Global Infodemiology of COVID-19: Analysis of Google Web Searches and Instagram Hashtags

Alessandro Rovetta1*; Akshaya Srikanth Bhagavathula2*, PharmD

1Research and Disclosure Division, Mensana srls, Brescia, Italy
2Institute of Public Health, College of Medicine and Health Sciences, United Arab Emirates University, Al Ain, United Arab Emirates

*all authors contributed equally

Corresponding Author:
Alessandro Rovetta
Research and Disclosure Division
Mensana srls
Via Moro Aldo 5
Brescia, 25124
Italy
Phone: 39 3927112808
Email: rovetta.mresearch@gmail.com

Abstract

Background: Although “infodemiological” methods have been used in research on coronavirus disease (COVID-19), an examination of the extent of infodemic moniker (misinformation) use on the internet remains limited.

Objective: The aim of this paper is to investigate internet search behaviors related to COVID-19 and examine the circulation of infodemic monikers through two platforms—Google and Instagram—during the current global pandemic.

Methods: We have defined infodemic moniker as a term, query, hashtag, or phrase that generates or feeds fake news, misinterpretations, or discriminatory phenomena. Using Google Trends and Instagram hashtags, we explored internet search activities and behaviors related to the COVID-19 pandemic from February 20, 2020, to May 6, 2020. We investigated the names used to identify the virus, health and risk perception, life during the lockdown, and information related to the adoption of COVID-19 infodemic monikers. We computed the average peak volume with a 95% CI for the monikers.

Results: The top six COVID-19–related terms searched in Google were “coronavirus,” “corona,” “COVID,” “virus,” “corona virus,” and “COVID-19.” Countries with a higher number of COVID-19 cases had a higher number of COVID-19 queries on Google. The monikers “coronavirus ozone,” “coronavirus laboratory,” “coronavirus 5G,” “coronavirus conspiracy,” and “coronavirus bill gates” were widely circulated on the internet. Searches on “tips and cures” for COVID-19 spiked in relation to the US president speculating about a “miracle cure” and suggesting an injection of disinfectant to treat the virus. Around two thirds (n=48,700,000, 66.1%) of Instagram users used the hashtags “COVID-19” and “coronavirus” to disperse virus-related information.

Conclusions: Globally, there is a growing interest in COVID-19, and numerous infodemic monikers continue to circulate on the internet. Based on our findings, we hope to encourage mass media regulators and health organizers to be vigilant and diminish the use and circulation of these infodemic monikers to decrease the spread of misinformation.

(J Med Internet Res 2020;22(8):e20673) doi: 10.2196/20673

KEYWORDS
COVID-19; coronavirus; Google; Instagram; infodemiology; infodemic; social media

Introduction

Globally, the internet is an extremely important platform for obtaining knowledge and information about the coronavirus disease (COVID-19) pandemic [1-3]. The Google Trends tool provides real-time insights into internet search behaviors on various topics, including COVID-19 [4]. Social media platforms such as Facebook, Twitter, and Instagram allow users to communicate their thoughts, feelings, and opinions by sharing short messages. A unique aspect of social media data from Instagram is that image-based posts are accessible, and the use of topic-related hashtags allows access to topic-related information for users [5]. In general, there is a growing interest...
in examining social data to understand and monitor public behavior in real time [6,7].

Research on the internet and social data are called infodemiology or infoeillance studies [8]. Infodemiology is defined as “the science of distribution and determinants of information in an electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform public health and public policy” [9]. Although several studies have been conducted using infodemiological methods in COVID-19 research, a limited number of studies have examined the extent of COVID-19–related misinformation on the internet [10-14]. We are defining an infodemic moniker to be a term, query, hashtag, or phrase that generates or feeds the misinformation circulating on the internet. These monikers can profoundly affect public health communication, giving rise to errors in interpretation, misleading information, xenophobia, and fake news [12-17]. In this context, we aimed to investigate the internet search behaviors related to COVID-19 and the extent of infodemic monikers circulating on Google and Instagram during the current pandemic period.

Methods

We used Google Trends and Instagram hashtags to explore internet search activities and behaviors related to the COVID-19 pandemic from February 20, 2020, to May 06, 2020. We investigated the following: names used to identify the virus, health and risk perception, lifestyles during the lockdown, and information related to the adoption of infodemic monikers related to COVID-19. The complete list of terms used to identify the most frequently searched queries in Google and hashtag suggestions for Instagram are presented in Multimedia Appendix 1.

