

Google Trends and Wikipedia Queries as Proxies for Measles Surveillance

OMAR ENZO SANTANGELO¹, ANNA SOLE PIZZAMIGLIO¹, CARLOTTA VELLA¹, DANILO GRECO²,
ALBERTO FIRENZE³, SANDRO PROVENZANO²

¹ Territorial Health and Social Care Authority of Lodi, ASST Lodi, Lodi, Italy; ² Local Health Unit of Trapani, ASP Trapani, Trapani, Italy; ³ Local Health Unit of Palermo, ASP Palermo, Palermo, Italy

Keywords

Infodemiology • Infoveillance • Measles • Infectious Diseases • Italy

Summary

Introduction. The increasing digitalization of information behaviors has created new opportunities in public health, promoting the development of infodemiology and infoveillance as complements to traditional epidemiological surveillance. This study aimed to assess the correlation and temporal association between official measles surveillance data published by the Italian National Institute of Health (ISS), Google Trends (GT) search volumes, and Wikipedia page views related to measles in Italy.

Methods. A retrospective ecological study with a time-series design was conducted using measles cases reported in Italy from January 2013 to November 2025. Monthly data from GT and Wikipedia were aggregated and compared with the monthly incidence reported in ISS bulletins. Linear regressions were performed to evaluate associations.

Results. The results showed a statistically significant associations between reported measles cases and both GT searches and Wikipedia page views. Linear regression analyses confirmed significant associations between digital indicators and official surveillance data, suggesting that online user activity closely reflects measles epidemiological trends.

Conclusion. These findings indicate that GT and Wikipedia may serve as useful and timely tools to complement traditional measles surveillance. Although caution is needed due to methodological limitations and potential information bias, integrating digital data could enhance early outbreak detection and support more rapid public health responses.

Introduction

In recent years, public healthcare has undergone a profound transformation driven by the progressive digitalization of the population's information behaviour, with increasing use of online platforms for health-related information searches. This is the context for infodemiology and infoveillance, two innovative approaches that apply the principles of epidemiology to the study of the production, dissemination, and use of health information in digital environments. Infodemiology, introduced by Gunther Eysenbach in the early 2000s, is defined as the science that analyses the distribution and determinants of information in digital contexts with the aim of supporting policy making and decision-making processes in public health [1, 2]. It is based on the assumption that online information flows significantly influence individual and collective behaviour, with potential direct impacts on population health outcomes. Within this theoretical framework, infoveillance represents the most operational dimension of infodemiology, configuring itself as a continuous and real-time information surveillance system, based on the analysis of online searches, social media and digital contents [2]. Numerous studies have demonstrated how these data sources can anticipate the progress of health events compared to traditional surveillance systems,

offering timely indicators relating to infectious diseases, health behaviours and risk perception in the general population [3-5]. The COVID-19 pandemic has further highlighted the crucial role of these tools in monitoring and combating the “infodemic”, *i.e.* the overabundance of information, often inaccurate or misleading, which can compromise the effectiveness of institutional healthcare responses [6]. At the same time, the evolution of digital platforms and the availability of large volumes of user-generated data from search engines like Google and social networks have expanded the application possibilities of digital epidemiological surveillance. Tools like Google Trends allow us to analyse the relative search volume (RSV) of specific terms or topics, providing indirect proxies for the population's attention, interest, and concerns [7, 8]. Several studies have highlighted a statistically significant correlation between online search trends and data from official surveillance systems for numerous infectious diseases [9-11], especially in contexts characterised by delays or incompleteness of traditional reporting systems. From this perspective, measles represents a particularly relevant case study. Despite the availability of a safe and effective vaccine, measles remains one of the most contagious infectious diseases and a significant cause of childhood morbidity and mortality globally, especially when vaccination coverage is insufficient [10, 12]. In Europe and Italy,

where the disease is subject to mandatory notification, epidemic outbreaks persist, particularly associated with poor adherence to vaccinations [12-14]. Traditional measles surveillance, based on clinical and laboratory reports, is often burdensome and characterized by a time delay in data availability; for this reason, the integration with digital data from the web and social media appears to be a promising strategy to improve early detection and monitoring of the disease progression [9-11, 15, 16]. Overall, infodemiology and infoveillance emerge as key tools for digital public health, capable of complementing traditional epidemiological surveillance systems, improving institutional communication, and strengthening the capacity to respond to health emergencies. However, their use requires careful consideration of methodological limitations, algorithmic biases, data representativeness, and ethical implications related to privacy protection and the responsible use of information [17]. In this scenario, harmonizing analysis methods and multidisciplinary integration are fundamental steps to fully exploit the potential of these innovative public health tools. This study was conducted with the primary objective of evaluating the temporal association between data from Google Trends (GT), Wikipedia searches, and conventional surveillance data relating to measles infection, reported in the bulletins of the Italian National Institute of Health (ISS). Specifically, the study aims to explore the applicability of digital surveillance tools as a support to traditional systems for monitoring communicable diseases, with specific reference to measles.

