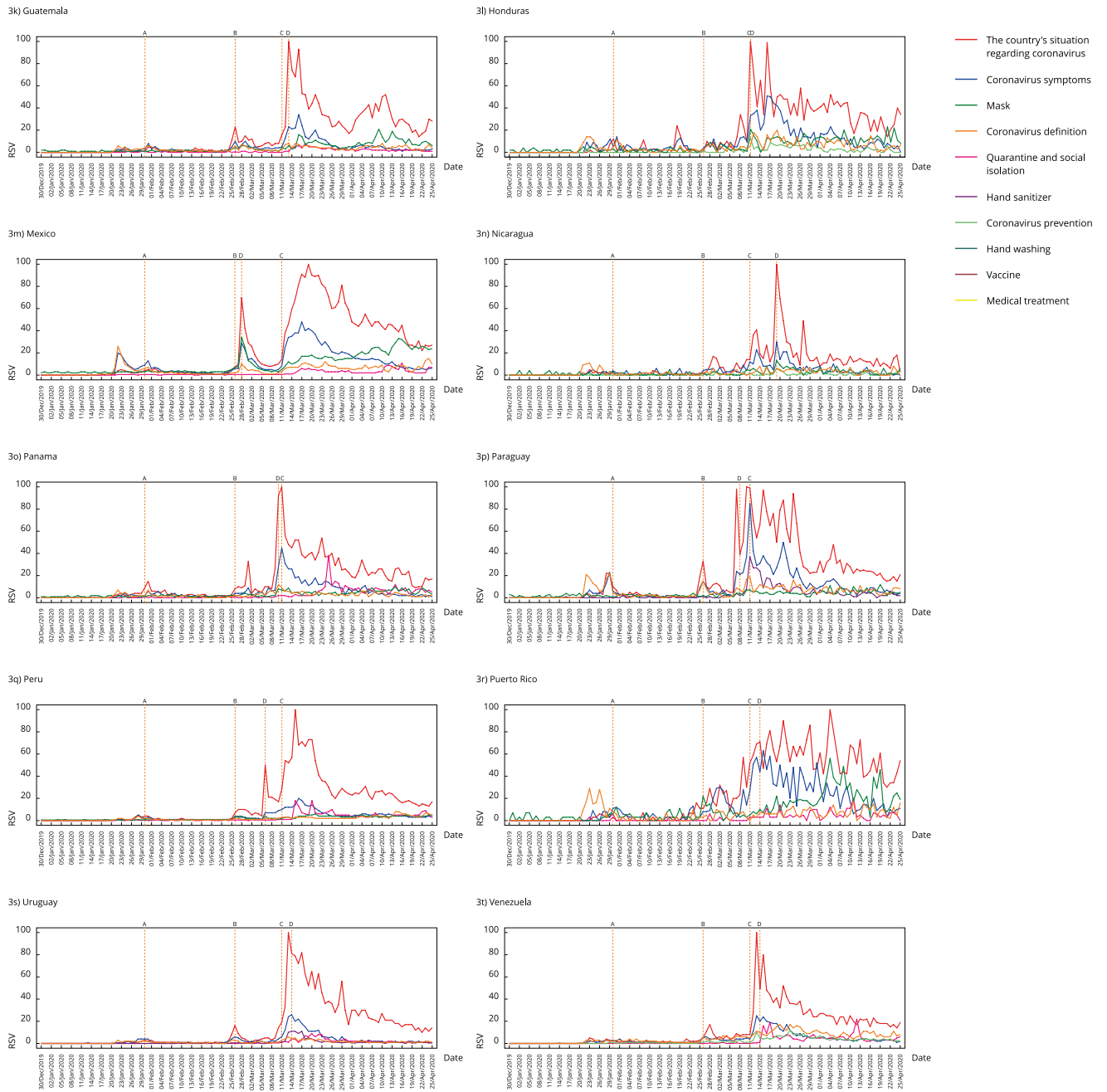
Figure 3

Relative Search Volume (RSV) for top five topics related to COVID-19 in Latin American and