The obtained infodemic monikers are characterized as follows:

1. **Generic**: The moniker confuses, due to a lack of specificity.
2. **Misinformative**: The moniker associates a certain phenomenon with fake news.
3. **Discriminatory**: The moniker encourages the association of a problem with a specific ethnicity and/or geographical region.
4. **Deviant**: The moniker does not identify the requested phenomenon.
5. **Other specificities**: We kept two additional points for special cases that prove to be exceptionally serious.

To determine the severity of the various infodemic monikers circulating on the internet, each moniker was assigned 1 to 2 points on the infodemic scale (I-scale) ranging from 0 (minimum) to 10 (maximum). Based on the sum of the I-scale scores, the infodemic monikers were classified as follows: not infodemic (0), slightly infodemic (1), moderately infodemic (2-4), highly infodemic (5-8), and extremely infodemic (9-10).

To assign points, we have adopted the following procedure:

- **Generic**: 1 point is assigned if the keyword is a scientific term but gives rise to misunderstanding (eg, “COVID” instead of “COVID-19”); 2 points are assigned if the keyword is a combination of two scientific terms that can be confused with previously used terms (eg, “SARS-CoV” instead of “SARS-CoV-2” or “SARS-COVID” instead of “SARS-CoV-2” and “COVID-19”).
- **Misinformative**: 1 point is assigned if the keyword can lead to both fake news pointing to individuals (eg, “coronavirus Bill Gates”); 2 points are assigned if the keyword is used to spread misinformation using unrelated or not officially confirmed sources (eg, “coronavirus laboratory”).
- **Discriminatory**: 1 point is assigned if the keyword refers to a specific country and incites unfounded, racial fear (eg, “coronavirus China”); 2 points are assigned if the keywords explicitly target a specific ethnicity (eg, “Chinese coronavirus”).
- **Deviant**: 1 point is assigned if the keyword expresses opinions to influence public opinion (eg, “ban china” or “china app”); 2 points are assigned if the keyword expresses a particular attitude to influence the public (eg, “china puppets” or “savage WHO”).
- **Other specificities**: 1 additional point is assigned when the adoption of a certain moniker is associated with real facts but involves serious health or economic risks (eg, “uv coronavirus”); 2 additional points are assigned when the adoption of a certain moniker involves only health risks (eg, “no sew mask” or “anti-mask protest”).

For each search keyword considered, Google Trends provided normalized data in the form of relative search volume (RSV) based on search popularity scale ranging from 0 (low) to 100 (highly popular). Using these RSV values, we computed the average peak volume (APC) with a 95% CI (for a Gaussian distribution) during the study period.

Instagram, a platform for image-based posts with hashtags, was also screened. We retrieved content based on hashtags through image classifiers every 3-4 days during the study period. All irrelevant content was excluded. The data collected included contents posted on Instagram and self-reported user demographic information. No personal information, such as emails, phone numbers, or addresses, was collected. The data from the Instagram hashtags were collected manually through the Instagram-suggested tags associated with specific countries.

All data used in the study were obtained from anonymous open sources. Thus, ethical approval was not required.

Results

The top five COVID-19–related infodemic and scientific terms used in Google searches were “coronavirus,” “corona,” “COVID,” “virus,” “corona virus,” and “COVID-19” (Figure 1). The most frequently used keywords globally were “coronavirus” (APC=1378, 95% CI 1246-1537), followed by “corona” (APC=530, 95% CI 477-610) and “COVID” (APC=345, 95% CI 292-398). Several keywords related to COVID-19 were identified (Table 1); of the top 10, four had an I-scale value >4: “corona,” “virus,” “corona virus,” and “coronavirus China.”
**Figure 1.** Top scientific and infodemic names related to coronavirus disease (COVID-19) in Google searches globally. RSV: relative search volume.

![Google Trends RSV](image)

