

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

Alessandro Rovetta

Submitted to: JMIRx Med
on: December 01, 2021

Disclaimer: © The authors. All rights reserved. This is a privileged document currently under peer-review/community review. Authors have provided JMIR Publications with an exclusive license to publish this preprint on its website for review purposes only. While the final peer-reviewed paper may be licensed under a CC BY license on publication, at this stage authors and publisher expressly prohibit redistribution of this draft paper other than for review purposes.

Table of Contents

Original Manuscript..... 4

Supplementary Files..... 16

 Figures 17

 Figure 1..... 18

 Figure 2..... 19

 Figure 3..... 20

 Multimedia Appendixes 21

 Multimedia Appendix 1..... 22

 Multimedia Appendix 2..... 22

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

Alessandro Rovetta¹

¹R&C Research Bovezzo (BS) IT

Corresponding Author:

Alessandro Rovetta
R&C Research
Via Brede Traversa II
Bovezzo (BS)
IT

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 brief 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 towards vaccinations in Italy from November 2020 to November 2021. The keyword "vaccine reservation" (VRQ) was chosen as it reflects a real intention of being vaccinated (V). Furthermore, the impact of the second most read Italian newspaper on vaccines-related web searches was investigated to evaluate the role of the mass media as a confounding factor.

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 $r^2 = .460$, $P < .001$, lag = 0 weeks; max $r^2 = .903$, $P < .001$, lag = 6 weeks). Cross-correlations between VRQ and news about COVID-19 vaccines have been markedly lower and characterized by greater lags (min $r^2 = .190$, $P = .001$, lag = 0 weeks; max $r^2 = .493$, $P < .001$, lag = -10 weeks). No correlation between news and vaccinations was sought since the lag would have been too high.

Conclusions: This research provides strong evidence in favor of using Google Trends as a surveillance and prediction tool for vaccine adherence against COVID-19 in Italy. 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.

(JMIR Preprints 01/12/2021:35356)

DOI: <https://doi.org/10.2196/preprints.35356>

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.

No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✓ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain visible to all users.
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in [http](#)

Original Manuscript

Original brief paperAlessandro Rovetta¹¹ R&C Research, Brescia, Italy.**Google Trends as a Predictive Tool for COVID-19 Vaccinations in Italy: A Retrospective Infodemiological Analysis.****Corresponding Author:**

Alessandro Rovetta,

R&C Research,

Brescia, IT

Phone: 39 3927112808

Email: rovetta.mresearch@gmail.com

ORCID: 0000-0002-4634-279X

WoS ID: AAT-9063-2020

Keywords

COVID-19, Epidemiology, Google Trends, Infodemiology, Infoveillance, Italy, Public Health, SARS-CoV-2, Vaccinations, Vaccines.

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 brief 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 towards vaccinations in Italy from November 2020 to November 2021. The keyword "vaccine reservation" (VRQ) was chosen as it reflects a real intention of being vaccinated (V). Furthermore, the impact of the second most read Italian newspaper (VRH) on vaccines-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 (δ) were used to compare Spearman coefficients. A regression model $V=f(VRH, VRQ)$ was built to validate the results found. Holm-Bonferroni correction was adopted (P^*). Standard errors are reported after " \pm ."

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 $r^2 = .460$, $P^* < .001$, lag = 0 weeks; max $r^2 = .903$, $P^* < .001$, 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 vs. VRH ($P^* < .001$ vs. $P = .026$, $P^* = .286$).

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.

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 employed by the scientific community to conduct infodemiological and epidemiological analyzes [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 API/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 drastically reducing the severity of the disease [13]. For Instance, the latest report from the Italian Medicines Agency (AIFA) highlighted a shallow risk associated with vaccines despite high protection against COVID-19 [14]. Even considering Omicron's more elusive variant of concern, the rates of hospitalizations, ICUs, 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 vaccines-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 synonyms queries had a much lower 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 datasets, 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 = 1 November 2020 – 27 November 2021, category = All categories, search type = Web Search. The "period" parameter has been changed to "Past 5 years" when performing a historical timeseries analysis. The "region" parameter was changed from "Italy" to "[the name of the region concerned]" when analyzing regional trends. The "interest over time" datasets 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 vaccine." 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 dataset 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 covering the period November 2020 - 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 very 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.