Caribbean countries. Data from December 12, 2019 to April 25, 2020.
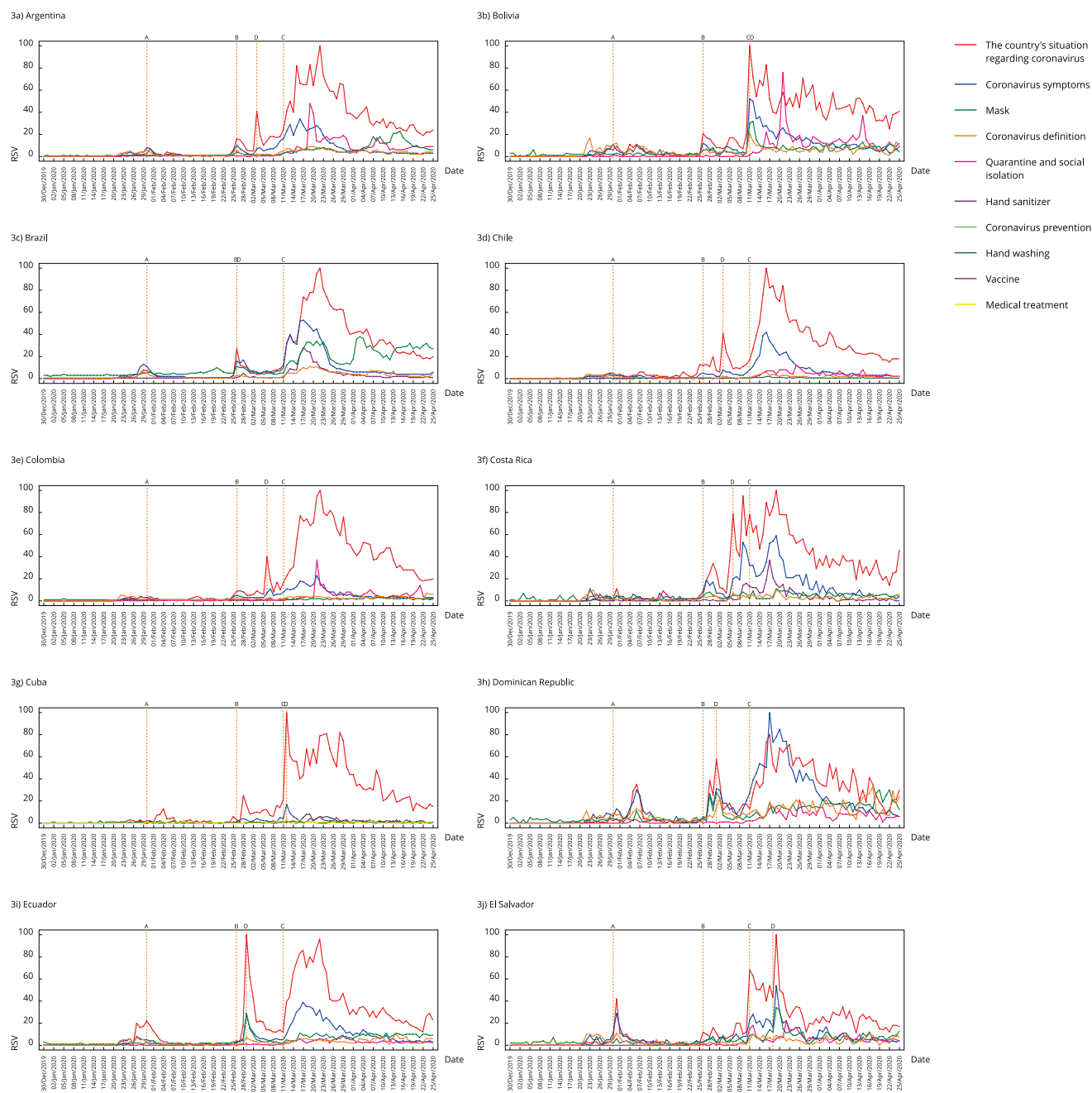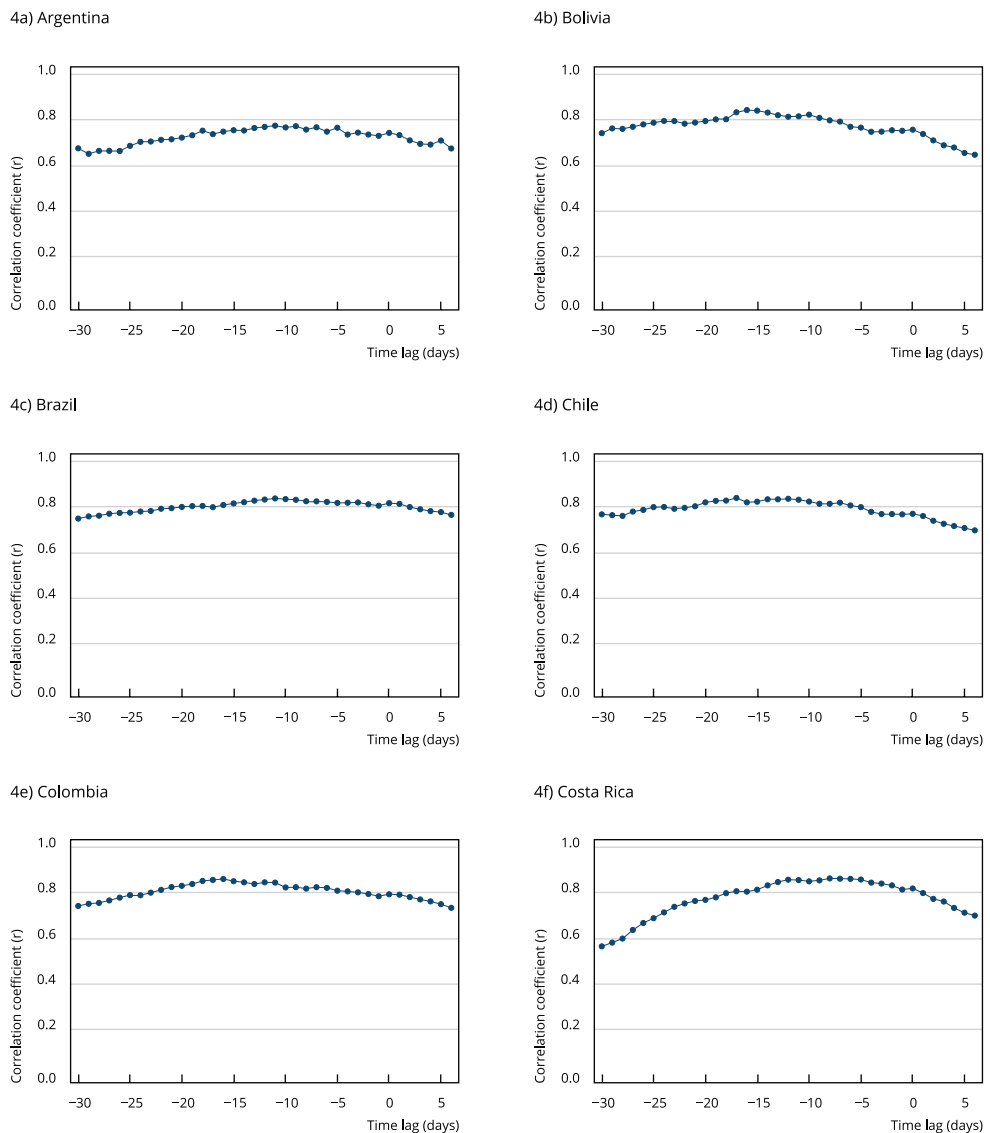