**Table 1.** Top infodemic and scientific Google searches related to coronavirus disease (COVID-19) globally.

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Total average peak volume</th>
<th>95% CI</th>
<th>I-scale(^a) score</th>
</tr>
</thead>
<tbody>
<tr>
<td>coronavirus</td>
<td>1378</td>
<td>1246-1537</td>
<td>4</td>
</tr>
<tr>
<td>corona</td>
<td>530</td>
<td>477-610</td>
<td>8</td>
</tr>
<tr>
<td>COVID</td>
<td>345</td>
<td>292-398</td>
<td>1</td>
</tr>
<tr>
<td>virus</td>
<td>239</td>
<td>212-292</td>
<td>6</td>
</tr>
<tr>
<td>corona virus</td>
<td>159</td>
<td>133-186</td>
<td>7</td>
</tr>
<tr>
<td>coronavirus Italy</td>
<td>54</td>
<td>45-62</td>
<td>4</td>
</tr>
<tr>
<td>COVID-19</td>
<td>53</td>
<td>45-60</td>
<td>0</td>
</tr>
<tr>
<td>coronavirus USA</td>
<td>32</td>
<td>29-36</td>
<td>4</td>
</tr>
<tr>
<td>coronavirus China</td>
<td>30</td>
<td>25-34</td>
<td>6</td>
</tr>
<tr>
<td>coronavirus Germany</td>
<td>23</td>
<td>20-27</td>
<td>4</td>
</tr>
<tr>
<td>corona Italy</td>
<td>13</td>
<td>12-14</td>
<td>7</td>
</tr>
<tr>
<td>corona Deutschland</td>
<td>12</td>
<td>10-14</td>
<td>7</td>
</tr>
<tr>
<td>SARS</td>
<td>9</td>
<td>8-10</td>
<td>5</td>
</tr>
<tr>
<td>corona China</td>
<td>9</td>
<td>7-11</td>
<td>9</td>
</tr>
<tr>
<td>corona Wuhan</td>
<td>1</td>
<td>0-2</td>
<td>9</td>
</tr>
<tr>
<td>SARS-CoV-2</td>
<td>1</td>
<td>0-1</td>
<td>0</td>
</tr>
</tbody>
</table>

\(^a\)I-scale: infodemic scale ranging from 0-10.

The country-wise dispersion of the scientific and infodemic names of COVID-19 used in Google searches are shown in **Figure 2.** Countries with a higher number of COVID-19 cases per 1 million people had greater Google search interest related to COVID-19 (eg, Italy, Spain, Ireland, Canada, France, and Qatar). These COVID-19–related search queries were significantly correlated with the incidence of COVID-19 cases in these countries (Pearson R=0.45, *P*<.001).
The top COVID-19 monikers related to fake news (eg, “coronavirus ozone,” “coronavirus laboratory,” and “coronavirus 5G”) that frequently circulated on the internet are presented in Table 2. Among these, “coronavirus conspiracy” and “coronavirus laboratory” had the highest I-scale scores globally (score=9). Additionally, the use of monikers with moderate-to-high infodemicity far exceeded the use of scientific names (Table 1)—52% of Google web searches were moderately infodemic (total APC=1487, 95% CI 1326-1647) and 34% highly infodemic (total APC=992, 95% CI 933-1050). The circulation of these infodemic monikers was further examined to understand the events associated with these searches. Infodemic monikers related to coronavirus origins, such as “SARS-CoV-2 made in the laboratory,” went viral (APC=41) when the National Associated Press Agency (Agenzia Nazionale Stampa Associata) of Italy posted a 2015 video about the origins of SARS-CoV-2 virus on March 25, 2020 [18]. In addition, the moniker reached breakout levels (RSV=100) on April 17, 2020, when the French Noble Prize winner Professor Luc Montagnier stated that the new coronavirus was the result of a laboratory accident in a high-security laboratory in Wuhan [19]. Detailed information on the different infodemic monikers and associated events are shown in Figure 3.

Table 2. Top global fake news–related Google searches on coronavirus disease (COVID-19).

<table>
<thead>
<tr>
<th>Keyword</th>
<th>Total average peak volume</th>
<th>95% CI</th>
<th>I-scale&lt;sup&gt;a&lt;/sup&gt; score</th>
</tr>
</thead>
<tbody>
<tr>
<td>coronavirus ozone</td>
<td>19</td>
<td>15-22</td>
<td>5</td>
</tr>
<tr>
<td>coronavirus laboratory</td>
<td>16</td>
<td>12-19</td>
<td>9</td>
</tr>
<tr>
<td>coronavirus 5G</td>
<td>10</td>
<td>8-13</td>
<td>8</td>
</tr>
<tr>
<td>coronavirus conspiracy</td>
<td>9</td>
<td>8-11</td>
<td>9</td>
</tr>
<tr>
<td>coronavirus bill gates</td>
<td>8</td>
<td>7-10</td>
<td>6</td>
</tr>
<tr>
<td>coronavirus milk</td>
<td>7</td>
<td>6-8</td>
<td>8</td>
</tr>
<tr>
<td>coronavirus military</td>
<td>4</td>
<td>4-5</td>
<td>8</td>
</tr>
<tr>
<td>coronavirus uv</td>
<td>3</td>
<td>3-4</td>
<td>5</td>
</tr>
</tbody>
</table>

<sup>a</sup>I-scale: infodemic scale ranging from 0-10.
The top searches related to health, precautions, and COVID-19 news are presented in Figure 4. Google searches related to COVID-19 news remained at the top throughout the pandemic period. However, searches related to “tips and cures” for COVID-19 spiked multiple times when the US president suggested that hydroxychloroquine (an unproven drug) was a “miracle cure” on April 4, 2020 (RSV=70) [20]; he also suggested injecting disinfectant to treat COVID-19 on April 24, 2020 (RSV=53) [21]. Other searches related to the use of medical masks and disinfectants (APC=23, 95% CI 21-25), lockdown (APC=19, 95% CI 16-22), and COVID-19 symptoms (APC=12, 95% CI 10-15) were less frequently used in Google searches.

