Original investigations / Commentaries

Google search volume predicts the emergence of COVID-19 outbreaks

Giuseppe Lippi[1], Camilla Mattiuzzi[2], Gianfranco Cervellin[3]
[1] Section of Clinical Biochemistry, University of Verona, Verona, Italy; [2] Service of Clinical Governance, Provincial Agency for Social and Sanitary Services, Trento, Italy; [3] Academy of Emergency Medicine and Care, Pavia, Italy

Summary. *Background and aim*: Digital epidemiology is increasingly used for supporting traditional epidemiology. This study was hence aimed to explore whether the Google search volume may have been useful to predict the trajectory of coronavirus disease 2019 (COVID-19) outbreak in Italy. *Materials and Methods:* We accessed Google Trends for collecting data on weekly Google searches for the keywords "tosse" (i.e., cough), "febbre" (i.e., fever) and "dispnea" (dyspnea) in Italy, between February and May 2020. The number of new weekly cases of COVID-19 in Italy was also obtained from the website of the National Institute of Health. *Results:* The peaks of Google searches for the three terms predicted by 3 weeks that of newly diagnosed COVID-19 cases. The peaks of weekly Google searches for "febbre" (fever), "tosse"(cough) and "dispnea" (dyspnea) were 1.7-, 2.2- and 7.7-fold higher compared to the week before the diagnosis of the first national case. No significant correlation was found between the number of newly diagnosed COVID-19 cases and Google search volumes of "tosse" (cough) and "febbre" (fever), whilst "dyspnea" (dyspnea) was significantly correlated (r= 0.50; p=0.034). The correlation between newly diagnosed COVID-19 cases and "tosse" (cough; r=0.65; p=0.008) or "febbre" (fever; 0.69; p=0.004) become statistically significant with a 3-week delay. All symptoms were also significantly inter-correlated. *Conclusions:* Continuously monitoring the volume of Google searches and mapping their origin can be a potentially valuable instrument to help predicting and identifying local recrudescence of COVID-19. (www.actabiomedica.it)

Key words: coronavirus, coronavirus disease 2019, COVID-19, epidemiology, infodemiology, Google Trends

Introduction

Coronavirus disease 2019 (COVID-19) is a critical infectious illness, sustained by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). According to official sources, this infectious disorder has probably emerged in the city of Wuhan (China) at the end of the 2019, and has then been conveyed all around the world, causing several hundred thousand deaths, and persuading the World Health Organization (WHO) to declare COVID-19 a pandemic disease, exactly 100 years after the notorious "Spanish flu" outbreak (1,2).

The still ongoing worldwide spread of the virus has caused an unprecedented healthcare (3), economic (4), and societal (5) burden on mankind. Besides the high infectivity, virulence and pathogenicity of SARS-CoV-2, one lesson that we have now clearly learnt is that the diagnostic challenges, along with the inefficient contact tracing, delayed cases isolation and retarded clinical management, have been major drawbacks in the management of this pandemic, which have thus fostered a rapid and massive spread of the virus all around the world, especially during the early phase after its emergence in China (6). Predicting, and hence anticipating the contagion curve, through implementation of efficient diagnostic tools, practicable contention measures and early patient care, shall now be considered cornerstones for preventing future epidemic waves and for efficient containment of local outbreaks (7).

Several lines of evidence have now become available that digital epidemiology could be regarded as a valuable support for traditional epidemiology, whereby it can contribute to identify emerging human pathologies, thus including infectious diseases, even earlier than using traditional epidemiological tools (8). Google Trends is one of such innovative and increasingly used informatics intruments. This freely available web-based application enables to collect direct information on the volume of Google searches for specific terms ("keywords"), using different languages, from different locations, and even across many different periods of time (9).

This study was hence aimed to explore whether or not the volume of Google searches for the most frequent symptoms of SARS-CoV-2 infection may have been useful to predict the trajectory of COVID-19 outbreak in Italy, and if the future usage of digital epidemiology could contribute to more efficient management of possible recrudescence of this diffuse and devastating pathology.