Materials and Methods

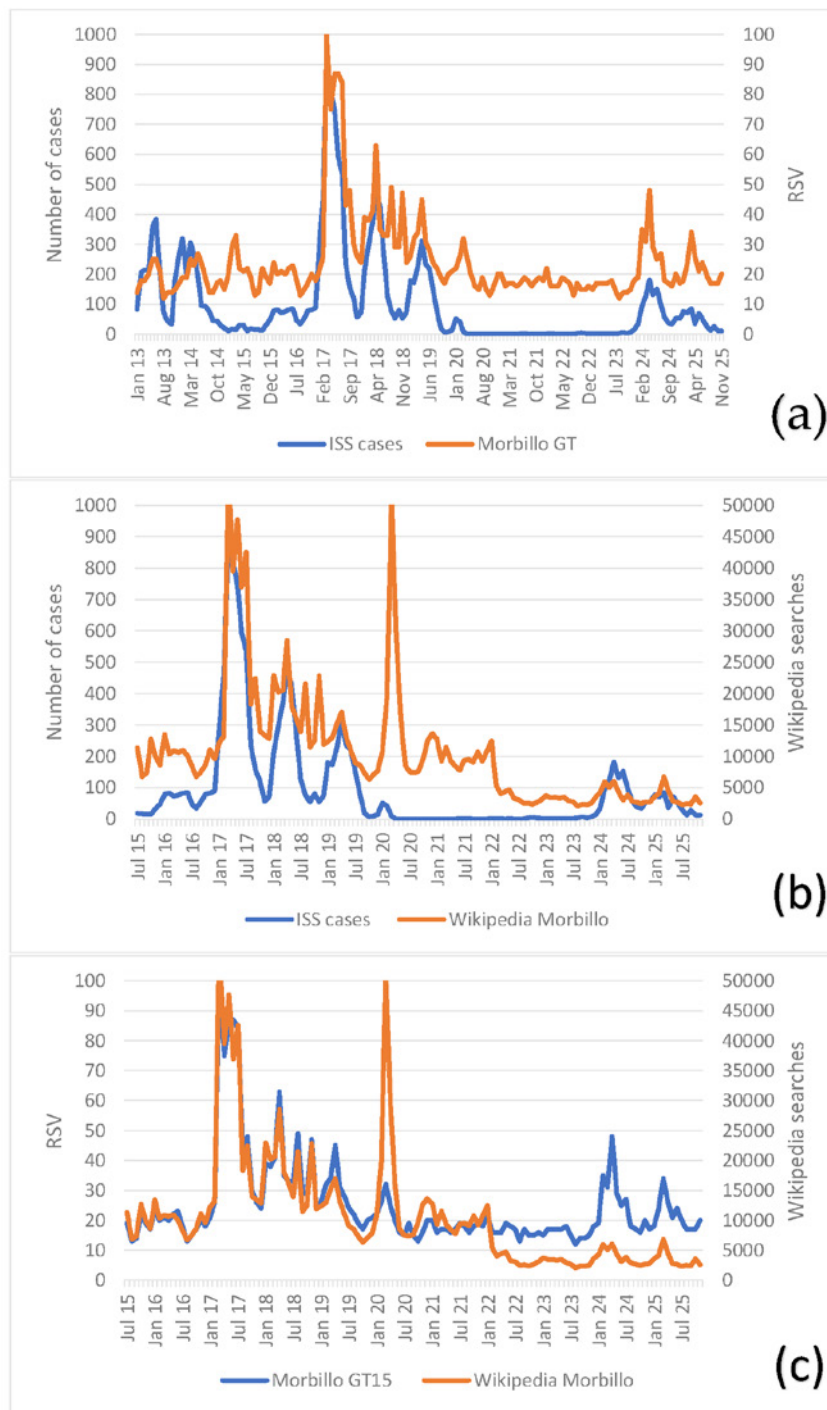
A retrospective ecological study with a time-series design was conducted. Measles cases were selected between January 2013 and November 2025. The ISS publishes a monthly bulletin containing the number of measles cases classified as possible, probable, and confirmed during the previous months [18]. Internet search data were obtained from Google Trends (GT) [11]. Google Trends provides a relative search volume (RSV), calculated as the proportion of searches for a specific term relative to the total number of searches performed in a given location and time period, normalized on a scale from 0 to 100. Data were extracted on December 31, 2025 using the Italian search terms in the “Health” category: “Morbillo” (Measles in English) and “Sintomi Morbillo” (Measles Symptoms in English), for the period between January 1, 2013 (when data became available) and November 30, 2025. Data were aggregated on a monthly basis. The resulting dataset, referred to as GT, included all available monthly RSV values during the study period. Since Wikipedia Trends data were available only from July 1, 2015, a second Google Trends dataset was created by selecting only RSV values from July 2015 to November 2025, corresponding to the period covered by Wikipedia data. This restricted dataset is referred to as GT15 throughout the manuscript and was used to