The top five nations most cited by global Instagram users in relation to COVID-19 were Italy (963,000 hashtags), Brazil (551,000 hashtags), Spain (376,000 hashtags), Indonesia (298,000 hashtags), and Turkey (244,000 hashtags) (Table 3). Over 2 million hashtags categorized as extremely infodemic, such as “corona,” “corona memes,” “coronado,” and “corona...
time,” have been in circulation. Nevertheless, the most globally used hashtags were those related to health and prevention, such as “stay home/safe” and “lockdown life” (>165 million). The contribution of the “covid-19” hashtag for COVID-19–related information was 35.6% (n=26,200,000), followed by “coronavirus” (n=22,500,000, 30.5%), “corona” (n=18,800,000, 25.6%), and “COVID” (n=5,900,000, 8%) (Table 4).

Table 3. Top 10 Instagram hashtags related to coronavirus disease (COVID-19), as of May 6, 2020.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Country</th>
<th>Monikers/hashtags, n^a</th>
<th>Virus Monikers/hashtags</th>
<th>Health Monikers/hashtags</th>
<th>n^a</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Italy</td>
<td>9.63</td>
<td>covid-19</td>
<td>health-stay home/safe</td>
<td>933</td>
</tr>
<tr>
<td>2</td>
<td>Brazil</td>
<td>5.51</td>
<td>coronavirus</td>
<td>lockdown life</td>
<td>718</td>
</tr>
<tr>
<td>3</td>
<td>Spain</td>
<td>3.76</td>
<td>corona</td>
<td>masks</td>
<td>135</td>
</tr>
<tr>
<td>4</td>
<td>Indonesia</td>
<td>2.98</td>
<td>covid</td>
<td>memes</td>
<td>25</td>
</tr>
<tr>
<td>5</td>
<td>Turkey</td>
<td>2.44</td>
<td>corona memes</td>
<td>gym/fitness</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>India</td>
<td>1.65</td>
<td>coronavirus Italy</td>
<td>art/hobbies</td>
<td>22</td>
</tr>
<tr>
<td>7</td>
<td>Malaysia</td>
<td>0.89</td>
<td>coronado</td>
<td>cooking</td>
<td>21</td>
</tr>
<tr>
<td>8</td>
<td>Dominican Republic</td>
<td>0.83</td>
<td>corona time</td>
<td>fashion</td>
<td>16</td>
</tr>
<tr>
<td>9</td>
<td>United States</td>
<td>0.75</td>
<td>coronavirus memes</td>
<td>hair/beard style</td>
<td>14</td>
</tr>
<tr>
<td>10</td>
<td>Argentina</td>
<td>0.74</td>
<td>coronavirus Brazil</td>
<td>fun/party</td>
<td>13</td>
</tr>
</tbody>
</table>

^aMultiples in 100,000.

Table 4. Top Instagram hashtags related to coronavirus disease (COVID-19) scientific and infodemic names.

<table>
<thead>
<tr>
<th>Hashtag</th>
<th>Count, n (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#2019nCOV</td>
<td>38,993 (0.1)</td>
</tr>
<tr>
<td>#SARSCoV2</td>
<td>49,011 (0.1)</td>
</tr>
<tr>
<td>#SARS</td>
<td>53,633 (0.1)</td>
</tr>
<tr>
<td>#COVID</td>
<td>5,890,625 (8.0)</td>
</tr>
<tr>
<td>#corona</td>
<td>18,849,998 (25.6)</td>
</tr>
<tr>
<td>#coronavirus</td>
<td>22,458,007 (30.5)</td>
</tr>
<tr>
<td>#COVID19</td>
<td>26,213,280 (35.6)</td>
</tr>
</tbody>
</table>

Discussion

Principal Findings

In light of the ongoing COVID-19 pandemic, we are the first to investigate the internet search behaviors of the public and the extent of infodemic monikers circulating on Google and Instagram globally. Our results suggest that (a) “coronavirus,” “corona,” “COVID,” “virus,” “corona virus,” and “COVID-19” are the top five terms used in the Google searches; (b) countries (eg, Italy, Spain, Ireland, Canada, and France) with a high incidence of COVID-19 cases (per million) have greater Google search queries about COVID-19; (c) “coronavirus laboratory,” “coronavirus 5G,” “coronavirus conspiracy,” and “coronavirus bill gates” are widely used infodemic monikers on the internet; (d) although COVID-19 news remains at the top, web searches related to “tips and cures” for COVID-19 spiked when the US president speculated about a “miracle cure” and the use of a disinfectant injection to treat COVID-19; (e) 66.1% (n=48,700,000) of Instagram users used “COVID-19” and “coronavirus” as a hashtag to disperse information related to COVID-19.