Materials and Methods

An electronic search was carried out in Google Trends (Google Inc. Mountain View, CA, US), using the three Italian keywords "tosse" (i.e., cough), "febbre" (i.e., "fever") and "dispnea" (dyspnea). This search terms were selected as they represent the most frequently symptoms reported at disease onset by patients with COVID-19 (73% for both cough and fever, 63% for dyspnea, respectively) (10). The search period ranged between the first week of February 2020 and the end of May 2020, and was limited to Google searches carried out within the Italian national territory. The weekly Google Trends score, which reflects the mean volume of weekly Google searches for these terms, was then downloaded and imported into a Microsoft Excel file (Microsoft, Redmond, WA, United States) for statistical analysis. The number of weekly new cases of COVID-19 in Italy was retrieved from the official website of the National Institute of Health (Istituto Superiore di Sanità), and tabulated into the same Excel worksheet. The volume of Google searches for the three keywords and for the new cases of COVID-19 were analyzed with Spearman's correlation, along with its 95% confidence interval (95%CI). The statistical analysis was carried out using Analyse-it (Analyse-it Software Ltd, Leeds, UK). The study was carried out in accordance with the Declaration of Helsinki, under the terms of relevant local legislation.
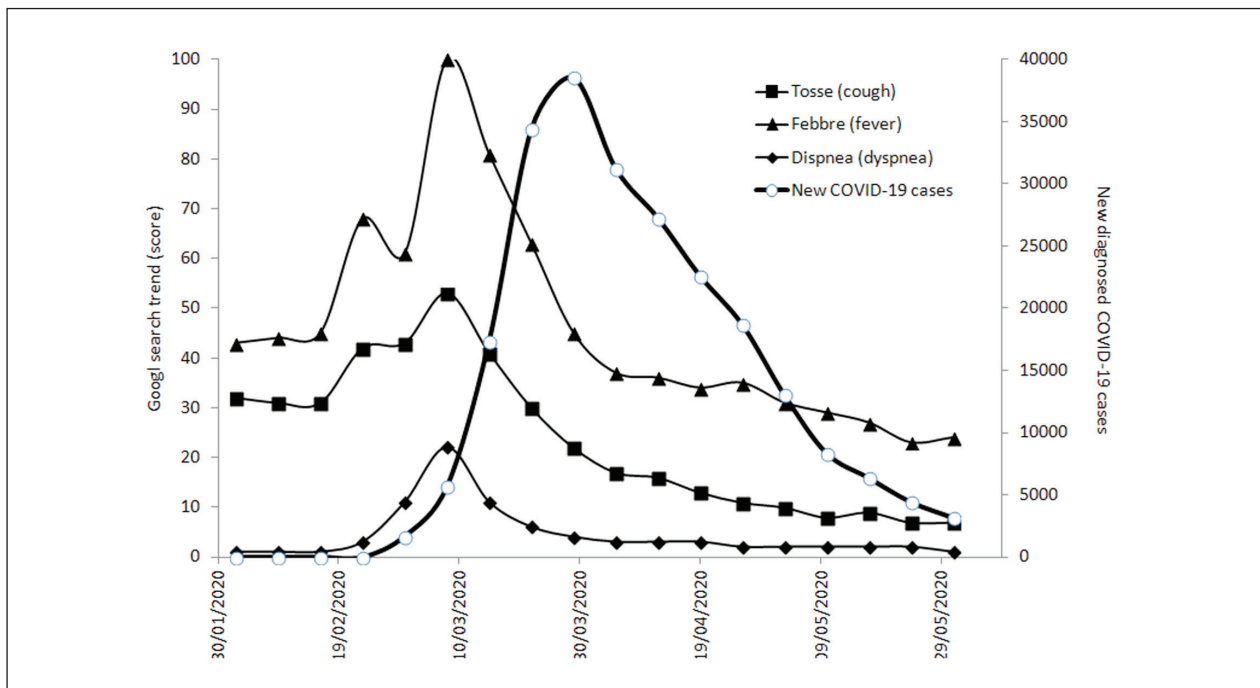
Results

