Focus on

Digital epidemiology and infodemiology of handfoot-mouth disease (HFMD) in Italy. Disease trend assessment via Google and Wikipedia

Omar Enzo Santangelo¹, Vincenza Gianfredi², Sandro Provenzano³, Fabrizio Cedrone⁴ ¹Regional Health Care and Social Agency of Lodi, ASST Lodi, Lodi, Italy; ²Department of Biomedical Sciences for Health, University of Milan, Milan, Italy; ³Local Health Unit of Trapani, ASP Trapani, Trapani, Italy; ⁴Hospital Management, Local Health Unit of Pescara, Pescara, Italy

Abstract. *Background and aim:* The study aimed to evaluate the epidemiological trend of hand, foot and mouth disease (HFMD) in Italy using data on Internet search volume. *Methods:* A cross-sectional study design was used. Data on Internet searches were obtained from Google Trends (GT) and Wikipedia. We used the following Italian search term: "Malattia mano-piede-bocca" (Hand-foot-mouth disease, in English). A monthly time-frame was extracted, partly overlapping, from July 2015 to December 2022. GT and Wikipedia were overlapped to perform a linear regression and correlation analyses. Statistical analyses were performed using the Spearman's rank correlation coefficient (rho). A linear regression analysis was performed considering Wikipedia and GT. *Results:* Search peaks for both Wikipedia and GT occurred in the months November-December during the autumn-winter season and in June during the spring-summer season, except for the period from June 2020 to June 2021, probably due to the restrictions of the COVID19 pandemic. A temporal correlation was observed between GT and Wikipedia search trends. *Conclusions:* This is the first study in Italy that attempts to clarify the epidemiology of HFMD. Google search and Wikipedia can be valuable for public health surveillance; however, to date, digital epidemiology cannot replace the traditional surveillance system (www.actabiomedica.it).

Key words: Hand-foot-mouth disease, HFMD, Italy, digital epidemiology, Google, Wikipedia, epidemiology, infectious diseases, infodemiology, medical informatics computing

Introduction

Hand, foot and mouth disease (HFMD) is a common viral infection that primarily affects infants and children. It is generally caused by Coxsackievirus A type 16; however, the infection can also be caused by many other strains of coxsackievirus and human enterovirus 71 (1,2). Usually, HFMD has an acute onset with a low-grade fever, reduced appetite, and general malaise. Moreover, the enanthem on the skin or mucosa is responsible for the pain of hands, feet, and mouth (1). In most of the cases, symptoms are mild, however; in

a smaller number of cases might have atypical features like concomitant aseptic encephalitis, meningitis, encephalomyelitis, myocarditis, delayed acute paralysis and even death in severe cases (3-5). This infection occurs worldwide without a particular geographical pattern. Because children (particularly those younger than five years of age) tend to be infected mostly, outbreaks within family can also be recorded. Moreover, high rate of infection is due to poor hygiene, social contacts frequency (6), and meteorological factors, as temperature and high humidity (cases tended to occur more frequently during late spring and early summer) (6-8).

Data from the large-scale surveillance reported that approximately 13 million cases of HFMD were recorded in China in the period 2008-2015. Out of these, 12,000 were severe cases and more than 3,300 deaths were documented (5). In America and Europe, the incidence of HFMD is not routinary monitored as there is no active surveillance system (9). Therefore, surveillance is based on passive reporting systems which tend to underestimate the real number of cases and do not give a clear epidemiological picture (9). In this perspective, online epidemiological surveillance might support in understanding trends. Actually, online epidemiological surveillance has received considerable attention over the last few years and could help for a first general overview in this field. These new data offer interesting insights to focus on as they trace Internet activities. In particular, in this case data refer to the Internet users' health-related searches. These new surveillance systems statistically combine traditional surveillance data with Internet activity in order to explore public interest and inform mathematical models that can predict the ongoing epidemics (10). As for instance, previous evidence compared web pages visualizations (most frequently Wikipedia web pages) (11-13), or search queries (most frequently Google Trends) (14-16) with surveillance data. This emerging surveillance approach, also known as digital epidemiology, has become even more attractive because faster and cheaper compared to the traditional surveillance systems. According to Marcel Salathe the term digital epidemiology is a field of study that uses data generated outside the public health system, i.e. data that were not generated with the main purpose of doing epidemiology (17). The assumption behind this approach is that an increase in disease cases is followed by Internet users' search on the disease, which finally corresponds to high internet search generated volume.