Exploring research using nontraditional data sources such as social media has several implications. First, our results demonstrated a potential application for the use of Instagram as a complementary tool to aid in understanding online search behavior; we also provided real-time tracking of infodemic monikers circulating on the internet. The strength of this study is the investigation of various infodemic monikers dispersed on the internet and correlating them with the events associated with that particular day. Although we used correlations to examine the possible linear association between search queries and the event, it should be noted that use of a search engine is voluntary and self-initiated search queries represent the users who are truly curious or worried about a situation. Thus, we believe that the unobtrusive search behavior of netizens may have resulted in an increase in search volume. By characterizing and classifying various infodemic monikers based on the degree of infodemicity (ie, via the I-scale), researchers can foster new methods of using social media data to monitor the monikers’
outcomes. The analysis and methods used in this study could aid public health and communication agencies in identifying and diminishing infodemic monikers circulating on the internet.

Findings from this study validate and extend previously published works that used Google keywords [1,12,13]. We also demonstrate the potential for the use of Instagram hashtags to monitor and predict both the cyber behavior and relaying of misinformation on the internet [22-24]. In 2017, Guidry et al [22] studied Ebola-related risk perception in Instagram users and identified that a significant proportion of posts had rampant misinformation about the Ebola disease during the outbreak. In addition, the percentage of Instagram posts and tweets posted by health organizations (eg. Centers for Disease Control and Prevention, World Health Organization, Médecins Sans Frontières [Doctors Without Borders]) that correct misinformation is less than 5% [22]. In general, negative information posted on the internet tends to receive a greater weight among netizens. Thus, this should be counter-balanced with evidence-based content from health organizations, particularly in the current pandemic situation. For example, when the US president suggested injecting disinfectant to treat COVID-19, the number of Google searches considering it as a cure sharply increased (APC=53) and resulted in 30 cases of disinfectant poisoning within 18 hours in New York City [25]. Health authorities should be vigilant and provide more positive and informative messages to combat the circulation of infodemic monikers on social media. Future studies will need to investigate the influence of infodemic monikers on individual cyber behavior.

**Limitations**

Our study used Google Trends, which only provides the search behavior of people using the Google search engine. Furthermore, our study focused on Google and Instagram for data retrieval. Future studies should consider studying the same topic on other social media platforms to capture a more diverse population of users. Instagram searches were conducted manually, introducing a potential for human error. Going forward, the use of an automated program can improve the accuracy of the data collected and analyzed. Lastly, Google Trends did not provide any information about the methodology used to generate search data and algorithms.

**Conclusion**

Using Google Trends and Instagram hashtags, the present study identified that there is a growing interest in COVID-19 globally and in countries with a higher incidence of the virus. Searches related to “COVID-19 news” are quite frequent and two thirds of Instagram users have used “COVID-19” and “coronavirus” as hashtags to disperse information related to the virus. Several infodemic monikers are circulating on the internet, with “coronavirus conspiracy” and “coronavirus laboratory” identified as the most dangerous (1-scale score=9). Given the prevalence of infodemic monikers, mass media regulators and health organizers should monitor and diminish the impact of these monikers. These governing bodies should also be encouraged to take serious actions against those spreading misinformation in social media.

**Conflicts of Interest**

None declared.

**Multimedia Appendix 1**

Terms used to identify the most frequently searched queries in Google and the hashtag suggestions for Instagram.

[DOCX File , 14 KB-Multimedia Appendix 1]

**References**


Abbreviations

APC: average peak volume
COVID-19: coronavirus disease
I-scale: infodemic scale
RSV: relative search volume

Edited by G Eysenbach, G Fagherazzi; submitted 25.05.20; peer-reviewed by W Aldhaleei, A Amresh, A Zink; comments to author 13.07.20; revised version received 02.08.20; accepted 03.08.20; published 25.08.20

Please cite as:
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
URL: http://www.jmir.org/2020/8/e20673/
doi: 10.2196/20673
PMID: