1 **Original Research**

2

3 **Global Infodemiology of COVID-19: Focus on Google Web Searches and**
4 **Instagram Hashtags**

5 *Alessandro Rovetta[1], Akshaya Srikanth Bhagavathula[2]*

6 [1]Mensana srls, Research and Disclosure Division, Via Moro Aldo 5 - 25124 Brescia, Italy (ORCID:
7 0000-0002-4634-279X)

8 [2]PhD Candidate, Institute of Public Health, College of Medicine and Health Sciences, United Arab
9 Emirates University, Al Ain, UAE. (ORCID: 0000-0002-0581-7808)

10

11

12 **Corresponding author**

13 Alessandro Rovetta,

14 Mensana srls, Research and Disclosure Division,

15 Via Moro Aldo 5 – 25124 Brescia, Italy.

16 Email: rovetta.mresearch@gmail.com

17 Phone: +39-3927112808

18 ORCID: https://orcid.org/0000-0002-4634-279X

19

20

21

22

23

24

25

26

27

28

29

30   **Abstract**

31   **Background:** Though 'infodemiological' methods have been used in COVID-19 research, an
32   examination of the extent of infodemic monikers (misinformation) use on the Internet remains
33   limited.

34   **Objective:** To investigate Internet search behavior related to COVID-19 and examine the
35   circulation of infodemic monikers through two platforms—Google and Instagram—during the
36   current global pandemic.

37   **Methods:** Using Google Trends and Instagram hashtags (#), we explored Internet search activities
38   and behaviors related to the COVID-19 pandemic from February 20, 2020, to May 06, 2020. We
39   investigated the names used to identify the virus, health and risk perception, life during the
40   lockdown, and information related to the adoption of COVID-19 infodemic monikers. We
41   computed the average peak volume (APC) with a 95% confidence interval (CI) during the study
42   period for the monikers.

43   **Results:** The top five COVID-19-related terms in the Google searches were "coronavirus",
44   "corona", "COVID", "virus", "corona virus", and "COVID-19". Countries with a higher number of
45   COVID-19 cases had a higher number of COVID-19 queries on Google.   The monikers
46   "coronavirus ozone", "coronavirus laboratory", "coronavirus 5G", "coronavirus conspiracy" and
47   "coronavirus bill gates" were widely circulated on the Internet. Searches about 'tips and cures' for
48   COVID-19 spiked in relation to the U.S. president speculating about a 'miracle cure' and
49   suggesting the injection of disinfectant to treat the virus. Around two-thirds (66.1%) of Instagram
50   users used the hashtags "COVID-19", and "coronavirus" to disperse virus-related information.

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

55   *Keywords: COVID-19, coronavirus, Google, Instagram, Infodemiology, social media.*

56

57

58

59

60

61

62

63

64

65

66

67

68

Introduction

70 Globally, the Internet is an extremely important platform for obtaining knowledge and information

71 about the novel coronavirus (COVID-19) pandemic [1-3]. The Google Trends tool provides real-

72 time insights into internet search behavior on various topics, including COVID-19 [4]. Social media

73 platforms such as Facebook, Twitter, and Instagram allow users to communicate their thoughts,

74 feelings, and opinions by sharing short messages. A unique aspect of social media data from

75 Instagram is that image-based posts are accessible, and the use of topic-related hashtags (#) allows

76 access to topic-related information for all Internet users [5]. In general, there is a growing interest in

77 examining social data to understand and monitor public behavior in real-time [6,7].

78 Research on the Internet and social data are called *Infodemiology* or *Infoveillance* studies

79 [8]. *Infodemiology* is defined as "the science of distribution and determinants of information in an

80 electronic medium, specifically the Internet, or in a population, with the ultimate aim to inform

81 public health and public policy" [9]. Although several studies have been conducted using

82 'infodemiological' methods in COVID-19 research, a limited number of studies have examined the

83 extent of COVID-19-related misinformation on the Internet [10-14]. The fake news, misleading,

84 and misinformation circulating on the Internet are referred to as "infodemic monikers". These

85 monikers can profoundly affect public health communication and also contribute to xenophobia

86 [12-17]. "*Infodemic monikers*" are defined as substantially erroneous information, which give rise

87 to interpretation mistakes, fake news, episodes of racism, or any other forms of misleading

88 information circulating on the internet [14]. In this context, we aimed to investigate the Internet

89 search behavior related to COVID-19 and the extent of infodemic monikers circulating on Google

90 and Instagram during the current pandemic period in the world.

91

92

93

Methods

95 We used Google Trends and Instagram hashtags to explore internet search activities and behaviors

96 related to the COVID-19 pandemic from February 20, 2020, to May 06, 2020. We investigated the

97 following: names used to identify the virus, health and risk perception, life during the lockdown,

98 and information related to the adoption of infodemic monikers related to COVID-19. The complete

99 list terms used to identify the most frequently searched queries in Google and the hashtag

100 suggestions for Instagram are presented in Supplementary File 1.

101 The obtained infodemic monikers are characterized as follows:

102   1. *Generic:* The moniker confuses, due to lack of specificity.

103   2. *Misinformative*: The moniker associates a certain phenomenon with fake news.

104   3. *Discriminatory:* The moniker encourages the association of a problem with a specific

105   ethnicity and/or geographical region.

106   4. *Deviant:* The moniker does not identify the requested phenomenon.

107   5. *Other specificities:* We keep two additional points for special cases that prove

108   exceptionally serious.

109 To determine the severity of various infodemic monikers circulating on the Internet, each

110 moniker was assigned 1 to 2 points on the infodemic scale (I-scale) ranging from 0 (minimum)

111 to 10 (maximum). Based on the sum of the I-scale scores, the infodemic monikers were

112 classified as follows:

113   • Not infodemic: 0

114   • Lowly infodemic: 1

115   • Moderately infodemic: >1- 4

116   • Highly infodemic: 5-8

117   •   Extremely infodemic: 9-10

118 For each search keyword considered, Google Trends provided normalized data in the form of

119 relative search volume (RSV) based on search popularity ranging from 0 (low) to 100 (highly

120 popular). Using these RSV values, we computed the average peak volume (APC) with a 95%

121 confidence interval (CI) during the study period.

122   Instagram, a platform for image-based posts with hashtags (#) was screened. We retrieved

123 content based on hashtags through image classifiers, every 3-4 days during the study period. All

124 irrelevant content was excluded. The data collected included contents posted on Instagram and self-

125 reported user demographic information. No personal information, such as emails, phone numbers,

126 or addresses, was collected. The data from the Instagram hashtags were collected manually, through

127 the Instagram-suggested tags associated with specific countries.

128   All data used in the study were obtained from anonymous open sources. Thus, ethical

129 approval was not required.

130 **Results**

131 The top five COVID-19 related infodemic and scientific terms used in Google searches were

132 "coronavirus", "corona", "COVID", "virus", "corona virus", and "COVID-19" [Figure 1]. The most

133 frequently used keywords globally were "coronavirus" (APC: 1378, 95% CI: 1246-1537), followed

134 by "corona" (APC: 530, 95% CI: 477-610) and "COVID" (APC: 345, 95% CI: 292-398). Several

135 keywords related to COVID-19 (Table 1) were identified, of the top 10, five had an I-scale value of

136 8: "corona", "corona Italy", "corona Deutschland", "corona China" and "corona Wuhan".

137 The country-wise dispersion of the scientific and infodemic names of COVID-19 used in Google

138 searches are shown in Figure 2. Countries with a higher number of COVID-19 cases per 1 million

139 population have recorded greater Google search queries related to COVID-19 (Italy, Spain, Ireland,

140   Canada, France, and Qatar). These COVID-19-related search queries were significantly correlated

141   with the incidence of COVID-19 cases across the countries (Pearson R = 0.45, $p<0.05$).

142   The top COVID-19-related infodemic monikers such as "coronavirus ozone", "coronavirus

143   laboratory", and "coronavirus 5G" frequently circulated on the Internet are presented in Table 2.

144   The following are infodemic monikers with the highest I-scores globally, "coronavirus conspiracy"

145   (I-score: 10), "coronavirus laboratory" (I-score: 9), and "coronavirus 5G" (I-score: 9). Additionally,

146   the use of monikers with moderate to high infodemicity far exceeded the use of scientific names

147   (Table 2): 57% of Google web searches are moderately infodemic (total APC: 109, 95% CI: 89 –

148   139) and 16% highly infodemic (total APC: 30, 95% CI: 25 – 34). The circulation of these

149   infodemic monikers was further examined to understand the events associated with these searches.

150   Infodemic monikers related to coronavirus origins, such as SARS-CoV-2 made in the laboratory"

151   went viral (APC: 41) when the National Association Press Agency (NAPA) from Italy posted a

152   2015 video about the origins of SARS-CoV-2 virus on March 25, 2020[18]. In addition, the

153   moniker reached the breakout level (RSV: 100) on April 17, 2020, when the French Noble Prize

154   winner Prof. Luc Montagnier stated that the new coronavirus is the result of a laboratory accident in

155   the Wuhan high-security laboratory in China [19]. Detailed information on the different infodemic

156   monikers and the associated events are shown in Figure 3.

157   The top searches related to health, precautions, and COVID-19 news are presented in Figure

158   4. Google searches related to COVID-19 news remained at the top throughout the pandemic period.

159   However, searches related to 'tips and cures' for COVID-19 spiked multiple times when the U. S.

160   president suggested that hydroxychloroquine (an unproven drug) was a 'miracle cure' on April 4,

161   2020 (RSV: 70) [20] and also injecting disinfectant to treat COVID-19 on April 24, 2020 (RSV:

162   53) [21]. Other searches related to the use of medical masks and disinfectants (APC: 23, 95% CI: 21

163   – 25), lockdown (APC: 19, 95% CI: 16 – 22), and COVID-19 symptoms (APC: 12, 95% CI: 10 –

164   15), are less frequently used in Google searches.

165   The top 10 COVID-19-related hashtags used on Instagram (country-specific) and groups and

166   topics associated with these hashtags are summarized in Table 3. Around one million users from

167   Italy used 'covid-19' as a hashtag 3.6 million times to communicate information related to health,

168   stay-home/safe (93.3 million times). These hashtags remained at the top for use for COVID-19-

169   related communication on Instagram. Similarly, Instagram users from Brazil (551,000), Spain

170   (376,000), Indonesia (298,000), and other countries were mostly used Instagram frequently to

171   distribute COVID-19 related information. Moreover, the contribution of the 'covid-19' hashtag for

172   COVID-19 related information was 35.6%, followed by 'coronavirus' (30.5%), 'corona' (25.6%),

173   and 'COVID' (8%) [Figure 5].

174   **Discussion**

175   In light of the ongoing COVID-19 pandemic, we are the first to investigate the Internet search

176   behavior of the public and the extent of Infodemic monikers circulated on Google and Instagram

177   globally. Our results suggest that (i). "coronavirus", "corona", "COVID", "virus", "corona virus",

178   and "COVID-19" are the top five terms used in the Google searches. (ii). countries (e.g., Italy,

179   Spain, Ireland, Canada, and France) with a high incidence of COVID-19 cases (per million) have

180   recorded greater Google search queries about COVID-19. (iii). "coronavirus ozone", "coronavirus

181   laboratory", "coronavirus 5G", "coronavirus conspiracy" and "coronavirus bill gates" are widely

182   used infodemic monikers on the Internet. However, the "coronavirus conspiracy" was the only

183   moniker that achieved the highest I-score of 10. (iv). though COVID-19 news remains at the top,

184   web searches related to 'tips and cures' for COVID-19 spiked when the U.S. president speculated

185   about a 'miracle cure' and the injection of a disinfectant to treat COVID-19. (v). Around two-thirds

186   (66.1%) of Instagram users use "COVID-19", and "coronavirus" as a hashtag to disperse the

187   information related to COVID-19.

188   Exploring research using nontraditional data sources such as social media has several

189   implications. First, our results demonstrated a potential application for using Instagram as a

190 complementary tool to aid in understanding online search behavior and also provided real-time

191 tracking of infodemic monikers circulated on the Internet. The strength of this study is the

192 investigation of various infodemic monikers dispersed on the internet and correlating them with the

193 events associated with that particular day. By characterizing and classifying various infodemic

194 monikers based on the degree of infodemicity scores (I-score), researchers can foster new methods

195 of using social media data to monitor the monikers' outcomes. The analysis and methods used in

196 this study could leverage public health and communication agencies in identifying and diminishing

197 infodemic monikers circulating on the Internet.

198 Findings from this study validate and extend previously published works that used Google

199 keywords [1,12,13]. We also demonstrate the potential for the use of Instagram hashtags to monitor

200 and predict both the cyber behavior and relaying of misinformation on the Internet [22-24]. In 2017,

201 *Guidry et al.* studied Ebola-related risk perception in Instagram users and identified that a

202 significant proportion of posts on Instagram had rampant misinformation about the Ebola disease

203 during the outbreak [22]. In addition, the percentage of Instagram posts and tweets posted by health

204 organizations (CDC, WHO, MSF) to correct misinformation are less than 5% [22]. In general,

205 negative information posted on the Internet tends to receive a greater weight among netizens. Thus,

206 this should be counter-balanced with evidence-based solution content from health organizations,

207 particularly in the current pandemic situation. For example, when the US president suggested

208 injecting disinfectant to treat COVID-19, the number of Google searches considering it as a cure

209 sharply increased (APC:53) and also implicated 30 cases of disinfectant poisoning within 18 hours

210 in New York City [25]. Health authorities should be vigilant and provide more positive and

211 informative messages to combat the circulation of infodemic monikers on social media. Future

212 studies will need to investigate the influence of infodemic monikers on individual cyber behavior.