Although this field, known as "infodemiology" or "infoveillance" is still in its infancy, it can be envisioned that its implications could overcome some of the problems of traditional systems because internet-based surveillance systems is based on real-time monitoring. Therefore, the aim of this study was to evaluate the epidemiological trend of the HFMD in Italy through the searches that users make on the internet, assessing whether a correlation/association between searches in Google and Wikipedia exist.

Material and methods

A cross-sectional study design was used. Data on Internet searches were obtained from Google Trends (GT) based on Google Search, the most widely used internet search engine (18). We used the following Italian search term: "Malattia manopiede-bocca" (Hand-foot-mouth disease, in English). We also searched for terms that might be associated with HFMD as they represent more common symptoms "febbre" (fever, in English), "vescicole sulle mani sui piedi e sulla bocca"/"vescicole mani piedi bocca" (vescicles on hand, feet and mouth, in English), "ulcera" (sore, in English); specifically regarding "vescicole sulle mani sui piedi e sulla bocca" or "vescicole mani piedi bocca" The data related to the search is not sufficient to be shown in Google Trends, the tool does not provide data. One monthly time-frame elapsing was extracted partly overlapping, from July 2015 to December 2022 (July 2015 is the beginning of data availability in Wikipedia). 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. From Wikipedia (19) it is possible to know how many times a specific page is viewed by users; data were extracted as monthly data corresponding to the monthly report of Google's RSV. Using this tool, the number of monthly views by users from the July 2015 to December 2022 of the Wikipedia' pages "Malattia manopiede-bocca" (Hand-foot-mouth disease, in English) were extracted. We also searched for pages that might be associated with HFMD as they represent more common symptoms "febbre" (fever, in English), "vescicola" (vescicle, in English), "ulcera" (sore, in English). The files in ".CSV" format has been downloaded.

We overlapped Google Trends and Wikipedia data to perform a linear regression and correlation analysis. 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 timeseries variables. Statistical analyses were performed using the Spearman's rank correlation coefficient (rho). According to a rule of thumb there is a strong correlation if rho> 0.7, moderate correlation if the value of rho is between 0.3 and 0.7 and weak correlation if rho< 0.3 (20). A linear regression was performed considering Wikipedia searches as dependent variable and Google trends RSV as independent variable, results are expressed as coefficient with 95% confidence intervals (95% CI). Potential autocorrelation was ascertained through the calculation of the Durbin-Watson (DW) statistics. The DW test is a statistic test used to detect the presence of autocorrelation in the residuals (prediction errors) from a regression analysis (21). The DW test statistic or d always lies between 0 and 4. If the d is substantially less than 2, there is evidence of positive serial correlation, while values greater than 2 suggest no autocorrelation. Representative linear model and correlation chart of the data were calculated, also calculating the R² of the model. The statistical significance level for the analyses was 0.05. The data were analyzed using the STATA statistical software, version 14 (22) and Microsoft Excel [®]. The data download and analyses have been done the 13th of February, 2023.

Results

The raw data for GT and Wikipedia are shown in Figure 1 and 2. Figure 3 and 4 show the correlation chart between Wikipedia searches and Google's RSV and linear regression between Wikipedia searches and Google's RSV for the search term HFMD, respectively. Figure 5 shows the RSV of GT in relation to the Italian Regions in the study period.