ensure temporal comparability between Google Trends and Wikipedia indicators. The files were downloaded in Comma-Separated Values (CSV) format. Wikipedia Trends [19] provides information on the number of views of specific Wikipedia pages over time. Data were extracted as daily values and subsequently aggregated on a monthly basis. The following Wikipedia pages were considered: “Morbillo” (Measles in English), “Vaccino del Morbillo” (Measles vaccine in English), “Vaccinazione MPR” (MMR Vaccination in English), “Esantema” (Exanthema in English), and “Macchie di Koplik” (Koplik spots in English), for the period between July 1, 2015 (when data became available) and November 30, 2025. Data from Wikipedia and Google Trends were temporally aligned with the monthly incidence of measles cases reported in the ISS epidemiological bulletins. Linear regression analyses were performed to evaluate the temporal associations between digital indicators and official measles surveillance data. Dependent variables included the monthly number of measles cases reported by the Italian National Institute of Health (ISS), Google Trends relative search volume (RSV) values for the terms “Morbillo” and “Sintomi Morbillo”, and monthly views of the Wikipedia page “Morbillo”. Independent variables included the selected Google Trends search terms, the analyzed Wikipedia pages, and, where applicable, ISS measles cases. Regression coefficients (β), 95% confidence intervals (95% CI), and corresponding *p-values* were calculated for each model. Statistical significance of individual regression coefficients was assessed using Student’s *t*-test. Potential autocorrelation was evaluated using the Durbin-Watson (DW) statistic. The DW test is used to detect autocorrelation in the residuals (prediction errors) of regression models [20, 21]. The DW statistic ranges between 0 and 4; values substantially lower than 2 indicate positive serial correlation, whereas values close to 2 suggest the absence of autocorrelation. Scatterplots with fitted linear regression lines and corresponding 95% confidence intervals were generated to visually represent the relationships between measles surveillance data and digital indicators. The coefficient of determination (R^2) was calculated to assess the proportion of variance explained by each linear model. The statistical significance level was set at 0.05. All analyses were conducted using STATA statistical software, version 13.

Results

The raw data on ISS measles cases, Google Trends relative search volume (RSV) for the term “Morbillo”, and views of the “Morbillo” Wikipedia page are presented in Figure 1. Figure 2 shows the scatterplots and the corresponding linear and exponential regression models describing the relationships between ISS cases, Google Trends search terms, and Wikipedia page views. In most analyses, linear models provided a better fit to the observed data compared with exponential models. Table I reports the results of the linear regression analyses,

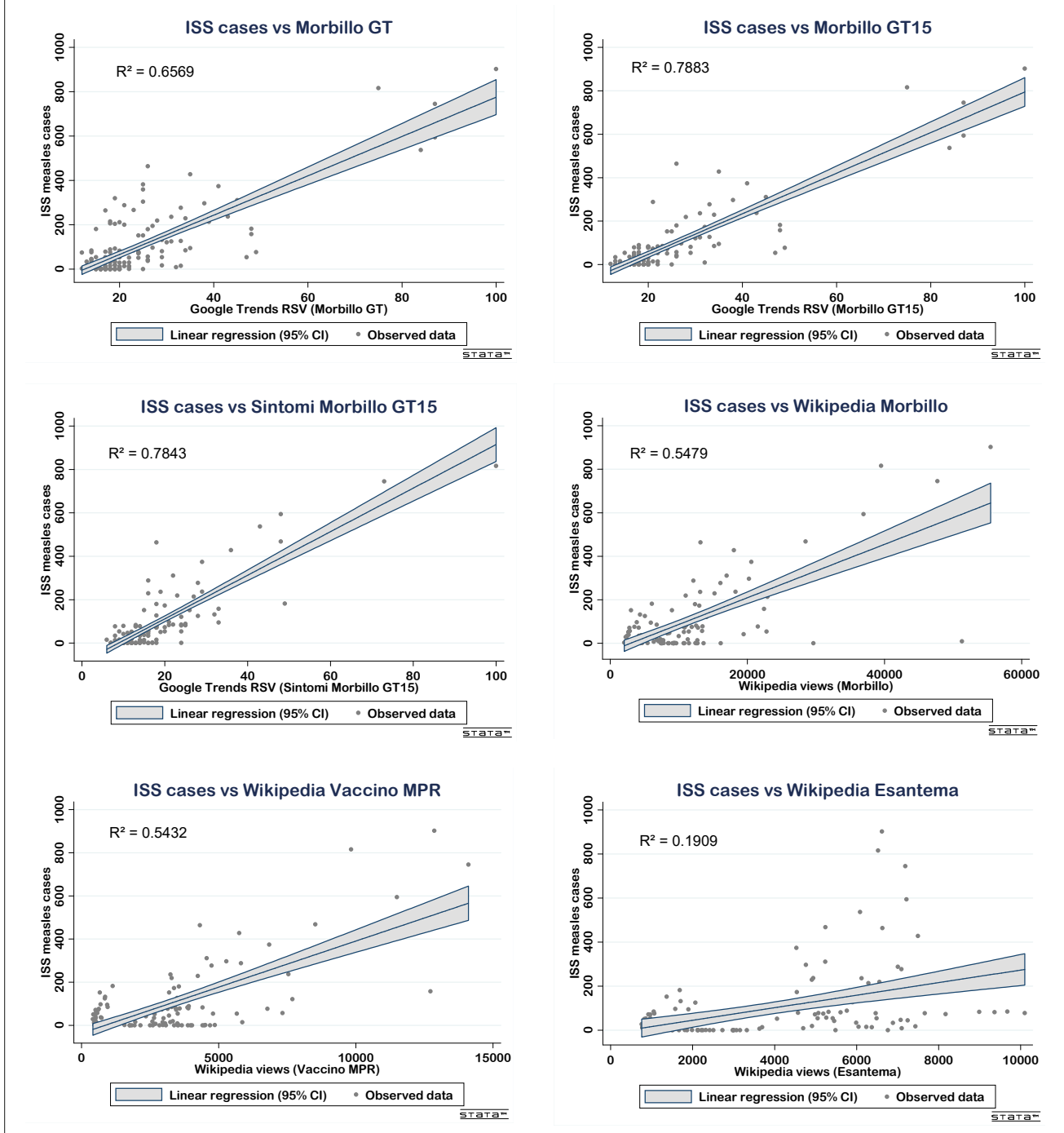
Fig. 1. Raw data for the number of ISS and RSV measles cases from GT (a). Raw data for the number of ISS cases and the number of searches for the measles Wikipedia page (b). Raw RSV measles data from GT15 and the number of searches for the measles Wikipedia page (c).



considering ISS measles cases, Google Trends (GT and GT15) values for “Morbillo”, and views of the “Morbillo” Wikipedia page as dependent variables. All models showed statistically significant positive associations between digital indicators and measles surveillance data, with all *p-values* <0.001. When ISS measles cases were considered as the dependent variable, Google Trends indicators showed a strong positive association with surveillance data. The search term “Morbillo GT” was

associated with ISS cases ($\beta=8.86$, 95% CI: 7.84-9.98), while the restricted dataset “Morbillo GT15” showed a slightly stronger association ($\beta=9.34$, 95% CI: 8.48-10.21). Similarly, searches for “Síntomi Morbillo” were positively associated with ISS cases, both for the complete dataset ($\beta=9.72$, 95% CI: 8.66-10.79) and for GT15 ($\beta=10.03$, 95% CI: 9.09-10.98). Wikipedia indicators also showed significant positive associations with ISS measles cases. The strongest association was observed for

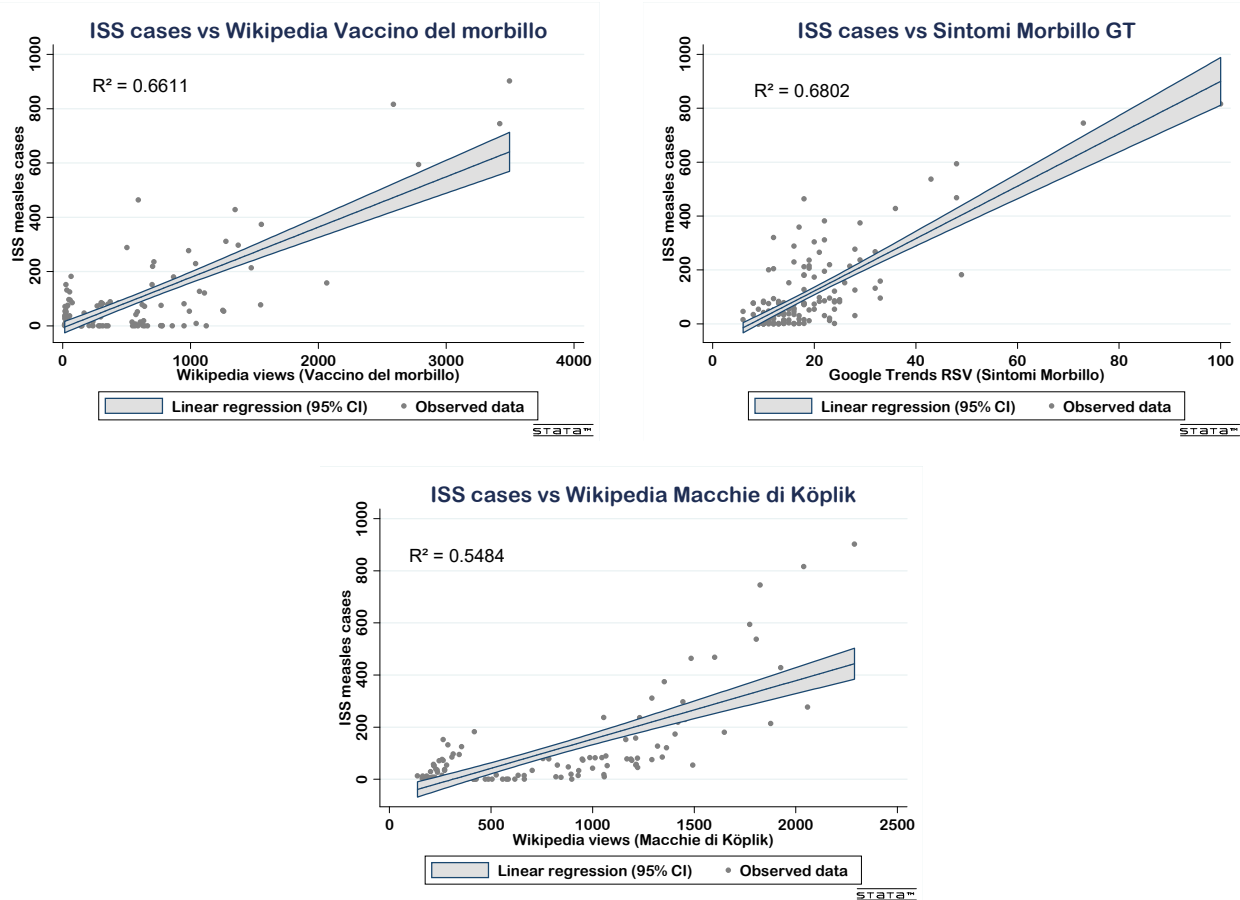
Fig. 2. Scatterplots showing the associations between official measles surveillance data (ISS cases) and digital indicators. Linear regression lines with corresponding 95% confidence intervals are shown. The corresponding R^2 are shown in the figures