213 **Limitations**

214 Our study used Google Trends, which provides the search behavior of people using their search

215 engines, but not others. We mainly focused on Google and Instagram for data retrieval. Future

216 studies should consider studying the same topic on other social media platforms to capture a more

217 diverse population of users. Instagram searches were conducted manually, introducing a variable of

218 error. Going forward, the use of an automated program can improve the accuracy of the data

219 collected and analyzed. Lastly, Google Trends did not provide any information about the methods

220 used to generate search data and algorithms.

221 **Conclusion**

222 Using Google Trends and Instagram hashtags, the present study identified that there is a growing

223 interest in COVID-19 globally and in countries with a higher incidence of the virus. Searches

224 related to 'COVID-19 news' are quite frequent and two-thirds (66.1%) of Instagram users have

225 used "COVID-19", and "coronavirus" as hashtags to disperse information related to the virus.

226 Several infodemic monikers are circulating on the Internet, with "coronavirus conspiracy" identified

227 as the most popular moniker (I-score of 10). Given the prevalence of the infodemic moniker use,

228 mass media regulators and health organizers should monitor and diminish the impact of these

229 monikers. These governing bodies should also be encouraged to take serious actions against those

230 spreading misinformation in social media.

231 **Acknowledgment:** None

232 **Conflict of Interest:** None

233 **Source of funding:** None

234 **Data availability:** All the data related to this study are presented in the Supplementary file.

235

236 **References**

237 1. Bento AI, Nguyen T, Wing C, Lozano-Rojas F, Ahn YY, Simon K. Evidence from internet
238 search data shows information-seeking responses to news of local COVID-19 cases. Proc Natl Acad
239 Sci USA; 2020. PMID: 32366658

240 2. Effenberger M, Kronbichler A, Shin JI, Mayer G, Tilg H, Perco P. Association of the COVID-19
241 pandemic with Internet Search Volumes: A Google TrendsTM Analysis. Int J Infect Dis; 2020
242 (95):192-197. PMID: 32305520

243 3. Lin YH, Liu CH, Chiu YC. Google searches for the keywords of "wash hands" predict the speed
244 of national spread of COVID-19 outbreak among 21 countries. Brain Behav Immun; 2020. PMID:
245 32283286

246 4. Google COVID-19. Available at: https://www.google.com/covid19/html. (Accessed: May
247 20,2020).

248 5. Giannoulakis S, Tsapatsoulis N. Evaluating the descriptive power of Instagram hashtags. J Innov
249 Digit Ecosyst; 2016 (3):114-129. DOI: 10.1016/j.jides.2016.10.001

250 6. Salathé, M. Digital epidemiology: what is it, and where is it going?. Life Sci Soc Policy; 2018 (1)
251 4:1. PMID: 29302758

252 7.Global social media research summary 2020. Available at: https://www.smartinsights.com/social-
253 media-marketing/social-media-strategy/new-global-social-media-research/ (Accessed: May 20,
254 2020).

255 8. Eysenbach G. Infodemiology and infoveillance tracking online health information and
256 cyberbehavior for public health. Am J Prev Med; 2011 40(5 Suppl 2):S154-S158. PMID: 21521589

257 9. Eysenbach G. Infodemiology: The epidemiology of (mis) information. Am J Prev Med; 2002
258 (113): 763-765. PMID: 12517369

259 10. Hernández-García I, Giménez-Júlvez T. Assessment of health information about COVID-19
260 prevention on the internet: infodemiological study. JMIR Public Health Surveill; 2020 (6): e18717.
261 PMID: 32217507

262 11. Park HW, Park S, Chong M. Conversations and Medical News Frames on Twitter:
263 Infodemiological Study on COVID-19 in South Korea. J Med Internet Res; 2020 (22): e18897.
264 PMID: 32325426

265 12. Cuan-Baltazar JY, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E.
266 Misinformation of COVID-19 on the Internet: Infodemiology study. JMIR Public Health and
267 Surveill; 2020 (6): e18444. PMID: 32250960

268 13. Rovetta A, Bhagavathula AS. COVID-19-Related Web Search Behaviors and Infodemic
269 Attitudes in Italy: Infodemiological Study. JMIR Public Health Surveill; 2020; (6): e19374. PMID:
270 32338613

271 14. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of Tweeters during
272 the COVID-19 pandemic: infoveillance study. J Med Internet Res; 2020 (22): e19016. PMID:
273 32287039

274 15. Shimizu K. 2019-nCoV, fake news, and racism. Lancet; 2020 (395):685-686. PMID: 32059801

275 16. Chung RY, Li MM. Anti-Chinese sentiment during the 2019-nCoV outbreak. Lancet; 2020
276 (395): 686-687. PMID: 32122469

277 17. Time. 2020 Feb 29. As Coronavirus Spreads, So Does Xenophobia and Anti-Asian Racism
278 URL: https://time.com/5797836/coronavirus-racism-stereotypes-attacks/ (Accessed April 9, 2020).