Search peaks for both Wikipedia and GT occurred in November-December during the autumnwinter season, and in June during the spring-summer season, except for the period from June 2020 to June 2021, probably due to the restrictions of the COVID19 pandemic, will be discussed more in the discussion chapter. A temporal correlation was observed between GT and Wikipedia search trends. Google Trends Internet search data showed strong correlation with Wikipedia, with a rho equal to 0.82 statistically significant (see Table 1). Figure 3 shows the correlation charts that best represented data, with a R² equal to 0.6722. Figure 4 shows the linear regression that best represented data, with a R^2 equal to 0.3774 (see also Table 2 for further details). As shown in Table 3, there was an increase in searches for the terms in question in Google Trends during the post-COVID pandemic period (2021-2022) compared to the preceding period



Figure 1. Interest over time of google HFMD search (from 01 July 2015 to 31 December 2022).



Figure 2. Search trend of Wikipedia page (normalized data, 100 maximum, 0 minimum) and Google RSV.



Figure 3. Correlation chart between Wikipedia searches and Google's RSV. Spearman's rank correlation coefficient was used. The red line shows model, the equation and the corresponding R^2 are shown in the red box.



Figure 4. Linear regression between Wikipedia searches and Google's RSV. The red line shows model, the equation and the corresponding R^2 are shown in the red box.



Figure 5. Italian Regions' interest in Google Trends for HFMD disease search. From 01 July 2015 to 31 December 2022. In this figure reported the top 5.

(2015-2019). Exactly the opposite instead happened for the search of the Wikipedia pages mentioned above. In the study period, Table 4 shows the interest by Region in Italy on the topic of HFMD as a disease, the regions of central-northern Italy showed greater research interest than those of the south or the islands (Sicily and Sardinia).

Discussion

In Italy there is no surveillance system for HFMD, which is an important disease for children under the age of 10 (1), with a potentially increasing disease burden for Italy, and more in general in Europe, as demonstrated in other regions of the world

	HFMD_Wiki	HFMD_GT	Fever_Wiki	Sores_Wiki	Vescicle_Wiki	Fever_GT	Sores_GT
HFMD_Wiki	1						
HFMD_GT	0.82*	1					
Fever_Wiki	0.36*	-0.03	1				
Sores_Wiki	0.55*	0.15	0.67*	1			
Vescicle_Wiki	0.55*	0.20	0.65*	0.80*	1		
Fever_GT	-0.31	-0.08	-0.07	-0.55*	-0.41*	1	
Sores_GT	0.08	0.18	-0.40*	-0.09	-0.11	-0.02	1

Table 1. Spearman's rank correlation coefficient.

**p-value*<0.001

Table 2. Linear regression models.

	Dependent variable: HFMD Wikipedia				
Independent variable	Coefficient	95% CI	p-value	Durbin-Watson	
HFMD (GT RSV)	130.66	95.11-166.22	<0.001	0.43	

Table 3. Excess relative search volume (RSV) for research terms in Google Trends and in Wikipedia (Italy). In the analyses, average estimates for 2015-2019 were assumed as the reference categories.

Research field	Delta of RSV Rate and Wikipedia page viewed
Google Trends	2015-2019 vs 2021-2022
HFMD	0.58%
Fever	49.23%
Sores	6.33%
Wikipedia	2015-2019 vs 2021-2022
HFMD	-61.40%
Vescicle	-61.20%
Fever	-47.74%
Sores	-83.07%

(4,5,9). Understanding its epidemic trend would allow to quickly implement preventive strategies which in turn would improve the management of a possible epidemic (23). Thanks to the improvement of the computing power of computers and the use of big data, infodemiological analysis is increasingly used, especially with the advent of the ongoing COVID-19 pandemic (24,25). The Italian national burden of disease for HFMD is very vague or not even known, therefore, in this pilot study we evaluated whether or not infodemiology could represent a valid option to monitor the epidemiological trend. As previously demonstrated (15-17), infodemiology - and more generally digital epidemiology - is an innovative opportunity for the monitoring of infectious diseases, preferably as a complementary strategy to traditional surveillance systems. Several studies have shown that this monitoring method is innovative, accurate, effective and reliable even in fields other than the epidemiology of infectious diseases (10,26-28).

Our study shows that in the period 2012-2022 the Internet search activities follow a constant trend with peaks in November-December during the autumnwinter season and in June during the spring-summer

Region	RSV		
Trentino-Alto Adige	100		
Lazio	94		
Umbria	93		
Marche	91		
Liguria	91		
Lombardy	90		
Abruzzo	85		
Emilia-Romagna	84		
Piedmont	83		
Veneto	82		
Friuli-Venezia Giulia	82		
Molise	75		
Sicily	74		
Tuscany	73		
Calabria	71		
Apulia	68		
Sardinia	65		
Campania	60		
Basilicata	59		
Aosta Valley	Not available		

Table 4. Italian Regions' interest in Google Trends for HFMD disease search. From 01 July 2015 to 31 December 2022.

season. An exception has been recorded for the period June 2020-June 2021 where searches carried out on Google and the Wikipedia pages visualization are at a minimum, probably because of COVID-19. As a matter of fact, during COVID-19 pandemic several preventive measures were in place, among them the lock-down that highly reduced frequency of contacts. Thus, the restrictive measures reduced also the circulation of many other airborne diseases, as demonstrated for measles and influenza (14,29), among the others. In fact, although virologically unrelated, SARS-CoV-2 and HFDM share several characteristics, such as airborne transmission via droplets or contaminated surfaces. Consequently, we can speculate that the reduction in the disease burden was accompanied by a reduction in Internet users' interest and then in related searches. In the context of the COVID-19 pandemic, the use of personal protective equipment such as masks, lockdown, and social isolation has increased the number of newborns who have not developed any immunity against HFMD, so we should expect an increase of important cases in the coming years. Moreover, our findings reveled a strong correlation and a statistically significant association between searches on Google and Wikipedia. We can interpret these results hypothesizing that Internet users search for general news on Google and then they go to in-depth sites, such as Wikipedia, to learn more about the disease, as demonstrated in earlier studies (14,17,26).

Considering searches on GT and Wikipedia (See Figures 1 and 2) it can be seen that between June 2020 and June 2021 the peaks seen in previous years are missing. After the reduction of restrictive measures (to prevent COVID-19), we see that searches on Wikipedia, and especially on Google, increased, reaching the maximum value of RSV in June 2022 and with an increasing trend for November-December 2022. We could interpret this data considering the last as a rebound effect. It has now been demonstrated that infodemiological research and the appropriate analysis of search trends in specific search engines, web platforms, and social media can reflect or anticipate the epidemiological characteristics of some disorders (30-32). Previous studies found that users consult Internet before referring to their doctor, even many days in advance (16, 33).

To our knowledge, this was the first study to investigate whether GT and Wikipedia searches on HFMD could represent a reliable proxy for the actual time course of HFMD infection in Italy. Of course, current assessments need to be handled with care and a precautionary approach needs to be used. However, the available data does not allow for a more accurate analysis.

Limitations of the study

The study has limitations: first, the data is limited to those who use Google and Wikipedia for healthrelated information. This may not represent the entire computing community; however, Google is used by more than 80% of internet users worldwide, covering the highest percentage of the population (33). Secondly, younger people are generally more digitally connected than older individuals, potentially affecting the generalizability to the whole population. Third, searches might be impacted by mediatic events that can affect Google searches. Fourth, in our study the data analyzed included the COVID-19 pandemic which may have influenced the quality and quantity of searches. Fifth, the proposed models may not be representative enough of reality because confounding factors are not taken into account. However, this is a pilot study that can offer a new insight regarding HFMD epidemiology. Despite our findings should be considered as pioneering, they might support policymakers by providing data on an unmonitored infectious disease.

Conclusions

The study, albeit with many limitations, described HFMD, GT and Wikipedia search trends and the correlations and associations between HFMD, Google and Wikipedia searches. We noted that the research volume trend of HFMD had atypical patterns during the period of COVID-19 pandemic restrictions. This could be due to considerably decrease of all diseases transmitted by direct contact or with mucus, saliva, or feces, considering the widespread use of masks and adherence to social distancing measures. However, the seasonal trend of searches with the peaks recorded during the change of season, as happens in other countries with a temperate climate, would suggest that the real epidemiology of the disease overlaps with the "digital epidemiology" of the phenomenon. Searches for infectious diseases and associated symptoms using Google search and Wikipedia can be useful for public health surveillance. However, to date, digital epidemiology cannot replace a classical surveillance system.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

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.

Authors Contribution: OES conceptualized, OES designed the study, OES performed data analysis and data extraction. OES, VG, SP and FC wrote the first draft. All authors have read and agreed to the published version of the manuscript. All authors reviewed the manuscript.

References

- 1. Esposito S, Principi N. Hand, foot and mouth disease: current knowledge on clinical manifestations, epidemiology, aetiology and prevention. Eur J Clin Microbiol Infect Dis. 2018;37(3):391-398. doi:10.1007/s10096-018-3206-x
- 2. CDC, Center for Diseases Control and Prevention. USA. Available online: https://www.cdc.gov/hand-foot-mouth /about/transmission.html (last access January 09, 2022)
- 3. Solomon T, Lewthwaite P, Perera D, Cardosa MJ, McMinn P, Ooi MH. Virology, epidemiology, pathogenesis, and control of enterovirus 71. Lancet Infect Dis. 2010;10(11):778-790. doi:10.1016/S1473-3099(10)70194-8
- Guerra AM, Orille E, Waseem M. Hand, Foot, and Mouth Disease. In: StatPearls. Treasure Island (FL): StatPearls Publishing; October 9, 2022.
- 5. Huang J, Liao Q, Ooi MH, et al. Epidemiology of Recurrent Hand, Foot and Mouth Disease, China, 2008-2015. Emerg Infect Dis. 2018;24(3):432-442. doi:10.3201 /eid2403.171303
- Ruan F, Yang T, Ma H, et al. Risk factors for hand, foot, and mouth disease and herpangina and the preventive effect of hand-washing. Pediatrics. 2011;127(4):e898-e904. doi:10.1542/peds.2010-1497
- Wang YR, Sun LL, Xiao WL, Chen LY, Wang XF, Pan DM. Epidemiology and clinical characteristics of hand foot, and mouth disease in a Shenzhen sentinel hospital from 2009 to 2011. BMC Infect Dis. 2013;13:539. doi:10.1186/1471-2334-13-539
- Nguyen NT, Pham HV, Hoang CQ, et al. Epidemiological and clinical characteristics of children who died from hand, foot and mouth disease in Vietnam, 2011. BMC Infect Dis. 2014;14:341. doi:10.1186/1471-2334-14-341
- Abedi GR, Watson JT, Nix WA, Oberste MS, Gerber SI. Enterovirus and Parechovirus Surveillance - United States, 2014-2016. MMWR Morb Mortal Wkly Rep. 2018; 67(18):515-518. doi:10.15585/mmwr.mm6718a2
- Santangelo OE, Gianfredi V, Provenzano S. Wikipedia searches and the epidemiology of infectious diseases: A systematic review. Data & Knowledge Engineering. 2022; 142: 102093. doi: 10.1016/j.datak.2022.102093
- Provenzano S, Gianfredi V, Santangelo OE. Insight the data: Wikipedia's researches and real cases of arboviruses in Italy. Public Health. 2021;192:21-29. doi:10.1016/j.puhe .2020.12.010

- Provenzano S, Santangelo OE, Giordano D, et al. Predicting disease outbreaks: evaluating measles infection with Wikipedia Trends. Recenti Prog Med. 2019;110(6):292-296. doi:10.1701/3182.31610
- Gianfredi V, Santangelo OE, Provenzano S. Correlation between flu and Wikipedia's pages visualization. Acta Biomed. 2021;92(1):e2021056. doi:10.23750/abm.v92i1.9790
- 14. 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(3):E586-E591. doi:10.15167/2421-4248 /jpmh2021.62.3.1704
- 15. Santangelo OE, Provenzano S, Grigis D, Giordano D, Armetta F, Firenze A. Can Google Trends and Wikipedia help traditional surveillance? A pilot study on Measles. Acta Biomed. 2020; 91 (4):e2020190. doi: 10.23750/abm .v91i4.8888
- 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(4):385-391. doi:10.7416 /ai.2019.2300
- Riccò M, Baldassarre A, Provenzano S, et al. Infodemiology of RSV in Italy (2017-2022): An Alternative Option for the Surveillance of Incident Cases in Pediatric Age?. Children (Basel). 2022;9(12):1984. doi:10.3390/children9121984
- Google Trends. Available online: https://trends.google.it /trends/?geo=IT (last access January 9, 2022)
- Wikipedia. Available from: https://tools.wmflabs.org /pageviews (last accessed January 9, 2022)
- Mukaka MM. Statistics corner: A guide to appropriate use of correlation coefficient in medical research. Malawi Med J. 2012;24(3):69-71.
- White KJ. The Durbin-Watson Test for Autocorrelation in Nonlinear Models 370 The Review of Economics and Statistics the Durbin-Watson Test for Autocorrelation in Nonlinear Models. Rev Econ Stat. 1992; 74:370–373.
- StataCorp. Stata Statistical Software. In: Station C, editor.: StataCorp LP; 2015.
- 23. Niu Q, Liu J, Zhao Z, et al. Explanation of hand, foot, and mouth disease cases in Japan using Google Trends before and during the COVID-19: infodemiology study. BMC Infect Dis. 2022;22(1):806. doi:10.1186/s12879-022-07790-9
- 24. Kow RY, Mohamad Rafiai N, Ahmad Alwi AA, et al. COVID-19 Infodemiology: Association Between Google Search and Vaccination in Malaysian Population. Cureus. 2022;14(9):e29515. doi:10.7759/cureus.29515

- 25. Springer S, Zieger M, Strzelecki A. The rise of infodemiology and infoveillance during COVID-19 crisis. One Health. 2021;13:100288. doi:10.1016/j.onehlt.2021.100288
- 26. Nucci D, Santangelo OE, Nardi M, Provenzano S, Gianfredi V. Wikipedia, Google Trends and Diet: Assessment of Temporal Trends in the Internet Users' Searches in Italy before and during COVID-19 Pandemic. Nutrients. 2021;13(11):3683. doi:10.3390/nu13113683
- 27. Cai O, Sousa-Pinto B. United States Influenza Search Patterns Since the Emergence of COVID-19: Infodemiology Study. JMIR Public Health Surveill. 2022;8(3):e32364. doi:10.2196/32364
- 28. Santangelo OE, Gentile V, Pizzo S, Giordano D, Cedrone F. Machine Learning and Prediction of Infectious Diseases: A Systematic Review. Machine Learning and Knowledge Extraction. 2023; 5(1):175-198. doi:10.3390/make5010013
- 29. 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(4):e2021318. doi:10.23750/abm.v92i4.11558
- 30. Gianfredi V, Bragazzi NL, Mahamid M, et al. Monitoring public interest toward pertussis outbreaks: an extensive Google Trends-based analysis. Public Health. 2018;165:9-15. doi:10.1016/j.puhe.2018.09.001
- Riccò M, Valente M, Marchesi F. Are symptoms associated with SARS-CoV-2 infections evolving over time?. Infect Dis Now. 2022;52(2):110-112. doi:10.1016/j.idnow .2022.01.006
- 32. Salathé M. Digital epidemiology: what is it, and where is it going?. Life Sci Soc Policy. 2018;14(1):1. doi:10.1186 /s40504-017-0065-7
- Malik S. A Comparative Study of two major Search Engines: Google and Yahoo . Orient J Comp Sci and Technol. 7(1).

Correspondence:

Received: 17 January 2023

- Accepted: 17 February 2023
- Vincenza Gianfredi, MD, PhD

Department of Biomedical Sciences for Health, University of Milan

- Via Pascal, 36, 20133 Milan, Italy
- E-mail: vincenza.gianfredi@unimi.it