

Original Paper

# Google Trends as a Predictive Tool for COVID-19 Vaccinations in Italy: Retrospective Infodemiological Analysis

---

Alessandro Rovetta

R&C Research, Bovezzo, Italy

**Corresponding Author:**

Alessandro Rovetta

R&C Research

Via Brede Traversa II

Bovezzo, 25073

Italy

Phone: 39 3927112808

Email: [rovetta.mresearch@gmail.com](mailto:rovetta.mresearch@gmail.com)

**Related Articles:**

Preprint: <https://preprints.jmir.org/preprint/35356>

Peer-Review Report by Artur Strzelecki (Reviewer M): <https://med.jmirx.org/2022/2/e38665/>

Peer-Review Report by Zubair Shah (Reviewer O): <https://med.jmirx.org/2022/2/e38724/>

Peer-Review Report by Angela Chang (Reviewer BL): <https://med.jmirx.org/2022/2/e38726/>

Authors' Response to Peer-Review Reports: <https://med.jmirx.org/2022/2/e38695/>

## Abstract

**Background:** Google Trends is an infoveillance tool widely used by the scientific community to investigate different user behaviors related to COVID-19. However, several limitations regarding its adoption are reported in the literature.

**Objective:** This paper aims to provide an effective and efficient approach to investigating vaccine adherence against COVID-19 via Google Trends.

**Methods:** Through the cross-correlational analysis of well-targeted hypotheses, we investigate the predictive capacity of web searches related to COVID-19 toward vaccinations in Italy from November 2020 to November 2021. The keyword “vaccine reservation” query (VRQ) was chosen as it reflects a real intention of being vaccinated (V). Furthermore, the impact of the second most read Italian newspaper (vaccine-related headlines [VRH]) on vaccine-related web searches was investigated to evaluate the role of the mass media as a confounding factor. Fisher r-to-z transformation ( $z$ ) and percentage difference ( $\delta$ ) were used to compare Spearman coefficients. A regression model  $V=f(VRH, VRQ)$  was built to validate the results found. The Holm-Bonferroni correction was adopted ( $P^*$ ). SEs are reported.

**Results:** Simple and generic keywords are more likely to identify the actual web interest in COVID-19 vaccines than specific and elaborated keywords. Cross-correlations between VRQ and V were very strong and significant (min  $r^2=0.460$ ,  $P^*<.001$ , lag 0 weeks; max  $r^2=0.903$ ,  $P^*<.001$ , lag 6 weeks). The remaining cross-correlations have been markedly lower ( $\delta>55.8\%$ ;  $z>5.8$ ;  $P^*<.001$ ). The regression model confirmed the greater significance of VRQ versus VRH ( $P^*<.001$  vs  $P=.03$ ,  $P^*=.29$ ).

**Conclusions:** This research provides preliminary evidence in favor of using Google Trends as a surveillance and prediction tool for vaccine adherence against COVID-19 in Italy. Further research is needed to establish the appropriate use and limits of Google Trends for vaccination tracking. However, these findings prove that the search for suitable keywords is a fundamental step to reduce confounding factors. Additionally, targeting hypotheses helps diminish the likelihood of spurious correlations. It is recommended that Google Trends be leveraged as a complementary infoveillance tool by government agencies to monitor and predict vaccine adherence in this and future crises by following the methods proposed in this paper.

(*JMIRx Med* 2022;3(2):e35356) doi: [10.2196/35356](https://doi.org/10.2196/35356)

**KEYWORDS**

COVID-19; epidemiology; Google Trends; infodemiology; infoveillance; Italy; public health; SARS-CoV-2; vaccinations; vaccines; social media analysis; social media