“Macchie di Koplik Wikipedia” ($\beta=0.22$, 95% CI: 0.18-0.26), followed by “Vaccino del Morbillo Wikipedia” ($\beta=0.18$, 95% CI: 0.16-0.21), “Vaccino MPR Wikipedia” ($\beta=0.04$, 95% CI: 0.03-0.05), “Esantema Wikipedia” ($\beta=0.02$, 95% CI: 0.01-0.03), and “Morbillo Wikipedia” ($\beta=0.012$, 95% CI: 0.010-0.014). To identify the digital indicators most strongly associated with official measles surveillance data, Internet-based variables were ranked according to the magnitude of the regression coefficients

(β) obtained from models using ISS cases as the dependent variable. Among Google Trends variables, the highest coefficients were observed for “Sintomi Morbillo GT15”, followed by “Sintomi Morbillo GT”, “Morbillo GT15”, and “Morbillo GT”. Among Wikipedia variables, the ranking was led by “Macchie di Koplik Wikipedia”, followed by “Vaccino del Morbillo Wikipedia”, “Vaccino MPR Wikipedia”, “Esantema Wikipedia”, and “Morbillo Wikipedia” (Tab. II). Further analyses showed that ISS

Fig. 2 (follows). Scatterplots showing the associations between official measles surveillance data (ISS cases) and digital indicators. Linear regression lines with corresponding 95% confidence intervals are shown. The corresponding R^2 are shown in the figures



measles cases were significantly associated with Google Trends activity for “Morbillo” ($\beta=0.07$, 95% CI: 0.06-0.08) and “Morbillo GT15” ($\beta=0.08$, 95% CI: 0.07-0.09). In addition, “Sintomi Morbillo GT” was strongly associated with “Morbillo GT” ($\beta=0.92$, 95% CI: 0.83-1.01), while “Sintomi Morbillo GT15” was significantly associated with “Morbillo GT15” ($p < 0.001$), confirming the consistency between general measles-related searches and symptom-related searches. Positive associations were also observed between Wikipedia indicators and “Morbillo GT15”. These included “Morbillo Wikipedia” ($\beta=0.001$, 95% CI: 0.0011-0.0015), “Vaccino del Morbillo Wiki” ($\beta=0.019$, 95% CI: 0.017-0.021), “Vaccino MPR Wiki” ($\beta=0.004$, 95% CI: 0.003-0.004), “Esantema Wiki” ($\beta=0.002$, 95% CI: 0.001-0.003), and “Macchie di Koplik Wiki” ($\beta=0.019$, 95% CI: 0.015-0.023). ISS cases were also associated with “Morbillo Wikipedia” activity ($\beta=44.62$, 95% CI: 37.39-51.86). Similarly, “Morbillo GT15” ($\beta=527.67$, 95% CI: 464.67-590.67) and “Sintomi del Morbillo GT15” ($\beta=453.25$, 95% CI: 362.01-544.49) significantly predicted Wikipedia searches for “Morbillo”. Overall, these findings demonstrate consistent positive relationships between online search activity, Wikipedia page views, and official measles surveillance data. Durbin-Watson statistics were calculated for all linear regression

models to assess residual autocorrelation. The obtained values ranged from 0.21 to 1.22, indicating the presence of positive autocorrelation in the residuals across most models. These findings suggest that the observations were not fully independent over time, which is consistent with the time-series nature of the analyzed data. Therefore, the observed associations should be interpreted with caution, as residual autocorrelation may have influenced the estimated regression coefficients and their statistical significance.

Discussion

The study of the application of GT and Wikipedia in the surveillance of communicable diseases in Italy has a high relevance for public health, as suggested by numerous previous studies [7, 8]. Measles outbreaks remain a global health problem, and surveillance is an essential tool for their control. However, traditional monitoring approaches based on conventional reporting systems present some critical issues, including time lags in data availability and potential underestimation of cases due to underreporting.

In 2009, Eysenbach defined infodemiology and

Tab. I. Linear regression analyses showing regression coefficients (B), 95% confidence intervals (C.I. 95%), *p*-values for individual regression coefficients (Student's *t*-test), and Durbin-Watson statistics.