(continues)

Figure 3 (continued)

A: declaration of a Public Health Emergency of International Concern (PHEIC) by the World Health Organization (WHO); B: first case of COVID-19 in Latin America (Brazil); C: declaration of a pandemic by WHO; D: first confirmed case in each country.

Lag correlation between daily new cases and Internet queries for topic regarding COVID-19

The correlation between daily new cases and RSV for COVID-19 topic was significant (p < 0.001) and strong in eighteen out of twenty countries (r ~ 0.72), moderate in El Salvador (r = 0.61) and weak in Nicaragua (r = 0.35). In the lag period, all correlations were significant (p < 0.05), with a high correlation (r = 0.79) 5.76 ± 13.35 days earlier for daily new cases (Figure 4).
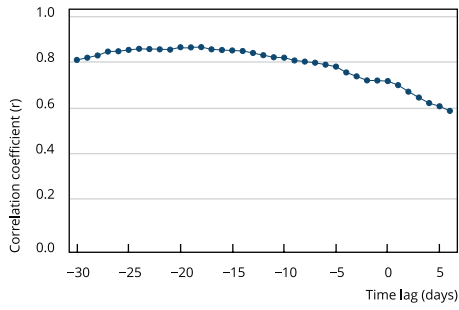
Figure 4

Lag correlations of Relative Search Volume (RSV) for COVID-19 and daily new cases in Latin American and Caribbean countries. Data from December 12, 2019 to April 25, 2020.
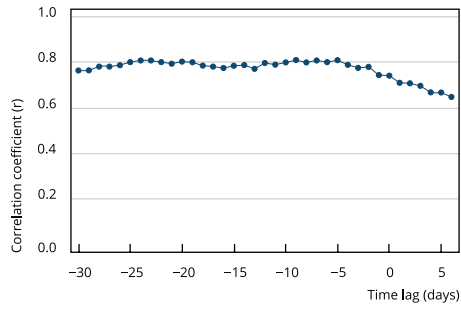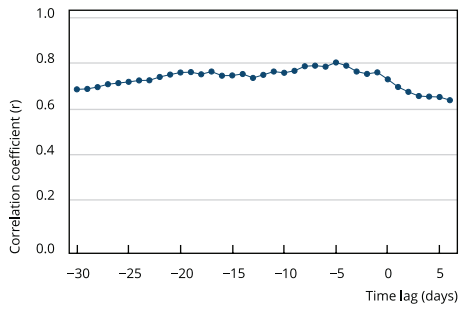

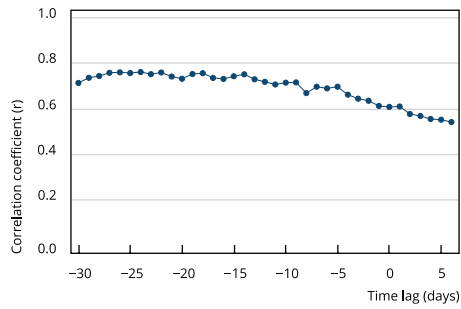(continues)

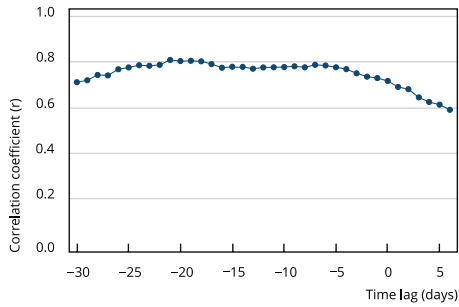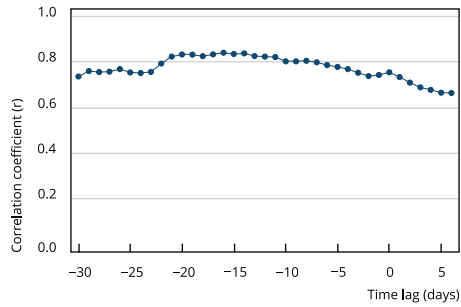**Figure 4 (continued)**