## Introduction

Google Trends is an online website created by Google LLC that allows the user to examine the popularity of exact search queries (keywords) in Google Search across specific regions, time lapses, and languages. Google Trends has often been used by the scientific community to conduct infodemiological and epidemiological analyses [1,2]. In particular, this infoveillance approach—aimed at studying distribution and determinants of information in an electronic medium, specifically the internet, or in a population, with the ultimate aim to inform public health and public policy—has been applied to various disciplines, including but not limited to psychology, economics, veterinary medicine, and pharmacy [3-7]. However, past studies have been often criticized for not providing sufficient documentation to guarantee the full reproducibility of the methods [8]. Moreover, some authors have shown severe limitations in its use as a surveillance tool, including anomalies in results and mass media influence [9,10]. Nonetheless, Google Trends remains a currently irreplaceable tool for infoveillance. In particular, its simplicity and efficiency make analyses much faster than other systems, such as investigating user posts via application programming interface and machine learning [10]. In this regard, various strategies have been proposed in the literature to address its weaknesses [10-12]. Taking the latter into account, in this brief paper, Google Trends is used to investigate vaccine adherence in Italy against COVID-19. Indeed, COVID-19 vaccines are essential to contain the infection, limiting the spread of new variants of concern and substantially reducing the severity of the disease [13]. For instance, the latest report from the Italian Medicines Agency highlighted a low risk associated with vaccines despite high protection against COVID-19 [14]. Even considering Omicron's more elusive variant of concern, the rates of hospitalizations, patients in intensive care units, and deaths are 10, 27, and 25 times higher for the unvaccinated, respectively [15]. At present, monitoring of vaccine adherence is epidemiologically essential, especially considering the growing no-vax movement [16]. Furthermore, the use of effective and efficient infoveillance techniques is also necessary for any future health crises. Therefore, this research proposes an approach capable of targeting the hypotheses and eliminating the anomalies of Google Trends, thus reducing the likelihood of running into spurious correlations and having statistically uncertain outcomes. Specifically, the ability to predict the COVID-19 vaccination trend in Italy based on vaccine-related web queries is examined.

## Methods

### Procedure Summary

The hypothesis to be verified is that the COVID-19 “vaccine reservation” query (VRQ) can predict the trends of national and regional vaccinations (V). To achieve this scope and quantify the impact of mass media on web queries, cross-correlations between VQR, V, and COVID-19 vaccine-related headlines (VRH) of the Italian newspaper “La Repubblica” were searched. In particular, “La Repubblica” was chosen for its large readership and its online historical database (which allows the user to easily search for published articles containing a list of

specific keywords). Besides, an appropriate regression model  $V=f(VRH, VRQ)$  was also constructed.

### Data Collection

The keyword “prenotazione vaccino” (vaccine reservation) was selected since it clearly expresses the desire to administer the dose of a vaccine. Synonyms of the word “prenotazione” (reservation) have been searched on the Treccani.it online dictionary. However, the synonym queries had a much lower relative search volume (RSV). Besides, even adding them to the original keyword through the “+” operator, the trends remained highly similar. Since the combination of queries makes it more likely that anomalies will appear in the data sets, a single query was chosen. The goodness of VRQ in identifying the web interest in COVID-19 vaccine queries is reported in the Results section. The Google Trends parameters have been set as follows: region: Italy; period: November 1, 2020, to November 27, 2021; category: all categories; and search type: web search. The “period” parameter has been changed to “Past 5 years” when performing a historical time series analysis. The “region” parameter was changed from “Italy” to “[the name of the region concerned]” when analyzing regional trends. The “interest over time” data sets were downloaded in “.csv format.” Following the previous methods, the keywords “disdire vaccino + cancellare vaccino + evitare vaccino + non vaccinarsi + green pass falso + comprare green pass” (revoke vaccine + cancel vaccine + avoid vaccine + do not get vaccinated + fake green pass + buy green pass) were searched to investigate users' web interest in methods of not getting vaccinated. The first keyword searched was “disdire vaccino.” The other terms have been selected by consulting various possible synonyms in the Treccani.it online dictionary and Google Trends-related queries. The final exact queries searched on Google Trends are reported as references [17,18]. Regarding national vaccinations, the data set was downloaded from the “GitHub” platform [19]. The keyword “vaccino, vaccini, astrazeneca, pfizer, moderna, johnson&johnson, vaxzevria, comirnaty, pikevax” was searched in the historical archive of the newspaper “La Repubblica” [20]. In particular, this query includes the generic and proper names of the COVID-19 vaccines administered in Italy during the investigated period. The number of articles containing the aforementioned keyword was counted from week to week until it covered the period November 2020 to November 2021. The filter has been set to “ricerca avanzata” (advanced search) and “almeno una [parola]” (at least one [word]). This newspaper was chosen since it represents the second most widely read newspaper in Italy and provides the most detailed news database online. Furthermore, a previous publication showed similar news trends across primary Italian mass media during COVID-19 [21]. Such a result aligns with the theory of news competition and increasing returns-to-scale, which prompts profit-motivated media to publish on hot topics (as of interest to a broad audience) [22]. For these reasons, the author of this paper considered the source “La Repubblica” sufficient to represent the Italian media clamor about vaccines.

## Ethical Considerations

This study does not involve human participants or animals. All Google Trends data is anonymized. Therefore, the research does not require approval from a committee.