279 18. Coronavirus: Il caso del video del Tgr Leonardo 2015 sul supervirus creato in Cina. (Article in
280 Italian) Available at: https://www.ansa.it/sito/notizie/politica/2020/03/25/coronavirus-il-caso-del-

281   video-del-tgr-leonardo-2015-sul-supervirus-creato-in-cina_7adf8316-6ca5-42cd-96de-
282   c18f7fb53595.html (Accessed 30 April 2020).

283   19. COVID-19: la théorie d'un virus fabriqué vivement contestée (Article in French). Available at:
284   https://www.lapresse.ca/actualites/sciences/202004/17/01-5269764-covid-19-la-theorie-dun-virus-
285   fabrique-vivement-contestee.php (Accessed 30 April 2020)

286   20. Trump Urges Coronavirus Patients to Take Unproven Drug. Available at:
287   https://www.nytimes.com/2020/04/04/health/coronavirus-drug-trump-hydroxycholoroquine.html
288   (Accessed 30 April 2020)

289   21. Trump suggests 'injection' of disinfectant to beat coronavirus and 'clean' the lungs.
290   https://www.nbcnews.com/politics/donald-trump/trump-suggests-injection-disinfectant-beat-
291   coronavirus-clean-lungs-n1191216 (Accessed 30 April 2020)

292   22. Guidry JP, Jin Y, Orr CA, Messner M, Meganck S. Ebola on Instagram and Twitter: How health
293   organizations address the health crisis in their social media engagement. Public Relat Rev; 2017
294   (43): 477-486.

295   23. Zarei K, Farahbakhsh R, Crespi N, Tyson G. A first Instagram dataset on COVID-19. arXiv
296   preprint arXiv:2004.12226. 2020.

297   24. Gupta R, Ariefdjohan M. Mental illness on Instagram: a mixed method study to characterize
298   public content, sentiments, and trends of antidepressant use. J Ment Health 2020. PMID: 32325006

299   25. Calls to poison centers spike after the presidents comments about using disinfectants to treat
300   coronavirus. Available at: https://www.forbes.com/sites/robertglatter/2020/04/25/calls-to-poison-
301   centers-spike--after-the-presidents-comments-about-using-disinfectants-to-treat-coronavirus.html
302   (Accessed 15 May 2020).

303

Table 1: Top infodemic and scientific Google searches related to COVID-19 in the world

Keyword	APC	95% CI	I-scale value
coronavirus	1378	1246 – 1537	3
corona	530	477 – 610	8
COVID	345	292 – 398	1
virus	239	212 – 292	7
corona virus	159	133 – 186	6
coronavirus Italy	54	45 – 62	4
COVID-19	53	45 – 60	0
coronavirus USA	32	29 – 36	4
coronavirus China	30	25 – 34	6
coronavirus Germany	23	20 – 27	4
corona Italy	13	12 – 14	8
corona Deutschland	12	10 – 14	8
SARS	9	8 – 10	6
corona China	9	7 – 11	8
corona Wuhan	1	0 – 2	8
SARS-CoV-2	1	0 – 1	0

Queries in APC: average peak volume; CI: confidence interval; I-scale: infodemic scale ranging from 0-10

Table 2: Top global infodemic Google searches related to COVID-19

Keyword	APC	95% CI	I-scale value
coronavirus ozone	19	15 – 22	6
coronavirus laboratory	16	12 – 19	9
coronavirus 5G	10	8 – 13	9
coronavirus conspiracy	9	8 – 11	10
coronavirus bill gates	8	7 – 10	6
coronavirus milk	7	6 – 8	6
coronavirus military	4	4 – 5	8
coronavirus uv	3	3 – 4	8

Queries in APC: average peak volume, and CI: confidence interval; I-scale: infodemic scale ranging from 0-10

Table 3: Top 10 Instagram hashtags related to COVID-19

Top 10 COVID-19-related Instagram hashtags (#)						
Rank	*Country group*	*Quantity†*	*Hashtag group*	*Quantity†*	*Hashtag topic group*	*Quantity†*
1	Italy	9.63	covid-19	306	Health-stay home/safe	933
2	Brazil	5.51	coronavirus	267	lockdown life	718
3	Spain	3.76	corona	188	masks	135
4	Indonesia	2.98	covid	69	memes	25
5	Turkey	2.44	corona memes	14	gym/fitness	24
6	India	1.65	coronavirus Italy	9.63	art/hobbies	22
7	Malaysia	0.89	coronado	8.19	cooking	21
8	Dominican Republic	0.83	corona time	7.12	fashion	16
9	USA	0.75	coronavirus memes	6.41	hair/beard style	14
10	Argentine	0.74	coronavirus Brazil	5.51	fun/party	13