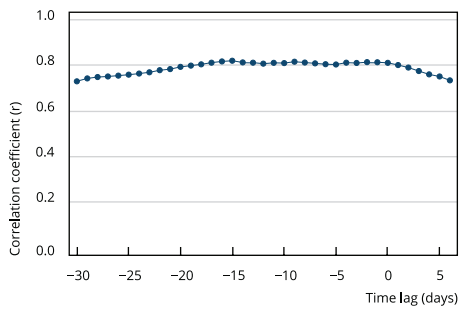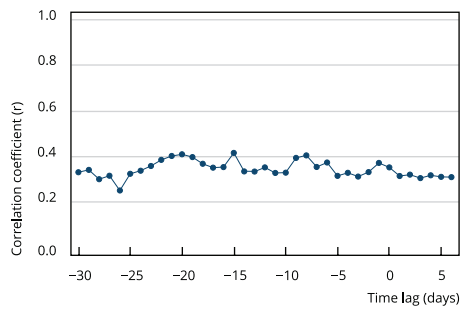4g) Cuba

4h) Dominican Republic
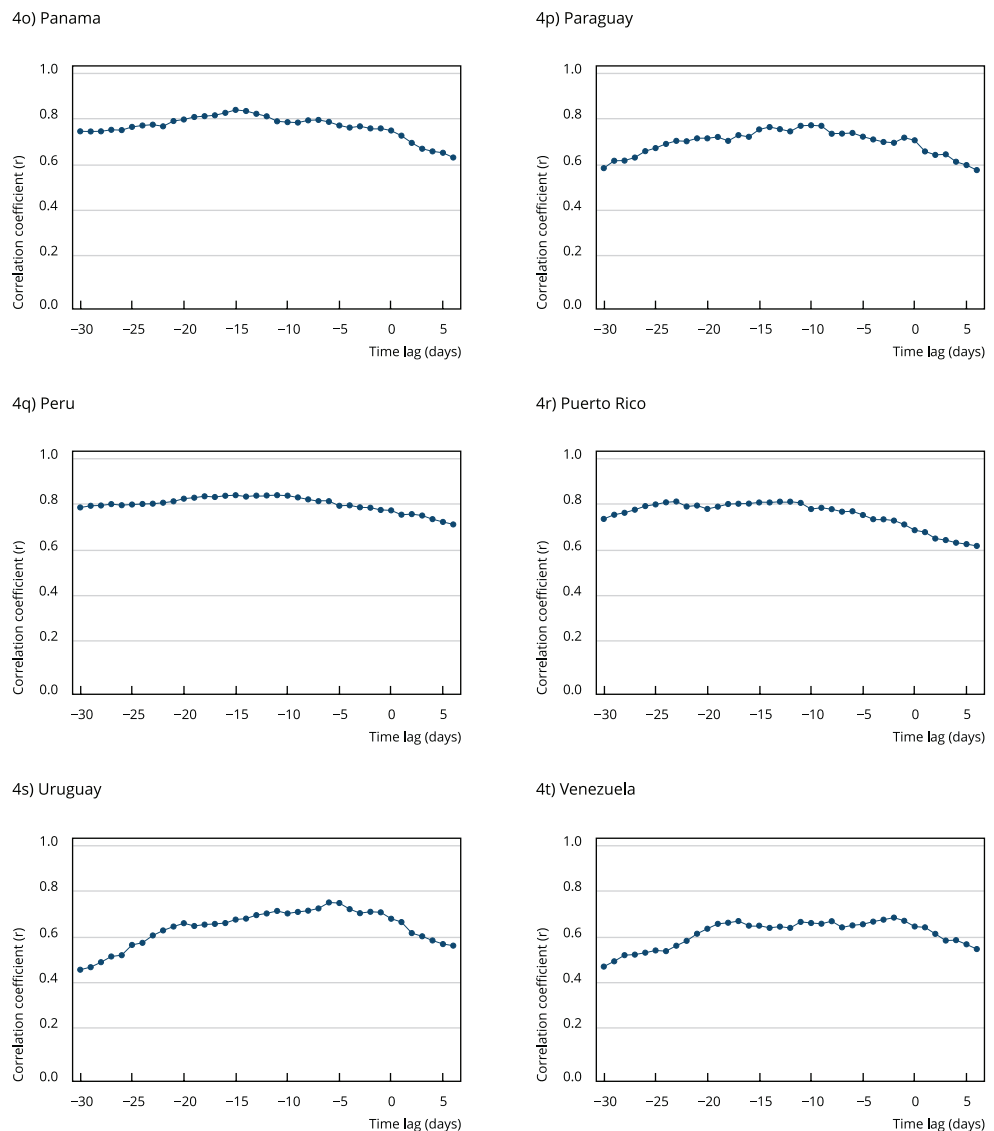
4i) Ecuador

4j) El Salvador

4k) Guatemala

4l) Honduras

4m) Mexico

4n) Nicaragua

(continues)

Figure 4 (continued)

4o) Panama

4p) Paraguay

4q) Peru

4r) Puerto Rico

4s) Uruguay

4t) Venezuela

Discussion

Interest about COVID-19 in Latin American and Caribbean countries, and its relationship with epidemiological indicators