The dynamics of the number of new COVID-19 cases per week and the weekly number of Google searches for "febbre" (fever), "tosse" (cough) and "dispnea" (dyspnea) in Italy throughout the study period (February-May 2020) is shown in figure 1. The search peaks for all the three symptoms anticipated by 3 weeks that of the number of newly diagnosed COVID-19 cases (March 8, 2020 vs. March 29, 2020). The weekly Google searches for "febbre" (fever), "tosse" (cough) and "dispnea" (dyspnea) exhibited a relative increase of 1.7-, 2.2- and 7.7-fold at their peak compared to the week before the first diagnosis of SARS-CoV-2 infection in an Italian resident. No significant correlation was found between the number of newly diagnosed COVID-19 cases and Google search volumes of "tosse" (cough; $p=0.277$) and "febbre" (fever; $p=0.852$), whilst a statistical significance was found with the volume of searches for "dispnea" (dyspnea; $r=0.50$; 95%CI, 0.05-0.79; $p=0.034$). Interestingly, the correlation between newly diagnosed COVID-19 cases and "tosse" (cough; $r=0.65$; 95%CI, 0.21-0.87; $p=0.008$) or "febbre" (fever; $r=0.69$; 95%CI, 0.27-0.89; $p=0.004$) become statistically significant with a 3-week delay.

A significant correlation could also be observed among the Google search volumes of all the three symptoms, as shown in figure 2. The Spearman's correlation was 0.94 (95%CI, 0.84-0.98; $p<0.001$) between "tosse" (cough) and "febbre" (fever), 0.49 (95%CI, 0.02-0.78; $p=0.041$) between "tosse" (cough) and "dispnea" (dyspnea), and 0.61 (95%CI, 0.20-0.84; $p=0.007$) between "febbre" (fever) and "dispnea" (dyspnea).

Discussion

When the most likely and frequent symptoms of an infectious disease are known in advance (e.g., cough,

Figure 1. Dynamics of the number of new coronavirus disease 2019 (COVID-19) cases per week and weekly number of Google searches for "febbre" (fever), "tosse" (cough) and "dispnea" (dyspnea) in Italy between February and May 2020

fever and dyspnea in COVID-19), a real-time analysis of Internet (e.g., Google) searches carried out within a specific geographical area may turn out to be a promising tool for predicting the dynamic of possible local outbreaks, as clearly highlighted by the results of our investigation. Overall, the weekly Google search volumes for the three most frequent COVID-19 symptoms increased between 1.7- and 7.7-fold compared to the week before the first symptomatic SARS-CoV-2 infection of an Italian citizen could be diagnosed in Codogno, Italy (February 19, 2020). It is also interesting to note that the weekly Google searches for fever already exhibited an increase of 2.3% and 4.7%, respectively, 2 and 3 weeks before the first official diagnosis of COVID-19 (Figure 1), which would contribute to suggest that the virus may have been already circulating for some weeks before the first official diagnosis has been reported. This can also be deducted from the fact that, according to the Italian National Institute of Health, the peak of 2019-2020 seasonal influenza has been recorded between the end of January and the beginning of February 2020, so that the increase of

Google searches cannot be simply attributed to the flu epidemic.