## Statistical Analysis

The shape of the data distribution was assessed both graphically and through the Shapiro-Wilk test. Since the data sets were not normal ( $P < .001$ ) and above or below threshold correlations were not of interest, we adopted the Spearman correlation ( $R$ ) [23]. To check the discrepancy between two time series, quantifiers such as percentage difference (used to compare the average RSV of two simultaneous series and indicated with “ $\delta$ ”) and percentage increase (used to compare the average RSV of two consecutive series and indicated with “ $\Delta$ ”) were exploited. The statistical significance of the discrepancies between average values was measured through the Welch  $t$  test ( $t$ ), which is also valid for large nonnormal data sets [24,25]. When two contiguous time series were compared, a graphic check was carried out to guarantee the absence of seasonality and trends. All data sets were normalized to 100 by multiplying individual values by the constant “100/data set maximum value.” The “Lag week” was defined as the number of weeks by which a time series was shifted to obtain the maximum correlation with another time series. By doing so, it was possible to estimate the predictive power of one time series over another and the latency between them. Finally, a multiple regression was used to build the function  $Y=f(\text{VRH}, \text{VRQ})$  to evaluate the impact of VRH and VRQ on  $V$  [26]. SEs for the regression coefficients are reported. Based on previous literature, any causal correlations between the media clamor and web searches should be sought within a maximum lag range of 3 weeks (from  $-3$  to  $3$ ) [9,11,21,27,28]. Indeed, the web interest in a topic must arise around the media hype peak to be considered a direct consequence or cause of the latter. Regarding the pairs (VRH,  $V$ ) and (VRQ,  $V$ ), the lag acceptability range was fixed at 0 to 8 weeks since it can take up to 2 months from vaccine booking to administration. Fisher  $r$ -to- $z$  transformation ( $z$ ) was used to compare Spearman coefficients. Since the search for cross-correlations is highly exploratory, the Holm-Bonferroni correction was adopted ( $m=50$  hypotheses). The original  $P$  values have been reported alongside the adjusted ones ( $P^*$ )—when  $P^* > .001$ —to allow the reader to interpret the data independently.

## Mass Media Clamor as a Confounding Factor

As previously discussed, there is solid evidence that mass media can substantially impact users' web interests. This fact increases the probability of spurious correlations due to a so-called

confounding factor, defined as a “hidden” variable (or set of variables) capable of distorting the true relationship between other apparently correlated (or uncorrelated) variables [29]. In this specific case, media hype can create highly confounding scenarios. For example, a COVID-19 outbreak can generate intense news fanfare, immediately followed by a user's growing web interest in the disease. After 7 days, an increase in COVID-19 cases is registered. Examining the sole couple (user interest, COVID-19 cases), it could seem like the online searches predicted the increase in infections. However, by introducing the “media hype” variable, it is observed that users' web interest is much more correlated with the latter than with COVID-19 cases [21]. For this reason, media coverage is introduced in this analysis as a possible confounding factor capable of distorting the relationship between  $V$  and VRQ. In this regard, it is fair to admit that other confounding factors not considered in this paper could alter such a relationship in complex ways. Nonetheless, at present, to the best of the author's knowledge, media influence is the only widely reported confounding factor in the literature regarding Google Trends. Furthermore, the main research hypothesis is well-targeted, thus reducing the likelihood of spurious correlations.

## Results

The adoption of the “vaccine reservation” query (VRQ) for our purpose is validated by the very strong correlation with the “covid vaccine” and “vaccine” queries (Multimedia Appendix 1, Figure S1) and the marked increase of its RSV in the period November 2020 to November 2021 compared to the past 4 years ( $\Delta=11,500\%$ ;  $t_{56}=6.8$ ;  $P^* < .001$ ). The keywords related to the desire not to get vaccinated registered an average RSV of 4% compared to “vaccine reservation.” VRQ's RSV has significantly exceeded that of searches for specific names such as “pfizer reservation,” “astrazeneca reservation,” “moderna reservation,” and “johnson&johnson reservation” ( $\delta=190\%$ ;  $t_{55}=6.6$ ;  $P^* < .001$ ). Table 1 shows very strong correlations between VRQ and the national vaccination ( $V$ ) trends (min  $r^2=0.460$ ;  $P^* < .001$ , lag 0 weeks; max  $r^2=0.903$ ;  $P^* < .001$ ; lag 6 weeks). Significant correlations were also highlighted between VRQ's RSV and the VRH of the newspaper “La Repubblica” (Multimedia Appendix 1, Table S1) and between VRH and  $V$  (Multimedia Appendix 1, Table S2). However, in these cases, the explained variations were markedly lower (max acceptable  $r^2=0.237$ ,  $P < .001$ ,  $P^* = .005$ , lag  $-3$  weeks; max acceptable  $r^2=0.286$ ,  $P < .001$ ,  $P^* = .002$ , lag 8 weeks). The differences between the Spearman coefficients were highly significant ( $z=6.16$ ,  $P^* < .001$ ;  $z=5.86$ ,  $P^* < .001$ ).

