Infodemiology of flu: Google trends-based analysis of Italians’ digital behavior and a focus on SARS-CoV-2, Italy

OMAR ENZO SANTANGELO1, SANDRO PROVENZANO2, VINCENZA GIANFREDI3
1Azienda Socio-Sanitaria Territoriale di Lodi, Lodi, Italy; 2 Azienda Ospedaliera Universitaria Policlinico “P. Giaccone”, Palermo, Italy; 3 School of Medicine, Vita-Salute San Raffaele University, Milan, Italy

Keywords
Flu • SARS-CoV-2 • Google trends • Big data • Medical informatics computing • Italy

Introduction

Influenza (or even flu) is a viral infectious disease causing a respiratory tract infection that cause a high burden in terms of direct and indirect costs, therefore it is still a public health concern [1]. Indeed, influenza viruses are characterized by the antigenic drift, that is responsible for the annual variability of the virus genome, which in turn is the reason why people can get the flu more than one time in their life [2]. Another characteristic of the flu viruses is the seasonality. Indeed, flu viruses are most common during the fall and winter, with a peak activity between December and February [2]. During the last season 2019-2020, an elevated influenza-like-illness have been detected. This excess of cases is due to a novel Coronavirus, the SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2), that was firstly identified in China, Wuhan, in the province of Hubei in December 2019 [3]. SARS-CoV-2 is responsible for a disease defined COVID-19 (where “CO” means corona, “VI” means virus, “D” means disease and “19” indicates the year in which it occurred) previously known as “2019 novel coronavirus”. Flu and the new influenza-like-illness are both respiratory illnesses caused by different viruses: influenza and the new SARS-CoV-2 [4].

The two infectious diseases spread from person-to-person via respiratory droplets emitted when people cough, sneeze or talk (close contact increases risk of transmission), landing in the upper respiratory tract of people nearby [5]. Moreover, the two types of illnesses have similar symptoms, making the differential diagnosis quite complex. In most of the cases there are symptoms variously intense including runny or stuffy nose, fever, cough, and more serious symptoms as pneumonia, bacterial infections, or hospitalizations; even if the new SARS-CoV-2 might range from a complete asymptomatic presentation to a highly complicated pulmonary and multi-organs failure, showing a more severe manifestation and causing thousands of deaths [6].

The first Italian patient tested positive to SARS-CoV-2 was detected in February in Lombardy region [7, 8]. Since the 23rd of February 2020, 225,435 total cases and 31,908 deaths have been recorder in Italy [9]. In the weeks ahead, we have seen the exponential increase of new cases and deaths for COVID-19 and the number of affected countries climb even higher. However, these numbers might be underestimated since collected through the classical surveillance systems, that are largely affected by under-diagnosis and under-reporting [10].

Nowadays, flourishing evidence is focusing on the adoption of potential novel surveillance systems based on disease-related internet activity traces [11-16] in order to monitor in a fast and cheap fashion the spread of (emerging and old) infectious diseases. Therefore, the aim of the current study was to assess if the frequency of the Italian general population searches for influenza, using the Google Trends, are aligned with Italian National Institute of Health (ISS - Istituto Superiore di Sanità) influenza cases and deaths. Moreover, we also assessed

Results. The strongest correlation between GT search and influenza cases was found at a lag of +1 week particularly for the period 2015-2019. A strong correlation was also found at a lag of +1 week between influenza death and GT search. About the correlation between GT search and SARS-CoV-2 new cases the strongest correlation was found at a lag of +3 weeks for the term flu.

Conclusion. In the last years research in health care has used GT data to explore public interest in various fields of medicine. Caution should be used when interpreting the findings of digital surveillance.
if there was a correlation between flu symptoms search volume and influenza cases and deaths. Lastly, due to the overlap with the spread of the new SARS-CoV-2, we evaluate the distribution over time and the correlation between Google search volume of flu and flu symptoms with reported cases of SARS-CoV-2 in Italy.