Independent variable: Morbillo GT				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin Watson
ISS cases	8.86	7.84-9.98	<0.001	0.56
Independent variable: Morbillo GT15				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin Watson
ISS cases	9.34	8.48-10.21	<0.001	0.93
Independent variable: Sintomi Morbillo GT				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
ISS cases	9.72	8.66-10.79	<0.001	0.67
Independent variable: Sintomi Morbillo GT15				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
ISS cases	10.03	9.09-10.98	<0.001	0.92
Independent variable: Morbillo Wikipedia				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
ISS cases	0.012	0.010-0.014	<0.001	0.43
Independent variable: Vaccino del Morbillo Wikipedia				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
ISS cases	0.18	0.16-0.21	<0.001	0.44
Independent variable: Vaccino MPR Wikipedia				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
ISS cases	0.04	0.03-0.05	<0.001	0.37
Independent variable: Esantema Wikipedia				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
ISS cases	0.02	0.01-0.03	<0.001	0.21
Independent variable: Macchie di Koplik Wikipedia				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
ISS cases	0.22	0.18-0.26	<0.001	0.25
Independent variable: ISS cases				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
Morbillo GT	0.07	0.06-0.08	<0.001	0.82
Independent variable: Sintomi Morbillo GT				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
Morbillo GT	0.92	0.83-1.01	<0.001	1.06
Independent variable: ISS cases				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
Morbillo GT15	0.08	0.07-0.09	<0.001	1.22
Independent variable: Sintomi Morbillo GT15				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
Morbillo GT15	0.94	0.84-1.03	<0.001	1.05
Independent variable: Morbillo Wikipedia				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
Morbillo GT15	0.001	0.0011-0.0015	<0.001	0.51
Independent variable: Vaccino del Morbillo Wiki				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
Morbillo GT15	0.019	0.017-0.021	<0.001	0.52
Independent variable: Vaccino MPR Wiki				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson

Tab. I (follows). Linear regression analyses showing regression coefficients (B), 95% confidence intervals (C.I. 95%), *p*-values for individual regression coefficients (Student's *t*-test), and Durbin-Watson statistics.

Independent variable: Morbillo GT				
Morbillo GT15	0.004	0.003-0.004	<0.001	0.55
Independent variable: Esantema Wiki				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
Morbillo GT15	0.002	0.001-0.003	<0.001	0.51
Independent variable: Macchie di Koplik Wiki				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
Morbillo GT15	0.019	0.015-0.023	<0.001	0.66
Independent variable: ISS cases				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
Morbillo Wikipedia	44.62	37.39-51.86	<0.001	0.73
Independent variable: Morbillo GT15				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
Morbillo Wikipedia	527.67	464.67-590.67	<0.001	0.52
Independent variable: Sintomi del Morbillo GT15				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
Morbillo Wikipedia	453.25	362.01-544.49	<0.001	0.61
Independent variable: Vaccino del Morbillo Wiki				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
Morbillo Wikipedia	12.38	11.31-13.45	<0.001	0.84
Independent variable: Vaccino MPR Wiki				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
Morbillo Wikipedia	2.92	2.58-3.26	<0.001	0.81
Independent variable: Esantema Wiki				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
Morbillo Wikipedia	2.34	1.78-2.91	<0.001	0.64
Independent variable: Macchie di Koplik Wiki				
Dependent variable	Coef.	C.I. 95%	<i>p</i> -value	Durbin-Watson
Morbillo Wikipedia	13.78	11.64-15.92	<0.001	0.83

infoveillance as “the science of the distribution and determinants of information in an electronic medium, especially the Internet, or in a population, with the ultimate goal of informing public health and public policy.” [1]. This emerging approach has been widely used to explore public interest in communicable diseases, such as influenza [3], and for the early detection of infectious epidemics [22].

Measles is one of the most contagious vaccine-preventable diseases ($R_0=12-18$) [12]. The inherent delays in traditional surveillance systems can limit the ability of health authorities to intervene promptly. Data from Google Trends and Wikipedia are collected and processed in near real time, allowing for significantly faster information availability than traditional systems. [9]. In this study, statistical analyses highlighted a clear temporal association between ISS surveillance data and search trends on Google and Wikipedia. In particular, associations were observed on a monthly basis, suggesting that this approach could

provide a useful time frame for more timely public health interventions. However, it cannot be ruled out that a weekly or daily analysis could allow for even earlier case reporting, as reported in other studies. [23, 24].

An additional methodological aspect to consider is the presence of positive residual autocorrelation, as indicated by the Durbin-Watson statistics. This finding is consistent with the temporal nature of the analyzed data and suggests that consecutive monthly observations were not fully independent. Consequently, although the observed associations between digital indicators and official surveillance data were statistically significant, the regression estimates should be interpreted with appropriate caution.

This study has some limitations. First, the influence of mass media (television, radio, and print) may lead to an increase in online searches regardless of actual case trends [25]. Furthermore, Google Trends does not provide sufficiently granular data to accurately identify the geographic location of outbreaks. Further limitations

Tab. II. Ranking of Internet-based indicators according to the strength of their association with ISS measles cases.