Most Latin American and Caribbean countries showed high interest for COVID-19, except Cuba, El Salvador, Honduras, Nicaragua and Puerto Rico. These countries presented approximately 2% of the interest shown by Peru (the country with the highest interest). In a previous study, interest about COVID-19 was higher in countries with the most confirmed cases at that moment (Italy, France and Spain) [12]. This suggests that the number of confirmed cases would be a factor related to popular interest for COVID-19. However, we found no significant correlation between epidemiological indicators (such as total confirmed cases) and the mean relative interest by country. Possibly intervening factors

or external causes of this relationship were not explored, thus not allowed to find it. Some factors could be the proportion of the population with Internet access, dissemination of these indicators through official and unofficial media, popularity and use of different sources of information, spread of news via social network sites, concern for national and international reality, Human Development Index (HDI), and others [14,27]. Future studies should investigate the role of these and other factors to better understand the applicability of different tools such as Google Trends, in Latin American and Caribbean countries.

Trends of interest in COVID-19 and its relationship with daily new cases and deaths

The three major peaks in searches observed, in most countries, could be explained by their proximity to high impact national and international events. The first peak would be related to the declaration of a Public Health Emergency of International Concern (PHEIC) by the WHO (January 30, 2020) [28] and the declaration of national alerts by health authorities in countries such as Colombia [15]. This is consistent with findings in United States, United Kingdom, Canada, Ireland, New Zealand, Australia [14], China [7,12,29], Taiwan [11], Republic of Korea, Japan, Germany, Iran, Italy [30,31,32], and Spain [29], which also showed an increase in their searches near that date. The second main peak occurred around the confirmation of the first case of COVID-19 in Latin America that took place in Brazil (February 26, 2020) [33]. This could be explained by the impact of the notification of cases of a country in the interest and the searches of neighboring countries, as in nearby states of the United States [34]. Future studies should deeply analyze the impact of notable events in each country on their epidemiological indicators for a better comprehension. The third peak would correspond to the declaration of a pandemic by the WHO (March 11, 2020) [28], the confirmation of the first cases and deaths, and the establishment of mitigation or suppression policies during March and April in each country, which is consistent with previous studies [11,14,29,30,31,32,34,35,36]. A trend of progressive decrease in interest in COVID-19 is also observed in most countries, even despite the increase in cases and deaths, similar to prior investigations [12,29,30,34,36]. This limits the use of Google Trends as an epidemiological surveillance tool in the long term, as has been suggested in background studies [7,13,15,30,32,37]. This can be explained by an increase in the global awareness due to an effective national dissemination [12], or a decrease in the interest of people over time, similar to viral news of rapid rise and rapid fall [14,29,30,38]. However, particular features of each country would impact the trends evolution, therefore, further research is needed.

Global and local response time and duration of public attention

Latin American and Caribbean countries began their searches for COVID-19 approximately 20 days after the declared outbreak in Wuhan, similar to Ireland and New Zealand which started 21 days later [14], but different from the United States, United Kingdom, Canada and Australia, whose searches began 3, 10, 14, and 10 days later, respectively [14]. This difference might be explained by the confirmation of cases in neighboring countries, since the first case in the Americas region was confirmed on January 21, 2020 (United States). The duration of public attention was highly variable between countries, but on average (65 days) was higher than the United States, United Kingdom, Canada, Ireland, New Zealand and Australia, which presented 27, 23, 13, 12, 11 and 22 days, respectively [14]. It is possible that a shorter attention time favors the spread of the outbreak [14]. However, there would be an intrinsic dynamism according to the period analyzed. The local response time was on average 17 days, but it was close to zero in Cuba, El Salvador, Guatemala, Nicaragua, Panama, Puerto Rico, Uruguay and Venezuela. This could be due to the impact of the first confirmed cases on the public interest, even contributing to the low number of confirmed cases in these countries, although it will be necessary to test this hypothesis in forthcoming studies.

Interest in topics related to COVID-19 and Trends of the top 5 topics related to COVID-19

The main topics related to COVID-19 with the highest mean interest in Latin America countries were the country's situation regarding COVID-19, symptoms, mask, definition of coronavirus, quarantine, and social isolation. This is comparable to the United States, United Kingdom, Canada, Australia and New Zealand, where searches about situation in the country, definition of coronavirus and its symptoms were predominant [14,34], unlike Iran [13] or Taiwan [11] where topics related to prevention, such as hand washing and use of antiseptics, were the most popular. This way, despite the differences between countries, there are common concerns, emphasizing symptoms, but showing a lesser interest in preventive measures, at least in the early stages [34]. The evolution of trends follows a variable pattern by country. Some countries such as Bolivia and Mexico had a rapid growth in all their searches, including preventive measures. In others, preventive measures occurred days or weeks later, and to a lesser extent. In Brazil, searches about masks remain relatively constant and in moderate magnitude, different from other countries that showed lower interest. This could be due to effective communication interventions by the health authorities, as a consequence the population did not need to search for additional information on the Internet, explaining the decrease in interest. Another explanation could be ineffective interventions, thus generating a constant search trend.