Materials and methods

A cross-sectional study design was used. The reported cases of flu were selected from October 2015 to April 2020. The reported deaths of flu were selected from October 2016 to April 2019. Every week from the 42nd week of the current year to the 17th week of the following year the ISS issues a bulletin with the flu cases reported in the previous week [17]. The reported cases of SARS-CoV-2 were selected from 24 February 2020 (9th week of 2020) to the end of 17th week of the following year [18]. The data have been aggregated by week.

Data on Internet searches have been obtained from Google Trends (GT) based on Google Search, the most widely used internet search engine, analyzes the popularity research topics in Google using graphs to compare the search volume of different queries over time and across different geographical locations [18]. We used the following Italian search terms in the “Health” category: “Influenza” (“Flu” in English) and “sintomi influenza” (“Symptoms of Flu” in English). Three time-frame elapsing have been extracted partly overlapping. The first from October 12, 2015 to April 28, 2019, named “2015-2019 period”, the second from October 12, 2015 to April 26, 2020, named “2015-2020 period” and finally the third from October 17, 2016 to April 28, 2019, named “2016-2019 period”. The data have been aggregated by week. According to the selected period, the relative search volume (RSV) changes, because it is a relative index. The file in “.CSV” format has been downloaded. GT produces relative search volume (RSV) scaled to the highest search proportion week, which is computed as the percentage of queries concerning a particular term for a specific location and time period, where 100 is the maximum value and 0 is the minimum value. Thus, RSV allows for directly comparing search volume across search terms.

The data coincide temporally with the weekly incidence reported in the epidemiological bulletins of the ISS; then, the data extracted from GT have been moved over time (Lag), one week in the future and one week in the past. Cross-correlation results are obtained as product-moment correlations between the two-time series. The advantage of using cross-correlations is that it accounts for time dependence between two time-series variables. Statistical analyses have been performed using the Spearman’s rank correlation coefficient (rho). The statistical significance level for the analyses has been fixed in 0.05. The data have been analyzed using the STATA statistical software, version 14 [19].

In the Tables, the wording “+1” means that we have moved the extracted data from Google one week in the future. In other words, Google anticipated the data by one week in relation to the comparison (for example the number of new cases of flu). Reverse speech for lag-1.

Results

Influenza-related digital behavior showed an increasing trend throughout the study period (from 2015 to 2019), with a peak during the epidemic year 2017, for influenza search term, and year 2019 for influenza symptoms search term. The temporal correlation between influenza cases reported by ISS and GT-based RSV was very large (rho > 0.70, highly statistically significant with p-values < 0.001) for the two study periods 2015-2019 and 2015-2020. The strongest correlation between Google trends search (for both flu and symptoms of flu) and the reported influenza cases from ISS was found at a lag of +1 week particularly for the period 2015-2019 (rho = 0.92 for flu and rho = 0.87 for symptoms), as showed in Table I. The correlation between influenza cases and Google trends search was still strong for the period 2015-2020 even if slightly attenuated compare to 2015-2019 (rho = 0.77 for flu and rho = 0.82 for symptoms, p-values < 0.001), as reported in Table I. In addition, a strong correlation was also found at a lag of +1 week between influenza death and Google trends search (rho = 0.84 for flu and rho = 0.81 for symptoms, p-values < 0.001), as described in Table II. These statistically significant patterns were depicted in Figure 1 (2015-2019 period) and Figure 2 (2015-2020 period) for influenza cases and in Figure 3 for deaths (2016-2019 period). When examining the

| Lag in week compared to deaths reported by ISS | 2015-2019 period | 2015-2020 period |
|---------------------------------------------|------------------|------------------|
| Flu                                         | 0.8257*          | 0.7521*          |
| Symptoms of Flu                             | 0.7657*          | 0.7991*          |

* p-value < 0.001.

| Lag in week compared to deaths reported by ISS | 2016-2020 period |
|---------------------------------------------|------------------|
| Flu                                         | 0.6015*          |
| Symptoms of Flu                             | 0.6177*          |