Data source	Internet variable (independent variable)	Ranking
Google Trends	Sintomi Morbillo GT15	1
Google Trends	Sintomi Morbillo GT	2
Google Trends	Morbillo GT15	3
Google Trends	Morbillo GT	4
Wikipedia	Macchie di Koplik Wikipedia	1
Wikipedia	Vaccino del Morbillo Wikipedia	2
Wikipedia	Vaccino MPR Wikipedia	3
Wikipedia	Esantema Wikipedia	4
Wikipedia	Morbillo Wikipedia	5

GT: Google Trends general dataset

GT15: Google Trends dataset restricted to the period July 2015–November 2025

arise from changes over time in the GT interface and algorithms, which are not fully documented and could influence the results. Therefore, the interpretation and generalization of the results should be made with caution. Another limitation is the presence of residual autocorrelation in the regression models, as indicated by the Durbin-Watson statistics. This reflects the time-series nature of the data and may have affected the independence of observations and the estimated standard errors. Therefore, although linear regression was considered appropriate for this exploratory analysis, future studies should consider applying dedicated time-series approaches, such as autoregressive or other longitudinal models, to better account for temporal dependence.

Previous experiences, such as Google Flu Trends, have shown how the use of big data for epidemiological forecasting can lead to significant overestimations, as occurred in the 2012–2013 influenza season. [26]. In the case of measles, online search volume is likely influenced more by media attention than by direct patient experience, given the relative rarity of cases. This makes digital surveillance potentially more sensitive, but also more susceptible to bias. A prudent and integrated use of these tools, alongside traditional surveillance systems, is therefore necessary.

Conclusion

In summary, the Italian healthcare system operates in a context characterized by limited resources, which necessitates the adoption of innovative strategies to improve efficiency and sustainability. Investing in technological innovation represents a strategic lever to strengthen the National Health Service and improve its capacity to respond to the population's needs. The results of this study suggest that surveillance systems based on Google Trends and Wikipedia could play a complementary role in public health by providing dynamic, near-real-time indicators of the spread of infectious diseases. The integration of these tools

with traditional surveillance systems could support policymakers in implementing preventive measures and targeted communication strategies [27, 28]. Although this field of research is still relatively young, its rapid expansion is expected, facilitated by the increasing availability of data and increased computational power [29]. Online searches, conducted in real time, could serve as an early warning system, signaling sudden changes in the population's interests and therefore deserving attention from health authorities [30]. Looking ahead, this approach could be used as a supplementary tool for epidemic forecasting and for a more efficient allocation of health resources [31, 32].

Acknowledgments

None.

Conflicts of Interest statement

The authors declare no conflicts of interest.

Authors' contributions

OES: Conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, visualization, supervision, project administration.

OES, ASP, CV, DG, SP, AF: Original draft preparation, writing, review and editing.

All authors have read and agreed to the published version of the manuscript.

References

- [1] Eysenbach G. Infodemiology: The epidemiology of (mis)information. *Am J Med* 2002;113:763–5. [https://doi.org/10.1016/s0002-9343\(02\)01473-0](https://doi.org/10.1016/s0002-9343(02)01473-0).
- [2] Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *J Med Internet Res* 2009;11:e11. <https://doi.org/10.2196/jmir.1157>.
- [3] Mavragani A. Infodemiology and Infoveillance: Scoping Review. *J Med Internet Res* 2020;22:e16206. <https://doi.org/10.2196/16206>.
- [4] Strzelecki A, Meinzenbach A, Zieger M. Infodemic and infodemiology in public health: Similarities and differences. *Healthcare Analytics*.2023;4:100243. <https://doi.org/10.1016/j.health.2023.100243>.
- [5] Zeraatkar K, Ahmadi M. Trends of infodemiology studies: a scoping review. *Health Info Libr J* 2018;35:91–120. <https://doi.org/10.1111/hir.12216>.
- [6] World Health Organization. 2020, 1st WHO Infodemiology Conference – Call for Papers and Recommendations. Available at: <https://www.who.int/initiatives/epi-win/infodemic-management/1st-who-infodemiology-conference>. Accessed on: 01/04/2026.
- [7] Milinovich GJ, Williams GM, Clements AC, Hu W. Internet-based surveillance systems for monitoring emerging infec-