Lag correlation between interest on COVID-19 and confirmed cases

A strong and positive correlation was found between searches for COVID-19 and the subsequent increase in the number of daily confirmed new cases after approximately 14 days. However, the correlation was moderate and weak in El Salvador and Nicaragua, respectively. One possible explanation is that the regions with the lowest correlation have less access to the Internet [11]. Compared with other studies, this lag time and the magnitude of correlation was greater or similar than other countries such as Taiwan ($r = 0.70$) from 1 to 3 days [11], China ($r = 0.96$) from 5 to 10 days [7], Germany, Austria, Australia, Japan, Egypt, United Kingdom ($r \sim 0.70$) with 11.5 days on average [30]. Furthermore, the instantaneous correlation (lag-0) in most Latin American and Caribbean countries was strong ($r \sim 0.75$), similar to other studies in Australia, Austria, Germany, Egypt, Japan, United Kingdom ($r \sim 0.60$), Spain, Italy [32], but different from countries such as the United States, United Kingdom, Canada, Ireland, Australia, and New Zealand, which presented weak and moderate correlations ($r \sim 0.4$) [14]. This variable correlation across studies could be explained by the fact that some studies used an accumulated number of cases for lag correlation analysis, but others used a number of daily new cases [30]. Another investigation, in Italy, found a correlation using more specific terms such as "cough" or "fever" symptoms, compared to general symptoms [39]. Although this suggests that people's interest is a possible predictor of first cases of an outbreak in each country [7], future studies should control other factors to elucidate in a better way the predictive value of Google Trends searches. Our findings give an idea that the Latin American and Caribbean population had more time to obtain information about COVID-19, a period of time that should have been used by governments because in these circumstances people are more sensitive, participative and receptive to information [13].

Limitations

Our study, like other studies using Google Trends, has some limitations. We used only one data source (Google Trends) and it could generate a possible selection bias. The study units were the countries; therefore, important subnational factors could not be evaluated. The representativeness of the sample studied cannot be guaranteed due to the heterogeneous Internet access in Latin American and Caribbean countries. The complexity and dynamism of epidemiological data diminish external validity at the future stages of COVID-19 pandemic. There is a possibility that the selection of the search terms is incomplete due to linguistic complexity and variability, and cultural worldview of each country, despite having performed a systematic process of identification of terms. In addition, our results cannot be replicated because the algorithm of Google Trends is not publicly accessible and it could change over time [34].

Strengths

Despite its limitations, this is the first study to analyze and explore the quantitative and qualitative relationship of epidemiological indicators of COVID-19 and the concerns of population on this subject, with the largest number of countries studied until now (20 countries) and from the same geographical region (Latin America and the Caribbean). In addition, we used several terms related to topics regarding COVID-19, having previously verified that they were the most used through multiple tests, unlike studies in China [7], Taiwan [11], among others, that used one or a few terms, and even other studies did not mention or make the terms used accessible [14,15], making it difficult to analyze the validity and replicability of their results, as mentioned in methodological Google Trends studies [4,5]. Furthermore, to the best of our knowledge, this is the first study to compare the mean interest in COVID-19 and its relationship with several epidemiological indicators in each country, simulating an ecological analysis.

Recommendations

Future studies should explore the correlation between adjusted epidemiological data and other search terms. Additionally, other data sources such as social network sites, national call centers for COVID-19, and the media should be added to the analysis.

Conclusions

Google Trends can be used to know the current status and evolution of the queries, thoughts, concerns, and needs of the study population in multiple periods. Health-policy makers could utilize these brief moments of active searching of the population to provide relevant information for either preventive reasons or to fight infodemics [1]. Our findings indicate that health authorities need to provide information in a massive way days before the confirmation of cases in the country, since people would be looking for information at that time. In this way, authorities can anticipate infodemics, which in many cases distort reality, inducing mass panic and dangerous practices, which are deeply rooted in public awareness despite future valid information. The correlation found between the number of new cases and interest for COVID-19 could set a framework for future predictive models. The high response time values in Latin American and Caribbean countries compared to high-income countries provides us with an idea of the low awareness about an international issue, which would go unnoticed unless it became a very close threat.

In summary, interest for COVID-19 in Latin America and the Caribbean is high in most countries. Its national trends were influenced by international events, such as the first case reported in the United States and Brazil, or the declaration of the outbreak as a PHEIC, then declared as a pandemic by the WHO, as well as national events, for instance, first confirmed cases and deaths in its territory or in neighboring countries, and the implementation of mitigation or suppression policies to a lesser extent. However, the degree of interest for COVID-19 does not clearly reflect the magnitude of epidemiological indicators in a country. The response time and the lag correlation in Latin America and the Caribbean were greater than European and Asian countries, suggesting that they had more time to provide information, failing to take an advantage of this opportunity. It is remarkable that searches in most Latin American and Caribbean countries have focused little on preventive measures. In most countries, a strong correlation between searches for COVID-19 and the number of daily new cases was found, therefore, the number of searches could serve as a predictive indicator of new cases and future studies are necessary to prove this hypothesis. This is the first infodemiological study to approach most of the countries in a continent, and we recommend continuing with more studies with a macroregional focus.