* p-value < 0.001.
correlation between Google trends search and SARS-CoV-2 new cases reported by the Ministry of Health, the strongest correlation was found at a lag of +3 weeks for the search term flu (rho = 0.80, p-value < 0.01) as showed in Table III. This statistical pattern is confirmed in Figure 4, where the Google research volume for flu and flu symptoms were plotted considering both influenza cases and new SARS-CoV-2 cases. In this figure, the search volume for flu and flu symptoms shows a double peak. The first is concurrent to the influenza cases peak, the second is precedent to the reported new SARS-CoV-2 cases.
Discussion

In this study we found a large correlation between flu – cases and deaths – occurred in Italy and reported by ISS and GT research for both flu and flu symptoms. This result remains consistent even using different time lag, becoming more stronger when a time lag of +1 week was adopted. Due to the overlap between clinical symptomatology and season during which flu and SARS-CoV-2 spread among population (in Italy), we further assessed the correlation between Google trends search and SARS-CoV-2 new cases reported by the Ministry of Health. A strong correlation was found in this analysis as well, with the strongest correlation at a lag of +3 weeks. This means that at the beginning of the SARS-CoV-2 pandemic, people affected by COVID-19 searched on Internet information related to flu, probably confusing the two diseases. Moreover, it confirms the hypothesis that people frequently use internet for searching health related information.

On Feb 22, 2020 an Editorial on the scientific journal The Lancet entitled “COVID-19: fighting panic with information” focused on the real risk of sanitary emergency saying there could be no way to prevent a COVID-19 pandemic in this globalized time, but verified information is the most effective prevention against the disease of panic [20]. Thus, from the first moment it became clear that we were struggling not only with an epidemic, but also with an infodemic [21]. A global epidemic of misinformation – spreading mainly through social media platforms and fake news – poses a serious problem for public health although the WHO is leading the effort to stem of public emergency. As a public health emergency of international concern, the COVID-19 has drawn global attention and response. In the scenario of COVID-19 pandemic [22], it is extremely important to promote the flu vaccination during the next campaign increasing the opinion, knowledge and attitude of health workers and the population with dedicated health policies [23-25].

This is true for several reasons, firstly, it could directly reduce the burden of the flu pandemic (diminishing and limiting the number of patients hospitalized because of flu), secondly, reducing the number of patients hospitalized because of flu, it will ameliorate the hospital organization of patients eventually positive to SARS-CoV-2. Thirdly, in flu immunized patients the differential diagnosis between flu and SARS-CoV-2 could be facilitate improving the clinical management of these patients [26]. In planning these measures, considerations should be given to minimizing the excess risk of morbidity and mortality from vaccine-preventable diseases (VPDs). Such outbreaks may result in VPD-related deaths and an increased burden on health systems already strained by the response to the COVID-19 outbreak [27].

In this context, the big data generated by web searches become increasingly important in the search for new surveillance systems based on digital epidemiology. According to Marcel Salathe the term digital epidemiology is a field of study that uses data that was generated outside the public health system, i.e. with data that was not generated with the primary purpose of doing epidemiology [28].
In a similar way to the results of the scientific literature our study shown that digital epidemiology, integrated to modern infectious disease surveillance systems, aim to employ the speed and scope of big data in an attempt to provide global health security [29].

Our study has strengths and limitations. Google Trends data helps identify developing interests in different public health topics including known and emerging infectious diseases (i.e. flu and SARS-CoV-2) or related clinical and diagnostic aspects and screening tests. Internet searches can be an important source for generating hypotheses about knowledge, attitudes, and practices in public health topics; evaluating changes in information seeking after targeted interventions to prevent the spread of emerging infectious diseases or stem vaccine-preventable diseases. In this field, public health interventions could be evaluated almost immediately and with a minimal expenditure.

The mass media (TV, radio, and social network) may influence the online population’s researches [30]. Indeed, the spike of Internet searches, for example, for “Flu” or “symptoms of Flu” may be attributed to various factors as an increased number of cases in the community and increased attention given by the mass media. Indeed, the data is only available for States and selected metropolitan areas limiting comparability with rural areas or areas with a low search volume, represented by the areas where Internet is less widespread among the population. Finally, Google Trends data are anonymous limiting the utility in examining subgroups or disparities among populations.

Thus, even considering the potential intrinsic limits of this analysis, our results show how this data might be extremely useful, encouraging the spread of future researches at each country level. The results of this study suggest that Google Trends based surveillance systems might be relevant for public health and for public health workers [31], because these novel systems have the potentiality to inform how the public is interested in searching health related information [32]. The info surveillance systems, based on the intrinsic characteristic of dynamicity, have the power to inform and provide near real-time data, useful to plan public health interventions [33]. Public health workforce should enforce communication and internet-based skills in order to fruitfully use a new and cheap technology able to support interventions design and implementation [34].

**Conclusions**

How key information must be communicated to the public during the next phase of the pandemic is critical. In the last years research in health care has used GT data to explore public interest and trends in various fields of medicine. It is evident that caution should be used when interpreting the findings of Google Trends digital surveillance.

**Acknowledgements**

Funding sources: this research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

**Conflict of interest statement**

The authors declare no conflict of interest.

**Authors’ contributions**

OES conceived, designed, coordinated and supervised the research project. OES, SP and VG performed the data quality control, optimized the informatics database, performed the statistical analyses and evaluated the results. OES, SP and VG wrote the manuscript. All Authors revised the manuscript and gave their contribution to improve the paper. All authors read and approved the final manuscript.

**References**

[1] Rizzo C, Bella A, Viboud C, Simonsen L, Miller MA, Rota MC, Salmaso S, Ciofi degli Atti ML. Trends for influenza-related deaths during pandemic and epidemic seasons, Italy, 1969-2001. Emerg Infect Dis 2007;13:694-9. https://doi.org/10.3201/eid1305.061309

[2] Centers for Disease Control and Prevention. Influenza (Flu). 2019. https://www.cdc.gov/flu/about/season/flu-season.htm (accessed 18 May 2020).

[3] Brussow H. The novel coronavirus - a snapshot of current knowledge. Microb Biotechnol 2020;13:607-12. https://doi.org/10.1111/1751-7915.13557

[4] Pascarella G, Strumia A, Pliego C, Bruno F, Del Buono R, Costa F, Scarlata S, Agrò FE. COVID-19 diagnosis and management: a comprehensive review. J Intern Med 2020. https://doi.org/10.1111/joim.13091

[5] Ningthoujam R. COVID 19 can spread through breathing, talking, study estimates. Curr Med Res Pract 2020. https://doi.org/10.1016/j.mpr.2020.05.003

[6] Huang Q, Deng X, Li Y, Sun X, Chen Q, Xie M, Liu S, Qu H, Liu S, Wang L, He G, Gong Z. Clinical characteristics and drug therapies in patients with the common type coronavirus disease 2019 in Hunan, China. Int J Clin Pharm 2020;1-9. https://doi.org/10.1007/s11096-020-01031-2

[7] Signorelli C, Odone A, Gianfredi V, Bossi E, Buccu D, Oradini-Alacreu A, Frascella B, Capraro M, Chiappa F, Blandi L, Ciceri F. COVID-19 mortality rate in nine high-income metropolitan regions. Acta Biomed 2020:91:7-18. https://doi.org/10.23750/abm.v91i19-S.10134

[8] Signorelli C, Odone A, Gianfredi V, Bossi E, Buccu D, Oradini-Alacreu A, Frascella B, Capraro M, Chiappa F, Blandi L, Ciceri F. The spread of COVID-19 in six western metropolitan regions: a false myth on the excess of mortality in Lombardy and the defense of the city of Milan. Acta Biomed 2020;91:23-30. https://doi.org/10.23750/abm.v91i2-9600.

[9] Ministry of Health. Covid-19 - Situazione in Italia. 2020. http://www.salute.gov.it/portale/nuovocoronavirus/dettaglioContenutiNuovoCoronavirus.jsp?area=nuovoCoronavirus&id=5351&lingua=italiano&menu=nuoto (accessed 17 May 2020).

[10] Gianfredi V, Moretti M, Fusco Moffa I. Burden of measles using disability-adjusted life years. Umbria 2013-2018. Acta Biomed 2020;91:48-54. https://doi.org/10.23750/abm.v91i3-S.9412
[11] Bragazzi NL, Barberis I, Rosselli R, Gianfredi V, Nucci D, Moretti M, Salvatori T, Martucci G, Martini M. How often people google for vaccination: qualitative and quantitative insights from a systematic search of the web-based activities using Google Trends. Hum Vaccin Immunother 2017;13:464-9. https://doi.org/10.1080/21645515.2017.1264742

[12] Gianfredi V, Bragazzi NL, Mahamid M, Bisharat B, Mahroum N, Amital H, Adawi M. Monitoring public interest toward pertussis outbreaks: an extensive Google Trends-based analysis. Public Health 2018;165:9-15. https://doi.org/10.1016/j.puhe.2018.09.001

[13] Gianfredi V, Bragazzi NL, Nucci D, Martini M, Rosselli R, Minelli L, Moretti M. Harnessing big data for communicable tropical and sub-tropical disorders: implications from a systematic review of the literature. Front Public Health 2018;6:90. https://doi.org/10.3389/fpubh.2018.00090

[14] Mahroum N, Bragazzi NL, Sharif K, Gianfredi V, Nucci D, Rosselli R, Brigo F, Adawi M, Amital H, Wadat A. Leveraging google trends, twitter, and Wikipedia to investigate the impact of a celebrity’s death from rheumatoid arthritis. J Clin Rheumatol 2018;24:188-92. https://doi.org/10.1097/RHU.0000000000000692

[15] Provenzano S, Santangelo OE, Giordano D, Alagna E, Piazza D, Genovese D, Calamusa G, Firenze A. Predicting disease outbreaks: evaluating measles infection with wikipedi trends. Recenti Prog Med 2019;109:292-6. https://doi.org/10.1701/3182.31610

[16] Santangelo OE, Provenzano S, Piazza D, Giordano D, Calamusa G, Firenze A. Digital epidemiology: assessment of measles infection through Google Trends mechanism in Italy. Ann Ig 2019;31:385-91. https://doi.org/10.7416/ai.2019.2300

[17] Istituto Superiore di Sanità. Rapporto Influen. Aggiornamento 8 maggio 2020. 2020. https://old.iss.it/site/RMI/influenet/pagine/rapportoinfluenet.aspx (accessed 17 May 2020).

[18] Google Trends. Scopri quali ricerche si fanno nel mondo. 2020. https://trends.google.it/trends/?geo=IT (accessed 17 May 2020).

[19] StaCorp. StaStat Statistical Software. In: Station C, editor. StaCorp LP 2015.

[20] COVID-19: fighting panic with information. Lancet 2020;395:537. https://doi.org/10.1016/S0140-6736(20)30379-2

[21] Gianfredi V, Provenzano S, Santangelo OE. What can internet users’ behaviours reveal about the mental health impacts of the COVID-19 pandemic? A systematic review. Public Health 2021;198:44-52. https://doi.org/10.1016/j.puhe.2021.06.024

[22] Gianfredi V, Pennisi F, Lume A, Ricciardi GE, Martinelli L, Azzurri CE. Digital epidemiology: assessment of mass vaccination centers in COVID-19 times: a rapid review of literature. Vaccines (Basel) 2021;9(6). https://doi.org/10.3390/vaccines9060574

[23] Gentile L, Cuda A, Dallagiocoma G, Provenzano S, Santangelo OE, Navaro M, D’Alloiso F, Gianfredi V. Opinion, knowledge and attitude of public health residents towards the new mandatory vaccination law in Italy. J Public Health (Berl) 2020. https://doi.org/10.1007/s10389-019-01171-4

[24] Gianfredi V, Dallagiocoma G, Provenzano S, Santangelo OE. Factors predicting health science students’ willingness to be vaccinated against seasonal flu during the next campaign. Ann Ist Super Sanita 2019;55:209-16. https://doi.org/10.4415/ANN_19_03_03

[25] Santangelo OE, Cederone F, Gentile L, Provenzano S, Dallagia coma G, Armetta F, Gianfredi V. Reasons behind flu vaccine acceptance and suggested interventions to promote flu vaccination acceptance among healthcare workers. J Infect Prev 2021;22:132-5.https://doi.org/10.1177/1757177420976804

[26] Gianfredi V, Santangelo OE, Provenzano S. The effects of COVID-19 pandemic on the trend of measles and influenza in Europe. Acta Biomed 2021;92:e2021318. https://doi.org/10.23750/abm.v92i1.11558

[27] World Health Organization. Guidance on routine immunization services during COVID-19 pandemic in the WHO European Region. 2020. http://www.euro.who.int/__data/assets/pdf_file/0004/433813/Guidance-routine-immunization-services-COVID-19-pandemic.pdf?ua=1 (accessed 20 May 2020).

[28] Salathe M. Digital epidemiology: what is it, and where is it going? Life Sci Soc Policy 2018;14:1. https://doi.org/10.1186/s40504-017-0065-7

[29] Eckmanns T, Fuller H, Roberts SL. Digital epidemiology and global health security; an interdisciplinary conversation. Life Sci Soc Policy 2019;15:2. https://doi.org/10.1186/s40504-019-0091-8

[30] Gianfredi V, Santangelo OE, Provenzano S. Correlation between flu and Wikipedia’s pages visualization. Acta Biomed 2021;92:e2021056. https://doi.org/10.23750/abm.v92i1.9790

[31] Gianfredi V, Odone A, Fiaccini D, Rosselli R, Battista T, Signorelli C. Trust and reputation management, branding, social media management nelle organizzazioni sanitarie: sfide e opportunità per la comunità igienistica italiana. J Prev Med Hyg 2019;60:E108-E9.

[32] Bragazzi NL, Gianfredi V, Villarini M, Rosselli R, Nasr A, Husein A, Martini M, Behzadifar M. Vaccines meet big data: state-of-the-art and future prospects. From the classical 3Is (“Isolate-Inactivate-Inject”) Vaccinology 1.0 to Vaccinology 3.0. Vaccinomics, and beyond: a historical overview. Front Public Health 2018;6:62. https://doi.org/10.3389/fpubh.2018.00062

[33] Gianfredi V, Grisi C, Nucci D, Parisi V, Moretti M. [Communi cation in health]. La comunicazione in sanità. Recent Prog Med 2018;109:374-83. https://doi.org/10.1701/2955.29706

[34] Gianfredi V, Balzarini F, Gola M, Mangan S, Carpegna LF, Colucci ME, Gentile L, Piscitelli A, Quattrone F, Scuri S, Mantovani LG, Auxilia F, Castaldi S, Capolongo S, Pelissero G, Odone A, Signorelli C. Leadership in public health: opportunities foryoung generations within scientific associations and the experience of the “Academy of Young Leaders”. Front Public Health 2019;7:378. https://doi.org/10.3389/fpubh.2019.00378

Received on July 27, 2020. Accepted on July 29, 2020.

Correspondence: Sandro Provenzano, Azienda Ospedaliera Universitaria Policlinico “P. Giaccone”, via del Vespro 129, 90127 Palermo (PA), Italia - Tel.: +390916553641 - Fax: +390916553697 - E-mail: provenzonosandro@hotmail.it

How to cite this article: Santangelo OE, Provenzano S, Gianfredi V. Infodemiology of flu: Google trends-based analysis of Italians’ digital behavior and a focus on SARS-CoV-2 Italy. J Prev Med Hyg 2021;62:E586-E591. https://doi.org/10.15167/2421-4248/jpmh2021.62.3.1704

© Copyright by Pacini Editore Srl, Pisa, Italy

This is an open access article distributed in accordance with the CC-BY-NC-ND (Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International) license. The article can be used by giving appropriate credit and mentioning the license, but only for non-commercial purposes and only in the original version. For further information: https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en