**Table 1.** Spearman cross-correlations (R) between the “vaccine reservation” query (VRQ) and vaccination administrations in Italy from November 2020 and November 2021. The highest correlation is obtained by shifting the VRQ 6 weeks ahead.

Lag week	R (VRQ vs V <sup>a</sup> ; 95% CI)	P value	P* value	N
-1	0.536 (0.297-0.711)	<.001	.002	47
0	0.678 (0.481-0.803)	<.001	<.001	48
1	0.777 (0.633-0.869)	<.001	<.001	48
2	0.833 (0.720-0.903)	<.001	<.001	48
3	0.887 (0.806-0.935)	<.001	<.001	48
4	0.927 (0.874-0.958)	<.001	<.001	48
5	0.946 (0.906-0.969)	<.001	<.001	48
6 <sup>b</sup>	0.950 (0.912-0.971)	<.001	<.001	48
7	0.946 (0.905-0.969)	<.001	<.001	48

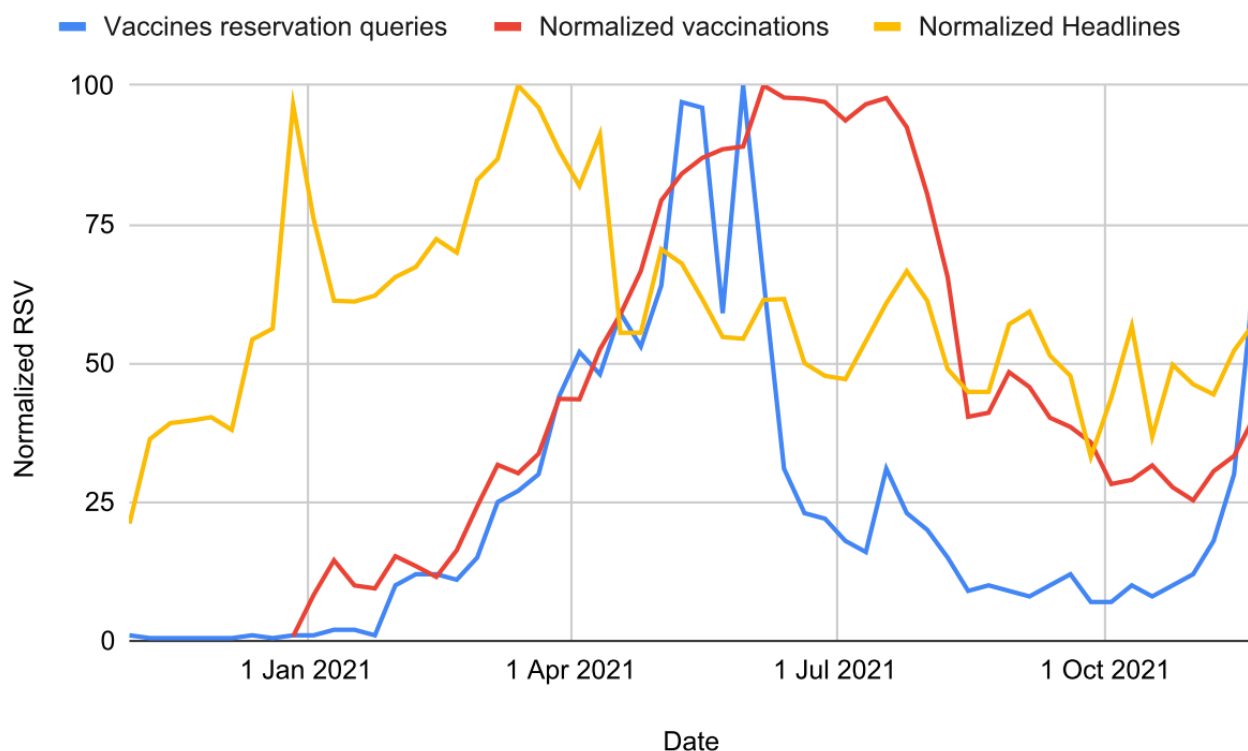
<sup>a</sup>V: vaccinations.