Contributors

C. J. Aragón-Ayala contributed to the conceptualization, methodology, formal analysis, investigation, data curation, and writing. J. Copa-Uscamayta contributed to the conceptualization, methodology, investigation, data curation, and writing. L. Herrera and F. Zela-Coila contributed to the conceptualization, methodology, investigation, and writing. C. U. Quispe-Juli contributed to the conceptualization, methodology, and writing.

Additional informations

ORCID: Carlos Jesús Aragón-Ayala (0000-0001-9536-2247); Julissa Copa-Uscamayta (0000-0002-2268-8577); Luis Herrera (0000-0001-7634-3016); Frank Zela-Coila (0000-0003-2805-9162); Cender Udai Quispe-Juli (0000-0003-0633-8339).

References

1. Tangcharoensathien V, Calleja N, Nguyen T, Purnat T, D'Agostino M, Garcia-Saiso S, et al. Framework for managing the COVID-19 infodemic: methods and results of an online, crowdsourced WHO technical consultation. J Med Internet Res 2020; 22:e19659.
2. Mavragani A. Infodemiology and infoveillance: scoping review. J Med Internet Res 2020; 22:e16206.
3. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. J Med Internet Res 2009; 11:e11.
4. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of Google Trends in health care research: a systematic review. PLoS One 2014; 9:e109583.
5. 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:e270.
6. 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; 7:185-9.
7. Li C, Chen LJ, Chen X, Zhang M, Pang CP, Chen H. Retrospective analysis of the possibility of predicting the COVID-19 outbreak from internet searches and social media data, China, 2020. Euro Surveill 2020; 25:2000199.
8. Choi SB, Kim J, Ahn I. Forecasting type-specific seasonal influenza after 26 weeks in the United States using influenza activities in other countries. PLoS One 2019; 14:e0220423.
9. Husnayain A, Fuad A, Lazuardi L. Correlation between Google Trends on dengue fever and national surveillance report in Indonesia. Glob Health Action 2019; 12:1552652.
10. Bragazzi NL, Mahroum N. Google Trends predicts present and future plague cases during the plague outbreak in Madagascar: infodemiological study. JMIR Public Health Surveill 2019; 5:e13142.
11. Husnayain A, Fuad A, Su EC-Y. 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; 95:221-3.
12. Strzelecki A, Rizun M. Infodemiological study using Google Trends on coronavirus epidemic in Wuhan, China. International Journal of Online and Biomedical Engineering 2020; 16:139-46.
13. 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:e18828.

14. Hu D, Lou X, Xu Z, Meng N, Xie Q, Zhang M, et al. More effective strategies are required to strengthen public awareness of COVID-19: evidence from Google Trends. J Glob Health 2020; 10:011003.

15. Ortiz-Martínez Y, Garcia-Robledo JE, Vásquez-Castañeda DL, Bonilla-Aldana DK, Rodriguez-Morales AJ. Can Google® trends predict COVID-19 incidence and help preparedness? The situation in Colombia. Travel Med Infect Dis 2020; 37:101703.

16. Google Trends. Explore Google Trends. https://trends.google.com/trends/explore (accessed on 16/May/2020).

17. Johns Hopkins University Center for Systems Science and Engineering. COVID-19 data repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University. https://github.com/CSSEGISandData/COVID-19 (accessed on 14/May/2020).

18. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. Lancet Infect Dis 2020; 20:533-4.

19. Rogers S. What is Google Trends data – and what does it mean? Medium 2016; 1 jul. https://medium.com/google-news-lab/what-is-google-trends-data-and-what-does-it-mean-b48f07342ee8.

20. Google Trends. FAQ about Google Trends data. https://support.google.com/trends/answer/4365533?hl=en&ref_topic=6248052 (accessed on 16/May/2020).

21. Boehm A, Pizzini A, Sonnweber T, Loeffler-Ragg J, Lamina C, Weiss G, et al. Assessing global COPD awareness with Google Trends. Eur Respir J 2019; 53:1900351.

22. Gazendam A, Nucci N, Ekhtiari S, Gohal C, Zhu M, Payne A, et al. Trials and tribulations: so many potential treatments, so few answers. Int Orthop 2020; [Epub ahead of print].

23. United Nations. World population prospects 2019. https://population.un.org/wpp/Download/Standard/Population/ (accessed on 16/Jul/2020).

24. 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: infoveillance study. J Med Internet Res 2020; 22:e19791.

25. Taylor R. Interpretation of the correlation coefficient: a basic review. J Diagn Med Sonogr 1990; 6:35-9.

26. Shin S-Y, Seo D-W, An J, Kwak H, Kim S-H, Gwack J, et al. High correlation of Middle East respiratory syndrome spread with Google search and Twitter trends in Korea. Sci Rep 2016; 6:32920.

27. Alicino C, Bragazzi NL, Faccio V, Amicizia D, Panatto D, Gasparini R, et al. Assessing Ebola-related web search behaviour: insights and implications from an analytical study of Google Trends-based query volumes. Infect Dis Poverty 2015; 4:54.

28. World Health Organization. Archived: WHO Timeline – COVID-19. https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19 (accessed on 30/May/2020).

29. 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 Health Surveill 2020; 6:e19702.