Unlike cough and fever, whose volume of Google searches was not significantly associated with the newly diagnosed COVID-19 cases because their curve was anticipated by nearly 3 weeks, the significant correlation between weekly Google searches for dyspnea and the number of new COVID-19 cases per week was predictable and is comprehensible. Dyspnea clearly reflects the development of severe lung involvement, in form of a frequently bilateral interstitial pneumonia, which typically develops 1 to 3 weeks after SARS-CoV-2 infection (7). The fact that the term "dispnea" (dyspnea) exhibited the sharpest increase of Google searches during the Italian outbreak (i.e., nearly 8-fold) is also paradigmatic if one considers that both fever and cough are non-specific symptoms, whose appearance is associated with many other infections or non-infectious diseases, whilst dyspnea reflects a deeper, and virtually more severe, involvement of the lower respiratory tract. The early detection of the "dispnea" (dyspnea) peak in terms of Google searches should
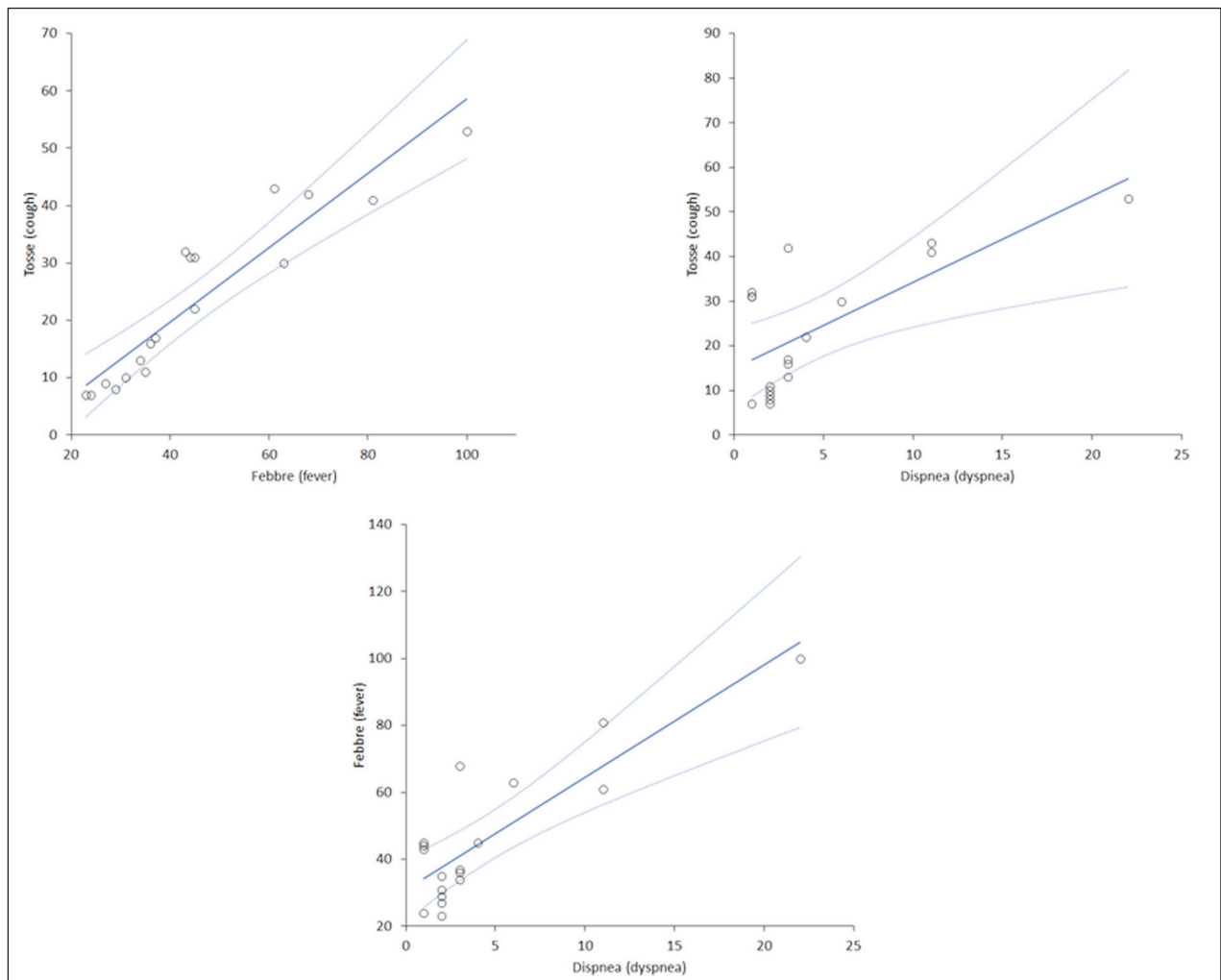
Figure 2. Spearman's correlations of Google search volumes for "tosse" (cough), "febbre" (fever) and "dispnea" (dyspnea) in Italy between February and May 2020

have probably persuaded health authorities to anticipate the establishment of restrictive measures and social distancing. The good inter-correlation found among the three symptoms (Figure 2) would also lead us to hypothesize that these Google searches in Italy were actually coordinated, and thus probably triggered by the same pathology (i.e., SARS-CoV-2 infection).