<sup>b</sup>The highest correlation is obtained by shifting the VRQ 6 weeks ahead.

The comparison of the trends is shown in Figure 1. All regional RSV trends have been similar to the national one (Multimedia Appendix 1, Figure S2) and were compatible with vaccination trends at the regional level [30]. Finally, the following regression model was built using appropriately translated time series based on the optimum lag previously identified (only values inside the acceptability range were considered):  $Sqrt(V) = A + B \times Log(VRH) + C \times Log(VRQ)$ , with  $A=-0.988$  (SE 1.930;  $P=.61$ ,  $P^*>.99$ ),  $B=2.67$  (SE 1.16;  $P=.03$ ,  $P^*=.29$ ),  $C=2.84$  (SE 0.22;  $P^*<.001$ ). We observe that VRQ significance was greater than VRH. The following assumptions were considered verified: residual normality (Shapiro-Wilk  $P=.38$ ), homoscedasticity

(White test  $P=.77$ ), and no multicollinearity (variance inflation factor [VIF]=1.46). Even considering an unlikely causal lag range of  $\pm 12$  weeks, VRQ is the most significant variable to predict vaccinations:  $Log(V) = A + B \times Log(VRH) + C \times Log(VRQ)$ , with  $A=0.381$  (SE 0.285;  $P=.19$ ,  $P^*>.99$ ),  $B=0.487$  (SE 0.180;  $P=.01$ ,  $P^*=.12$ ), and  $C=0.353$  (SE 0.041;  $P^*<.001$ ). Furthermore, despite that  $B>C$ , the 95% CIs are largely overlapping (overlap 0.308). The following assumptions were considered verified: residual normality (Shapiro-Wilk  $P=.86$ ), homoscedasticity (White test  $P=.23$ ), and no multicollinearity (VIF=2.45).

**Figure 1.** Comparison between the vaccine-related headlines of the newspaper "La Repubblica," national vaccinations, and national queries on vaccination reservations from November 2020 to November 2021. RSV: relative search volume.





## Discussion

### Principal Findings

This study shows a marked and significant cross-correlation between web queries on vaccine reservations and actual vaccinations against COVID-19 in Italy. Based on the lower cross-correlations between vaccine-related news and vaccine web searches, the mass media may have only partially influenced web searches related to vaccine booking. Nevertheless, even assuming a positive impact of the mass media on these queries, this does not compromise the adoption of Google Trends as a predictive tool for vaccinations: indeed, the mass media could push users to search for online information on vaccines and then book their administration. Furthermore, COVID-19 vaccine reservation is easily obtainable through a user-friendly online procedure proposed by the regional health organizations (eg, [31]). This fact helps explain the strong correlation between web searches and vaccinations. Therefore, it is likely that the cross-correlations found between vaccine-related queries and vaccinations are not spurious. Alongside this, it is necessary to consider that the Italian mass media have even risked compromising the effectiveness of the vaccination campaign against COVID-19 by providing infodemic news on rare side effects [32]. Hence, it is plausible that, given the high number of vaccinations achieved at the national level, more authoritative sources have also been consulted by users. The capacity to provide accurate predictions on vaccination trends several weeks in advance is an extremely relevant epidemiological tool for developing future containment strategies [33]. These findings show that Google Trends can be

exploited for this purpose if used properly. The search for simple well-targeted keywords on Google Trends is more likely to return the actual scenario of web interest on a certain topic. Specifically, it is essential not to use too complex or specific names, which tend to be ignored by users, and to try to express a precise action (in this case, the vaccine reservation).

Among the limitations of this paper, it is fair to emphasize that no definitive causal evidence has been provided, and unknown confounders may have skewed the results in unpredictable ways. Moreover, the variability of time lags between online booking and vaccine administration was not considered in this study. Finally, although well targeted, there are no guarantees that all the keywords relating to the desire not to be vaccinated have been selected. In this regard, given the broad antivaccination movement, many users may not have expressed an online interest in not getting vaccinated.

### Conclusions

This research provides preliminary evidence in favor of using Google Trends as a surveillance and prediction tool for vaccine adherence against COVID-19 in Italy. Further research is needed to establish appropriate use and limits of Google Trends for vaccination tracking. However, these findings prove that the search for suitable keywords is a fundamental step to reduce confounding factors. Additionally, targeting hypotheses helps diminish the likelihood of spurious correlations. It is recommended that Google Trends be leveraged as a complementary infoveillance tool by government agencies to monitor and predict vaccine adherence in this and future crises by following the methods proposed in this manuscript.