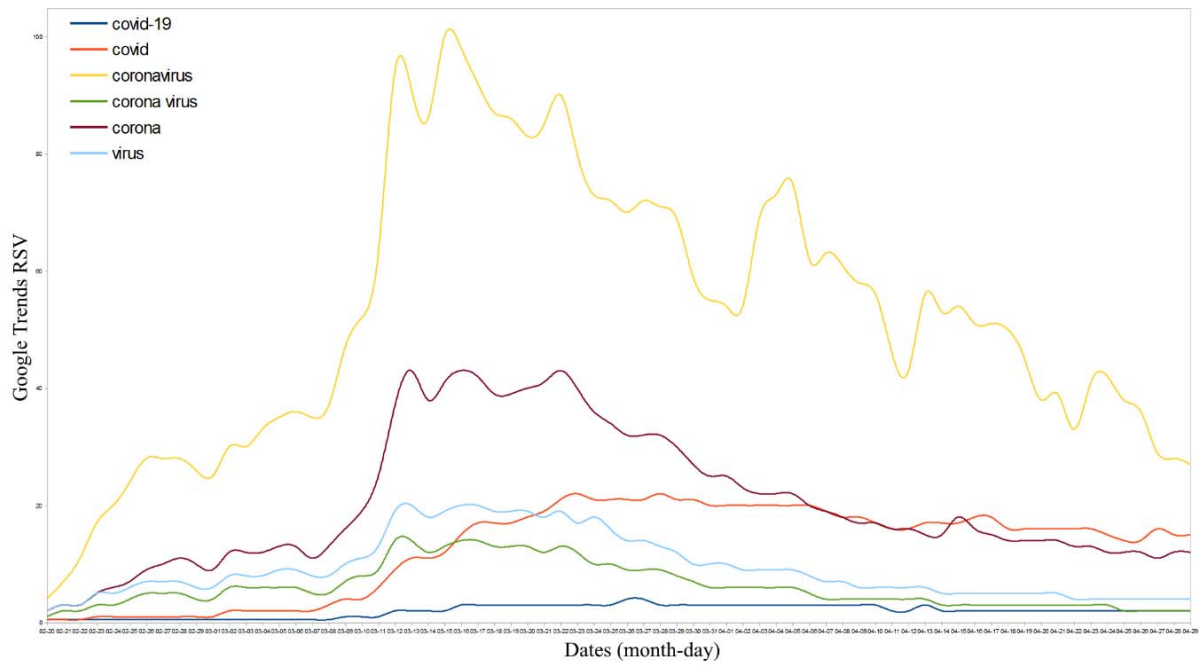Searches identified until: May 6, 2020; †multiples in 100,000