Notably, our data are in keeping with those reported by a few number of other studies. For example, Yuan et al found that several COVID-19-related specific search terms in Google Trends ("COVID", "COVID pneumonia", "COVID heart") were significantly correlated with both daily incidence and mortality of COVID-19 in the US, but with a nearly 12-day

and 19-day delay (11), which is a period very similar to that observed in our study for the first two respiratory symptoms (cough and fever). Analogous data have been reported by Panuganti et al (12), who also found a nearly 3-week delay and an ensuing good correlation between COVID-19 incidence and volume of Google searches for fever (r=0.749), cough (r=0.629) and shortness of breath (r=0.732) in the US. In another interesting Italian study, Ciaffi et al attempted to correlate the official nationwide data on intensive care (ICU) admissions and deaths for COVID-19 with the volume of Google searches for "tosse" (cough) and "febbre" (fever) (13). Even in this case a lag period of 1 to 2 weeks could be clearly observed between Google

search volumes and ICU admissions or deaths, after which all the correlations become statistically significant (all ≥0.71). Finally, Higgins et al convincingly showed that the volume of Google and Baidu (i.e., a Chinese search engine) searches for many symptoms characterizing SARS-CoV-2 infection were capable to predict by 12 days both the new daily confirmed cases of SARS-CoV-2 infections, as well as and the number of COVID-19 deaths worldwide (14).

We can hence conclude that continuously monitoring the volume of Google searches, and accurately mapping their geographic origin (i.e., region and/or province), may be regarded as a potentially valuable instrument to help identifying local recrudescence or the probable "second wave" of COVID-19 (15), at least until this infectious disease will be certainly defeated.

Conflict of interest: Each author declares that he or she has no commercial associations (e.g. consultancies, stock ownership, equity interest, patent/licensing arrangement etc.) that might pose a conflict of interest in connection with the submitted article

References

1. Cucinotta D, Vanelli M. WHO Declares COVID-19 a Pandemic. Acta Biomed 2020;91:157-160.
2. Franchini AF, Auxilia F, Galimberti PM, Piga MA, Castaldi S, Porro A. COVID 19 and Spanish flu pandemics: All it changes, nothing changes. Acta Biomed 2020;91:245-250.
3. Signorelli C, Fara GM. COVID-19: Hygiene and Public Health to the front. Acta Biomed 2020;91:7-8.
4. Nicola M, Alsafi Z, Sohrabi C, et al. The socio-economic implications of the coronavirus pandemic (COVID-19): A review. Int J Surg 2020;78:185-193.
5. Lippi G, Henry BM, Bovo C, Sanchis-Gomar F. Health risks and potential remedies during prolonged lockdowns for coronavirus disease 2019 (COVID-19). Diagnosis (Berl) 2020;7:85-90.
6. Mattiuzzi C, Lippi G. Which lessons shall we learn from the 2019 novel coronavirus outbreak? Ann Transl Med 2020;8:48.
7. Lippi G, Sanchis-Gomar F, Henry BM. Coronavirus disease 2019 (COVID-19): the portrait of a perfect storm. Ann Transl Med 2020;8:497.
8. Lippi G, Mattiuzzi C, Cervellin G. Is Digital Epidemiology the Future of Clinical Epidemiology? J Epidemiol Glob Health 2019;9:146.
9. 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;7:185-189.
10. Argenziano MG, Bruce SL, Slater CL, et al. Characterization and clinical course of 1000 patients with coronavirus disease 2019 in New York: retrospective case series. BMJ 2020;369:m1996.
11. Yuan X, Xu J, Hussain S, Wang H, Gao N, Zhang L. Trends and prediction in daily incidence and deaths of COVID-19 in the United States: a search-interest based model. Explor Res Hypothesis Med 2020;5:1-6.
12. Panuganti BA, Jafari A, MacDonald B, DeConde AS. Predicting COVID-19 Incidence Using Anosmia and Other COVID-19 Symptomatology: Preliminary Analysis Using Google and Twitter. Otolaryngol Head Neck Surg. 2020 Jun 2:194599820932128. doi: 10.1177/0194599820932128. Epub ahead of print.
13. Ciaffi J, Meliconi R, Landini MP, Ursini F. Google trends and COVID-19 in Italy: could we brace for impact? Intern Emerg Med. 2020 May 25:1–5. doi: 10.1007/s11739-020-02371-7.
14. Higgins TS, Wu AW, Sharma D, Illing EA, Rubel K, Ting JY; Snot Force Alliance. Correlations of Online Search Engine Trends With Coronavirus Disease (COVID-19) Incidence: Infodemiology Study. JMIR Public Health Surveill 2020;6:e19702.
15. Wise J. Covid-19: Risk of second wave is very real, say researchers. BMJ 2020;369:m2294.