### Conflicts of Interest

None declared.

### Multimedia Appendix 1

Supplementary figures and tables.

[\[DOCX File, 391 KB-Multimedia Appendix 1\]](#)

### References

1. Sulyok M, Ferenci T, Walker M. Google Trends data and COVID-19 in Europe: correlations and model enhancement are European wide. *Transbound Emerg Dis* 2021 Jul 17;68(4):2610-2615. [doi: [10.1111/tbed.13887](https://doi.org/10.1111/tbed.13887)] [Medline: [33085851](https://pubmed.ncbi.nlm.nih.gov/33085851/)]
2. Springer S, Zieger M, Strzelecki A. The rise of infodemiology and infoveillance during COVID-19 crisis. *One Health* 2021 Dec;13:100288 [FREE Full text] [doi: [10.1016/j.onehlt.2021.100288](https://doi.org/10.1016/j.onehlt.2021.100288)] [Medline: [34277922](https://pubmed.ncbi.nlm.nih.gov/34277922/)]
3. 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 Mar 27;11(1):e11 [FREE Full text] [doi: [10.2196/jmir.1157](https://doi.org/10.2196/jmir.1157)] [Medline: [19329408](https://pubmed.ncbi.nlm.nih.gov/19329408/)]
4. Zitting K, Lammers-van der Holst HM, Yuan RK, Wang W, Quan SF, Duffy JF. Google Trends reveals increases in internet searches for insomnia during the 2019 coronavirus disease (COVID-19) global pandemic. *J Clin Sleep Med* 2021 Feb 01;17(2):177-184. [doi: [10.5664/jcsm.8810](https://doi.org/10.5664/jcsm.8810)] [Medline: [32975191](https://pubmed.ncbi.nlm.nih.gov/32975191/)]
5. Brodeur A, Clark AE, Fleche S, Powdhavee N. COVID-19, lockdowns and well-being: evidence from Google Trends. *J Public Econ* 2021 Jan;193:104346 [FREE Full text] [doi: [10.1016/j.jpubeco.2020.104346](https://doi.org/10.1016/j.jpubeco.2020.104346)] [Medline: [33281237](https://pubmed.ncbi.nlm.nih.gov/33281237/)]
6. Ho J, Hussain S, Sparagano O. Did the COVID-19 pandemic spark a public interest in pet adoption? *Front Vet Sci* 2021;8:647308. [doi: [10.3389/fvets.2021.647308](https://doi.org/10.3389/fvets.2021.647308)] [Medline: [34046443](https://pubmed.ncbi.nlm.nih.gov/34046443/)]
7. Hanna A, Hanna L. What, where and when? Using Google Trends and Google to investigate patient needs and inform pharmacy practice. *Int J Pharm Pract* 2019 Feb;27(1):80-87. [doi: [10.1111/ijpp.12445](https://doi.org/10.1111/ijpp.12445)] [Medline: [29602180](https://pubmed.ncbi.nlm.nih.gov/29602180/)]

8. Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, et al. The use of google trends in health care research: a systematic review. *PLoS One* 2014;9(10):e109583 [FREE Full text] [doi: [10.1371/journal.pone.0109583](https://doi.org/10.1371/journal.pone.0109583)] [Medline: [25337815](https://pubmed.ncbi.nlm.nih.gov/25337815/)]
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 Sep;7(3):185-189 [FREE Full text] [doi: [10.1016/j.jegh.2017.06.001](https://doi.org/10.1016/j.jegh.2017.06.001)] [Medline: [28756828](https://pubmed.ncbi.nlm.nih.gov/28756828/)]
10. Rovetta A. Reliability of Google Trends: analysis of the limits and potential of web infoveillance during COVID-19 pandemic and for future research. *Front Res Metr Anal* 2021;6:670226. [doi: [10.3389/frma.2021.670226](https://doi.org/10.3389/frma.2021.670226)] [Medline: [34113751](https://pubmed.ncbi.nlm.nih.gov/34113751/)]
11. Sato K, Mano T, Iwata A, Toda T. Need of care in interpreting Google Trends-based COVID-19 infodemiological study results: potential risk of false-positivity. *BMC Med Res Methodol* 2021 Jul 18;21(1):147 [FREE Full text] [doi: [10.1186/s12874-021-01338-2](https://doi.org/10.1186/s12874-021-01338-2)] [Medline: [34275447](https://pubmed.ncbi.nlm.nih.gov/34275447/)]
12. Rovetta A, Castaldo L. A new infodemiological approach through Google Trends: longitudinal analysis of COVID-19 scientific and infodemic names in Italy. *BMC Med Res Methodol* 2022 Jan 30;22(1):33 [FREE Full text] [doi: [10.1186/s12874-022-01523-x](https://doi.org/10.1186/s12874-022-01523-x)] [Medline: [35094682](https://pubmed.ncbi.nlm.nih.gov/35094682/)]
13. Harder T, Külper-Schiek W, Reda S, Treskova-Schwarzbach M, Koch J, Vygen-Bonnet S, et al. Effectiveness of COVID-19 vaccines against SARS-CoV-2 infection with the Delta (B.1.617.2) variant: second interim results of a living systematic review and meta-analysis, 1 January to 25 August 2021. *Euro Surveill* 2021 Oct;26(41):2100920 [FREE Full text] [doi: [10.2807/1560-7917.ES.2021.26.41.2100920](https://doi.org/10.2807/1560-7917.ES.2021.26.41.2100920)] [Medline: [34651577](https://pubmed.ncbi.nlm.nih.gov/34651577/)]
14. Rapporto annuale sulla sicurezza dei vaccini anti-COVID-19 27/12/2020 - 26/12/2021. Agenzia Italiana del Farmaco. URL: [https://www.aifa.gov.it/documents/20142/1315190/Rapporto\\_annuale\\_su\\_sicurezza\\_vaccini%20anti-COVID-19.pdf](https://www.aifa.gov.it/documents/20142/1315190/Rapporto_annuale_su_sicurezza_vaccini%20anti-COVID-19.pdf) [accessed 2022-02-11]
15. Istituto Superiore di Sanità. COVID-19: sorveglianza, impatto delle infezioni ed efficacia vaccinale. Facebook. URL: <https://www.facebook.com/ISS.social/posts/400586958532877> [accessed 2022-02-11]
16. Cadeddu C, Sapienza M, Castagna C, Regazzi L, Paladini A, Ricciardi W, et al. Vaccine hesitancy and trust in the scientific community in Italy: comparative analysis from two recent surveys. *Vaccines (Basel)* 2021 Oct 19;9(10):1206 [FREE Full text] [doi: [10.3390/vaccines9101206](https://doi.org/10.3390/vaccines9101206)] [Medline: [34696314](https://pubmed.ncbi.nlm.nih.gov/34696314/)]
17. Vaccine reservation query. Google Trends. URL: <https://trends.google.com/trends/explore?date=2020-11-01%202021-11-27&geo=IT&q=prenotazione%20vaccino> [accessed 2021-11-27]
18. Vaccine hesitancy query. Google Trends. URL: <https://tinyurl.com/5dewb4j3> [accessed 2022-03-23]
19. owid/covid-19-data. GitHub. URL: [https://github.com/owid/covid-19-data/blob/master/public/data/vaccinations/country\\_data/Italy.csv](https://github.com/owid/covid-19-data/blob/master/public/data/vaccinations/country_data/Italy.csv) [accessed 2021-11-27]
20. Archivio. La Repubblica. 2021. URL: <https://ricerca.repubblica.it/repubblica/archivio/repubblica/2021> [accessed 2021-11-28]
21. Rovetta A, Castaldo L. Influence of mass media on Italian web users during the COVID-19 pandemic: infodemiological analysis. *JMIRx Med* 2021;2(4):e32233 [FREE Full text] [doi: [10.2196/32233](https://doi.org/10.2196/32233)] [Medline: [34842858](https://pubmed.ncbi.nlm.nih.gov/34842858/)]
22. Strömberg D. Mass media competition, political competition, and public policy. *Rev Econ Stud* 2004;71(1):284. [doi: [10.1111/0034-6527.00284](https://doi.org/10.1111/0034-6527.00284)]
23. Rovetta A. Raiders of the lost correlation: a guide on Using Pearson and Spearman coefficients to detect hidden correlations in medical sciences. *Cureus* 2020 Nov 30;12(11):e11794 [FREE Full text] [doi: [10.7759/cureus.11794](https://doi.org/10.7759/cureus.11794)] [Medline: [33409040](https://pubmed.ncbi.nlm.nih.gov/33409040/)]
24. Kwak SG, Kim JH. Central limit theorem: the cornerstone of modern statistics. *Korean J Anesthesiol* 2017 Apr;70(2):144-156 [FREE Full text] [doi: [10.4097/kjae.2017.70.2.144](https://doi.org/10.4097/kjae.2017.70.2.144)] [Medline: [28367284](https://pubmed.ncbi.nlm.nih.gov/28367284/)]
25. Fagerland MW. t-tests, non-parametric tests, and large studies--a paradox of statistical practice? *BMC Med Res Methodol* 2012 Jun 14;12:78 [FREE Full text] [doi: [10.1186/1471-2288-12-78](https://doi.org/10.1186/1471-2288-12-78)] [Medline: [22697476](https://pubmed.ncbi.nlm.nih.gov/22697476/)]
26. Multiple Linear Regression Calculator. Statistics Kingdom. URL: [https://www.statskingdom.com/410multi\\_linear\\_regression.html](https://www.statskingdom.com/410multi_linear_regression.html) [accessed 2022-02-11]
27. Ming W, Huang F, Chen Q, Liang B, Jiao A, Liu T, et al. Understanding health communication through Google Trends and news coverage for COVID-19: multinational study in eight countries. *JMIR Public Health Surveill* 2021 Dec 21;7(12):e26644 [FREE Full text] [doi: [10.2196/26644](https://doi.org/10.2196/26644)] [Medline: [34591781](https://pubmed.ncbi.nlm.nih.gov/34591781/)]
28. Rovetta A, Bhagavathula AS. Global infodemiology of COVID-19: analysis of Google web searches and Instagram hashtags. *J Med Internet Res* 2020 Aug 25;22(8):e20673 [FREE Full text] [doi: [10.2196/20673](https://doi.org/10.2196/20673)] [Medline: [32748790](https://pubmed.ncbi.nlm.nih.gov/32748790/)]
29. Skelly AC, Dettori JR, Brodt ED. Assessing bias: the importance of considering confounding. *Evid Based Spine Care J* 2012 Feb;3(1):9-12 [FREE Full text] [doi: [10.1055/s-0031-1298595](https://doi.org/10.1055/s-0031-1298595)] [Medline: [23236300](https://pubmed.ncbi.nlm.nih.gov/23236300/)]
30. Mastrolonardo R. Vaccino Covid: dati e grafici sulle somministrazioni in Italia, regione per regione. Sky TG24. URL: <https://tg24.sky.it/cronaca/approfondimenti/dati-vaccini-covid-italia> [accessed 2021-11-29]
31. Prenotazione Vaccinazioni Anti COVID-19 in Lombardia. URL: <https://prenotazionevaccinicovid.regione.lombardia.it/> [accessed 2021-11-29]
32. Rovetta A. The impact of COVID-19 on conspiracy hypotheses and risk perception in Italy: infodemiological survey study using Google Trends. *JMIR Infodemiology* 2021;1(1):e29929 [FREE Full text] [doi: [10.2196/29929](https://doi.org/10.2196/29929)] [Medline: [34447925](https://pubmed.ncbi.nlm.nih.gov/34447925/)]