Figure 1: Top global scientific and infodemic names related to COVID-19 in the Google
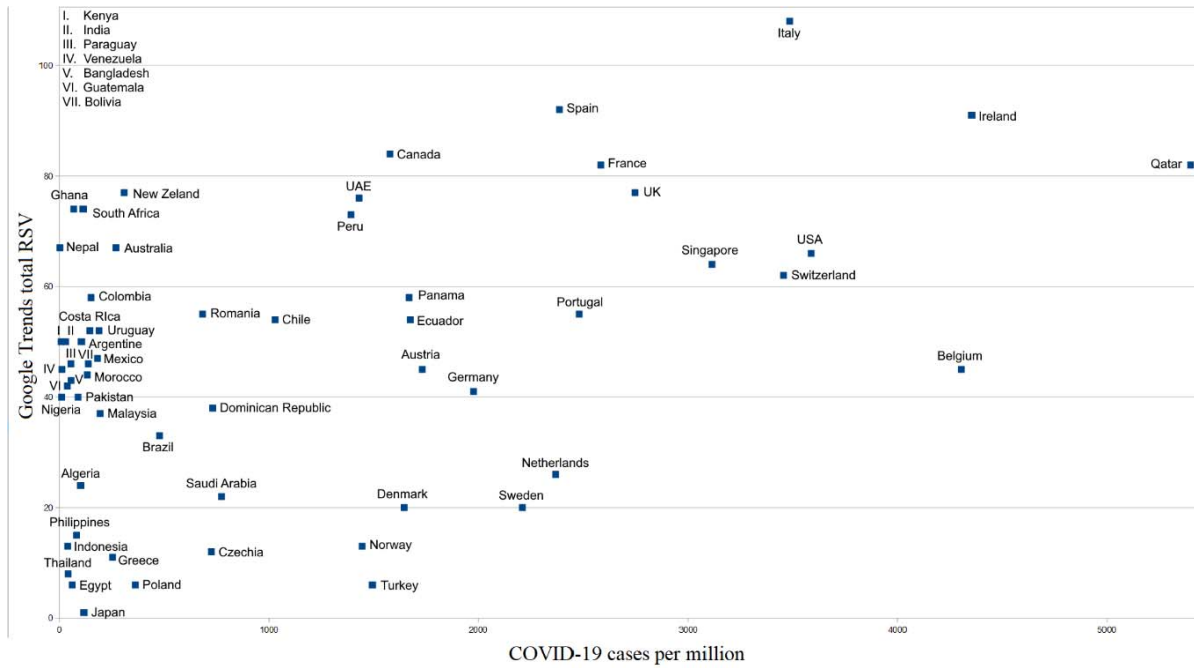
Figure 2: Countries-wise dispersion of scientific and infodemic names of COVID-19
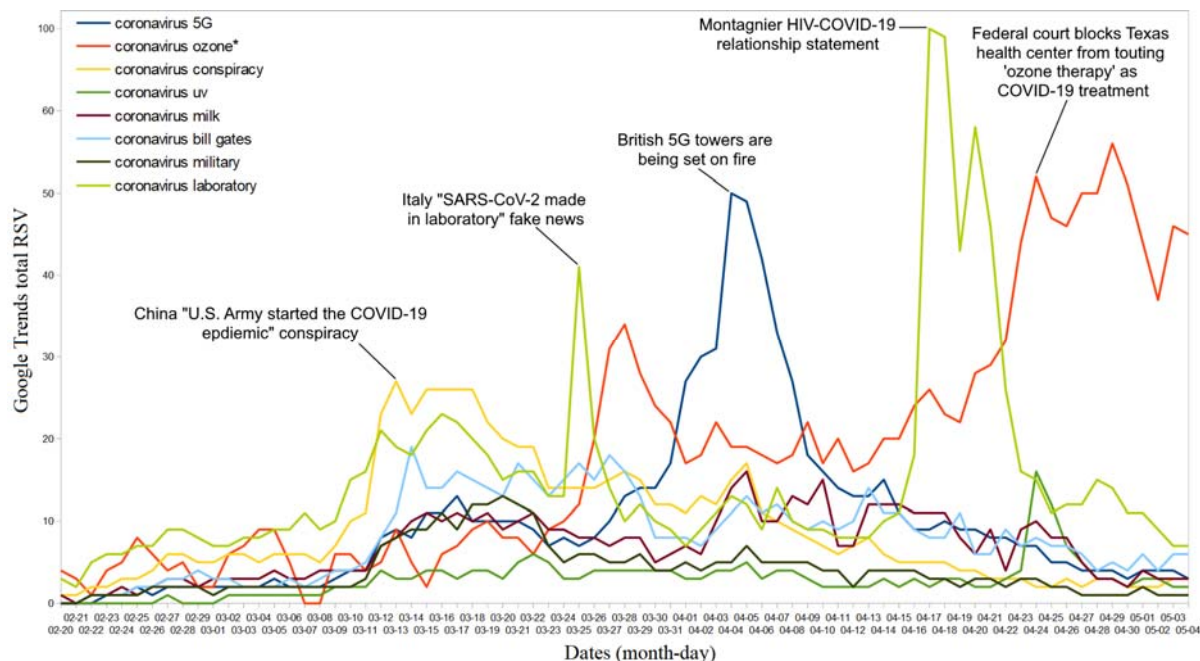
Figure 3: Top high and extreme infodemic global web searches related to COVID-19.