- tious diseases. *Lancet Infect Dis* 2014;14:160-8. [https://doi.org/10.1016/S1473-3099\(13\)70244-5](https://doi.org/10.1016/S1473-3099(13)70244-5).
- [8] Pollett S, Boscardin WJ, Azziz-Baumgartner E, Tinoco YO, Soto G, Romero C, Kok J, Biggerstaff M, Viboud C, Rutherford GW. Evaluating Google Flu Trends in Latin America: Important Lessons for the Next Phase of Digital Disease Detection. *Clin Infect Dis* 2017;64:34-41. <https://doi.org/10.1093/cid/ciw657>.
 - [9] Teng Y, Bi D, Xie G, Jin Y, Huang Y, Lin B, An X, Feng D, Tong Y. Dynamic Forecasting of Zika Epidemics Using Google Trends. *PLoS One*. 2017;12:e0165085. <https://doi.org/10.1371/journal.pone.0165085>.
 - [10] Giordano D, Provenzano S, Santangelo OE, Piazza D, Ferraro D, Colomba C, Corsello G, La Placa S, Firenze A. Active immunization status against measles, mumps, rubella, hepatitis B in internationally adopted children, surveyed at the university hospital of Palermo, Sicily. *Ann Ig* 2018;30:431-5. <https://doi.org/10.7416/ai.2018.2243>.
 - [11] Google Trends. Available at: <https://trends.google.com/trends/>. Accessed on: 31/12/2025.
 - [12] Moss WJ. Measles. *Lancet* 2017;390:2490-502. [https://doi.org/10.1016/S0140-6736\(17\)31463-0](https://doi.org/10.1016/S0140-6736(17)31463-0).
 - [13] Coughlan S, Connell J, Cohen B, Jin L, Hall WW. Suboptimal measles-mumps-rubella vaccination coverage facilitates an imported measles outbreak in Ireland. *Clin Infect Dis* 2002;35:84-6. <https://doi.org/10.1086/340708>.
 - [14] Probert WS, Glenn-Finer R, Espinosa A, Yen C, Stockman L, Harriman K, Hacker JK. Molecular Epidemiology of Measles in California, United States-2019. *J Infect Dis* 2021;224:1015-23. <https://doi.org/10.1093/infdis/jiab059>.
 - [15] Hilbert M, López P. The world's technological capacity to store, communicate, and compute information. *Science* 2011;332:60-5. <https://doi.org/10.1126/science.1200970>.
 - [16] Alicino C, Bragazzi NL, Faccio V, Amicizia D, Panatto D, Gasparini R, Icardi G, Orsi A. Assessing Ebola-related web search behaviour: insights and implications from an analytical study of Google Trends-based query volumes. *Infect Dis Poverty* 2015;4:54. <https://doi.org/10.1186/s40249-015-0090-9>.
 - [17] Velasco E, Agheneza T, Denecke K, Kirchner G, Eckmanns T. Social media and Internet-based data in global systems for public health surveillance: a systematic review. *Milbank Q* 2014;92:7-33.
 - [18] Istituto Superiore di Sanità. Available at: <https://www.epicentro.iss.it/morbillo/bollettino>. Accessed on: 10/01/2026.
 - [19] Wikipedia. Available at: <https://it.wikipedia.org/wiki/Wikipedia>. Accessed on: 31/12/2025.
 - [20] Chen Y. Spatial Autocorrelation Approaches to Testing Residuals from Least Squares Regression. *PLoS One* 2016;11:e0146865. <https://doi.org/10.1371/journal.pone.0146865>.
 - [21] 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;74:370-3.
 - [22] 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>.
 - [23] 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>.
 - [24] Provenzano S, Santangelo OE, Giordano D, Alagna E, Piazza D, Genovese D, Calamusa G, Firenze A. Predicting disease outbreaks: evaluating measles infection with Wikipedia Trends. *Recenti Prog Med* 2019;110:292-6. <https://doi.org/10.1701/3182.31610>.
 - [25] Odone A, Signorelli C. When vaccine hesitancy makes headlines. *Vaccine* 2017;35:1209-10. <https://doi.org/10.1016/j.vaccine.2015.11.051>.
 - [26] Lazer D, Kennedy R, King G, Vespignani A. Big data. The parable of Google Flu: traps in big data analysis. *Science*. 2014;343(6176):1203-5. <https://doi.org/10.1126/science.1248506>.
 - [27] Niu Q, Liu J, Zhao Z, Onishi M, Kawaguchi A, Bandara A, Harada K, Aoyama T, Nagai-Tanima M. 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;22:806. <https://doi.org/10.1186/s12879-022-07790-9>.
 - [28] Provenzano S, Santangelo OE, Gianfredi V. Infodemiology and infoveillance: framework for contagious exanthematous diseases, of childhood in Italy. *Pathog Glob Health* 2024;118:317-24. <https://doi.org/10.1080/20477724.2024.2323844>.
 - [29] Santangelo OE, Gentile V, Pizzo S, Giordano D, Cedrone F. Machine Learning and Prediction of Infectious Diseases: A Systematic Review. *Mach Learn Knowl Extr* 2023;5:175-98. <https://doi.org/10.3390/make5010013>.
 - [30] Lan R, Campana F, Tardivo D, Catherine JH, Vergnes JN, Hadj-Saïd M. Relationship between internet research data of oral neoplasms and public health programs in the European Union. *BMC Oral Health* 2021;21:648. <https://doi.org/10.1186/s12903-021-02022-z>.
 - [31] Steinert Y, Naismith L, Mann K. Faculty development initiatives designed to promote leadership in medical education. A BEME systematic review: BEME Guide No. 19. *Med Teach* 2012;34:483-503. <https://doi.org/10.3109/0142159X.2012.680937>.
 - [32] Santangelo OE, Gianfredi V, Provenzano S. Wikipedia searches and the epidemiology of infectious diseases: A systematic review. *Data & Knowledge Engineering* 2022;142:102093. <https://doi.org/10.1016/j.datak.2022.102093>.

Received on May 15, 2026. Accepted on August 17, 2026.

Correspondence: Santangelo Omar Enzo, Address: Territorial Health and Social Care Authority of Lodi, ASST Lodi, 26900 Lodi, Italy. E-mail: omarenzosantangelo@hotmail.it

How to cite this article: Santangelo OE, Pizzamiglio AS, Vella C, Greco D, Firenze A, Provenzano S. Google Trends and Wikipedia Queries as Proxies for Measles Surveillance. *J Prev Med Hyg* 2026;67:E326-E334. <https://doi.org/10.82082/2421-4248/jpmh2026.67.3.3956>

© 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>