33. Bian L, Gao Q, Gao F, Wang Q, He Q, Wu X, et al. Impact of the Delta variant on vaccine efficacy and response strategies. *Expert Rev Vaccines* 2021 Oct;20(10):1201-1209 [FREE Full text] [doi: [10.1080/14760584.2021.1976153](https://doi.org/10.1080/14760584.2021.1976153)] [Medline: [34488546](https://pubmed.ncbi.nlm.nih.gov/34488546/)]

## Abbreviations

- RSV:** relative search volume  
**VIF:** variance inflation factor  
**VRH:** vaccine-related headlines  
**VRQ:** vaccine reservation query

*Edited by E Meinert; submitted 01.12.21; peer-reviewed by A Strzelecki, Z Shah, A Chang; comments to author 29.01.22; revised version received 15.02.22; accepted 06.03.22; published 19.04.22*

*Please cite as:*

*Rovetta A*

*Google Trends as a Predictive Tool for COVID-19 Vaccinations in Italy: Retrospective Infodemiological Analysis*

*JMIRx Med* 2022;3(2):e35356

URL: <https://med.jmirx.org/2022/2/e35356>

doi: [10.2196/35356](https://doi.org/10.2196/35356)

PMID:

©Alessandro Rovetta. Originally published in JMIRx Med (<https://med.jmirx.org>), 19.04.2022. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIRx Med, is properly cited. The complete bibliographic information, a link to the original publication on <https://med.jmirx.org/>, as well as this copyright and license information must be included.