* the ozone-coronavirus association concerns both the alleged therapy against COVID-19 and the stratospheric phenomenon. Although the second association is not directly infodemic, it can contribute to the spread of the first.
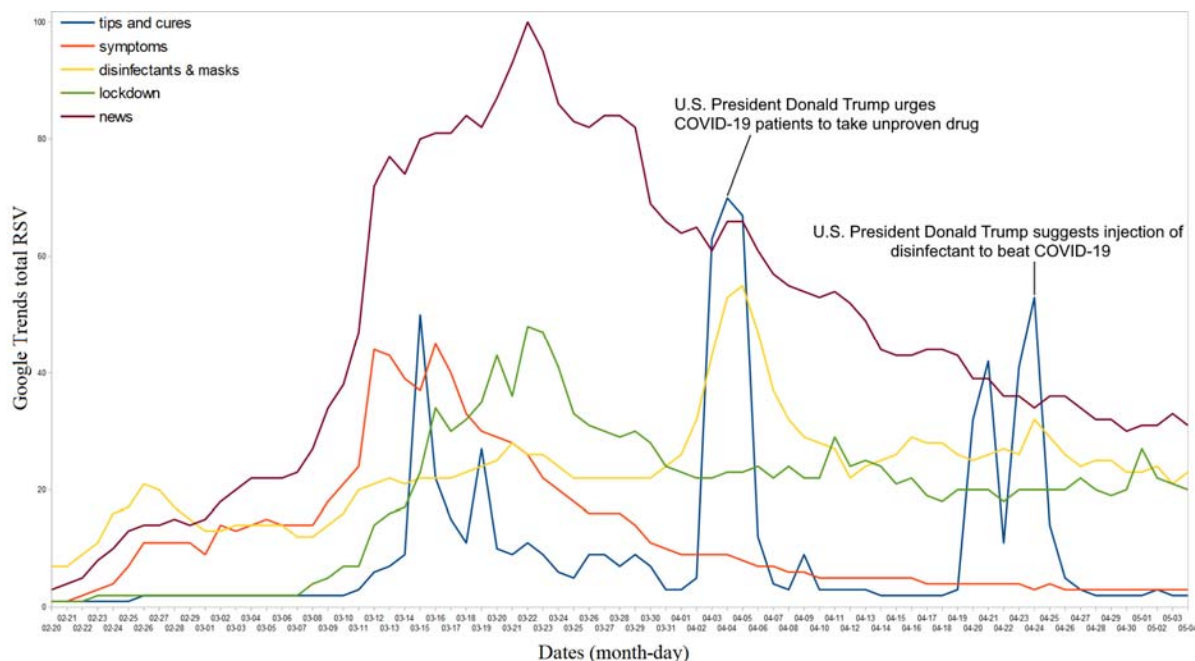

Figure 4: Top global web searches related to health, precautions and COVID-19 news.

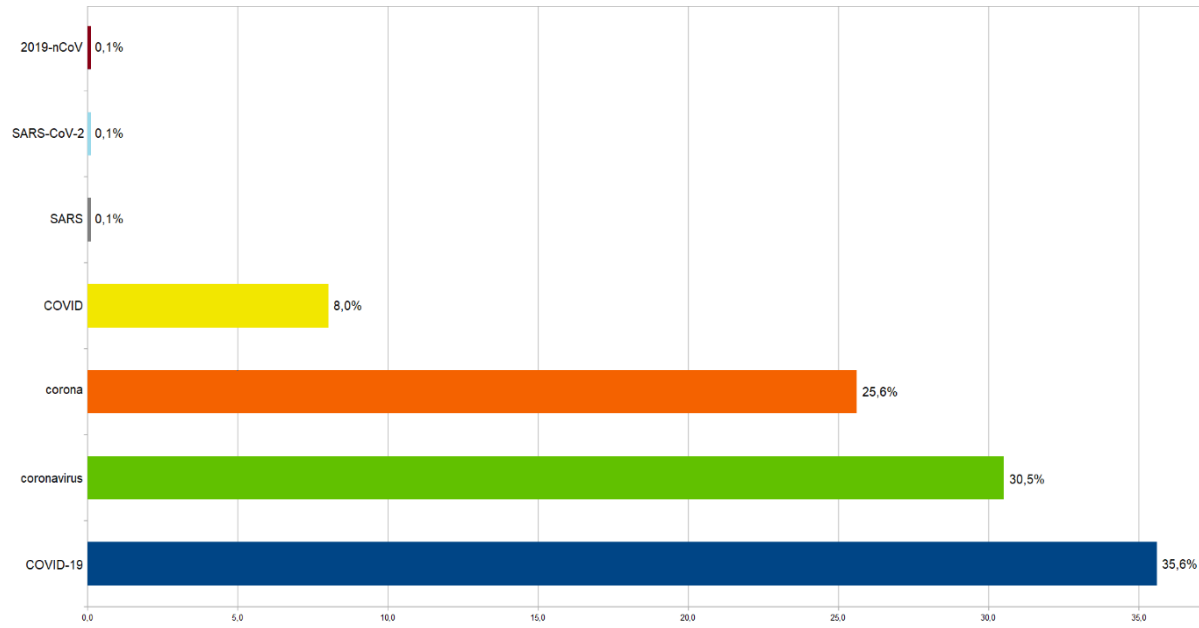**Figure 5:** Top Instagram hashtags related to COVID-19 scientific and infodemic names.