30. Effenberger M, Kronbichler A, Shin JI, 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; 95:192-7.

31. Rovetta A, Bhagavathula AS. COVID-19-related web search behaviors and infodemic attitudes in Italy: infodemiological study. JMIR Public Health Surveill 2020; 6:e19374.

32. 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; 44:117.

33. Ravelo J, Jerving S. COVID-19 – a timeline of the coronavirus outbreak. https://www.devex.com/news/sponsored/covid-19-a-timeline-of-the-coronavirus-outbreak-96396 (accessed om 03/Jul/2020).

34. Bento AI, Nguyen T, Wing C, Lozano-Rojas F, Ahn Y-Y, 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; 117:11220-2.

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 2020; 88:950-1.

36. Springer S, Menzel LM, Zieger M. Google Trends provides a tool to monitor population concerns and information needs during COVID-19 pandemic. Brain Behav Immun 2020; 87:109-10.

37. Yuan X, Xu J, Hussain S, Wang H, Gao N, Zhang L. Trends and prediction in daily new cases and deaths of COVID-19 in the United States: an internet search-interest based model. Explor Res Hypothesis Med 2020; 5:1-6.

38. Mavragani A. Tracking COVID-19 in Europe: infodemiology approach. JMIR Public Health Surveill 2020; 6:e18941.

39. Ciaffi J, Meliconi R, Landini MP, Ursini F. Google trends and COVID-19 in Italy: could we brace for impact? Intern Emerg Med 2020; 15:1555-9.

Resumen

La infodemiología ha sido utilizada ampliamente para evaluar epidemias. A la luz de la pandemia reciente, usamos datos de Google Search para investigar el interés en línea sobre COVID-19 y temas relacionados en 20 países de Latinoamérica y el Caribe. Los datos se obtuvieron de Google Trends, desde el 12 de diciembre 2019 al 25 de abril de 2020, respecto a la COVID-19 y otros temas relacionados, y se correlacionaron con datos oficiales sobre casos de COVID-19, así como con indicadores nacionales epidemiológicos. Los países latinoamericanos y del Caribe con mayor interés en la COVID-19 fueron: Perú (100%) y Panamá (98,39%). No se encontró correlación entre este interés y los indicadores nacionales epidemiológicos. Los tiempos de respuesta global y local fueron 20,2 ± 1,2 días and 16,7 ± 15 días, respectivamente. La duración de la atención pública fue 64,8 ± 12,5 días. Los temas más populares relacionados con la COVID-19 fueron: la situación del país (100 ± 0) y síntomas de coronavirus (36,82 ± 16,16). La mayoría de los países mostraron una correlación significativa fuerte o moderada (r = 0,72) entre las búsquedas relacionadas con la COVID-19 y los nuevos casos diarios. Asimismo, se descubrió que la correlación de retraso significativa más alta se produjo el día 13,35 ± 5,76 (r = 0,79). El interés mostrado por países latinoamericanos y del Caribe sobre la COVID-19 fue alto. El grado de interés en línea en un país no refleja claramente la magnitud de sus indicadores epidemiológicos. El tiempo de respuesta y la correlación de retraso fueron mayores que en los países europeos y asiáticos. Se descubrió poco interés respecto a las medidas preventivas. Una fuerte correlación entre las búsquedas de COVID-19 y nuevos casos diarios sugiere una utilidad predictiva.

Infecciones por Coronavirus; Epidemiología; Internet

Resumo

A infodemiologia tem sido largamente utilizada para avaliar as epidemias. À luz da mais recente pandemia, utilizamos dados da Google Search para explorar o interesse online sobre COVID-19 e temas correlatos em 20 países da América Latina e Caribe. Foram recuperados e correlacionados dados do Google Trends entre 12 de dezembro de 2019 e 25 de abril de 2020, sobre COVID-19 e temas relacionados, com dados oficiais sobre casos de COVID-19 e com indicadores epidemiológicos nacionais. Os países latino-americanos e caribenhos com maior interesse na COVID-19 foram Peru (100%) e Panamá (98.39%). Não foi encontrada correlação entre esse interesse e os indicadores epidemiológicos nacionais. Os tempos de resposta global e local foram de 20,2 ± 1,2 dias e 16,7 ± 15 dias, respectivamente. A duração da atenção do público foi de 64,8 ± 12,5 dias. Os temas mais populares relacionados à COVID-19 foram: a situação do respectivo país (100 ± 0) e os sintomas relacionados ao novo coronavírus (36,82 ± 16,16). A maioria dos países mostrou uma associação significativa forte ou moderada (r = 0,72) entre buscas relacionadas à COVID-19 e casos novos diários. Além disso, a correlação defasada significativa mais alta foi encontrada no dia 13,35 ± 5,76 (r = 0,79). Os países latino-americanos e caribenhos demonstraram forte interesse na COVID-19. O grau de interesse na internet em um país não reflete claramente a magnitude de seus indicadores epidemiológicos. O tempo de resposta e a correlação defasada foram maiores que nos países europeus e asiáticos. Houve pouco interesse em medidas preventivas. A correlação forte entre busca por COVID-19 e casos novos diários sugere utilidade preditiva.

Infecções por Coronavírus; Epidemiologia; Internet