Statistical analysis

The shape of the data distribution was assessed both graphically and through the Shapiro-Wilk test. Since the datasets were not normal ($P < .001$) and above or below threshold correlations were not of interest, we adopted the Spearman correlation (R) [27]. To check the discrepancy between two timeseries, quantifiers such as percentage difference (used to compare the average RSV of two simultaneous series and indicated with " δ ") and percentage increase (used to compare the average RSV of two consecutive series and indicated with " Δ ") 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 non-normal datasets [28, 29]. When two contiguous timeseries were compared, a graphic check was carried out to guarantee the absence of seasonalities and trends. All datasets were normalized to 100 by multiplying individual values by the constant "100/dataset 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 $Y=f(VRH, VRQ)$ to evaluate the impact of VRH and VRQ on V [23]. Standard errors for the regression coefficients are reported after " \pm ." Based on previous literature, any causal correlations between the media clamor and web searches should be sought within a maximum of ± 3 weeks (acceptability range) [9, 11, 21, 24, 25]. Indeed, the web interest in a topic must arise around the media hype peak to be considered a direct consequence of the latter. Regarding the pairs (VRH, V) and (VRQ, V), the lag acceptability range was fixed at 0 – 8 weeks since it can take up to two 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 discussed above, there is solid evidence that mass media can significantly 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 (un)correlated variables [26]. 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 growing users' web interest in the disease. After seven days, an increase in COVID-19 cases is registered. Examining the sole couple (user interest, COVID-19 cases), it could seem that 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 (Figure S1) and the marked increase of its relative search volume in the period November 2020 - November 2021 compared to the past four years ($\Delta=11,500\%$, $t=6.8$, $P^* < .001$). The keywords related to the desire not to get vaccinated registered an average RSV of 4% compared to "vaccine reservation." VRQ's relative search volume has significantly exceeded that of searches for specific names such as "pfizer reservation," "astrazeneca reservation," "moderna reservation," and "johnson&johnson reservation" ($\delta = 190\%$, $t = 6.6$, $P^* < .001$). Table 1 shows very strong correlations between VRQ and the national vaccination (V) trends (min $r^2 = .460$, $P^* < .001$, lag = 0 weeks; max $r^2 = .903$, $P^* < .001$, lag = 6 weeks). Significant correlations were also highlighted between VRQ's relative search volume and the vaccines-related headlines (VRH) of the newspaper "La Repubblica" (Table S1), and between VRH and V (Table S2). However, in these cases, the explained variations were markedly lower (max acceptable $r^2 = .237$, $P < .001$, $P^* = .005$, lag = -3 weeks; max acceptable $r^2 = .286$, $P < .001$, $P^* = .002$, lag = 8 weeks). The differences between Spearman coefficients were highly significant ($z=6.16$, $P^* < .001$; $z=5.86$, $P^* < .001$).

Lag week	R (VRQ vs V)	95% CI low	95% CI up	P	P*	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	0.950	0.912	0.971	<.001	<.001	48
7	0.946	0.905	0.969	<.001	<.001	48

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

The comparison of the trends is shown in Figure 1. All regional RSV trends have been similar to the national one (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 identified above (only values inside the acceptability range were considered): $Sqrt(V)=A+B\times Log(VRH)+C\times Log(VRQ)$, with $A=-0.988\pm1.930$ ($P=.611$, $P^*>.999$), $B=2.67\pm1.16$ ($P=.026$, $P^*=.286$), $C=2.84\pm0.22$ ($P^*<.001$). We observe that VRQ significance is greater than VRH. The following assumptions were considered verified: residual normality (Shapiro-Wilk $P=.377$), homoscedasticity (White test $P=.767$), no multicollinearity ($VIF=1.46$). Even considering an unlikely causal lag range of ± 12 weeks, VRQ is the most significant variable to predict vaccinations: $Log(V)=A+B\times Log(VRH)+C\times Log(VRQ)$, with $A=0.381\pm0.285$ ($P=.188$, $P^*>.999$), $B=0.487\pm0.180$ ($P=.010$, $P^*=.120$), $C=0.353\pm0.041$ ($P^*<.001$). Furthermore, despite $B>C$, the 95% confidence intervals are largely overlapping (overlap=0.308). The following assumptions were considered verified: residual normality (Shapiro-Wilk $P=.861$), homoscedasticity (White test $P=.227$), no multicollinearity ($VIF=2.45$).

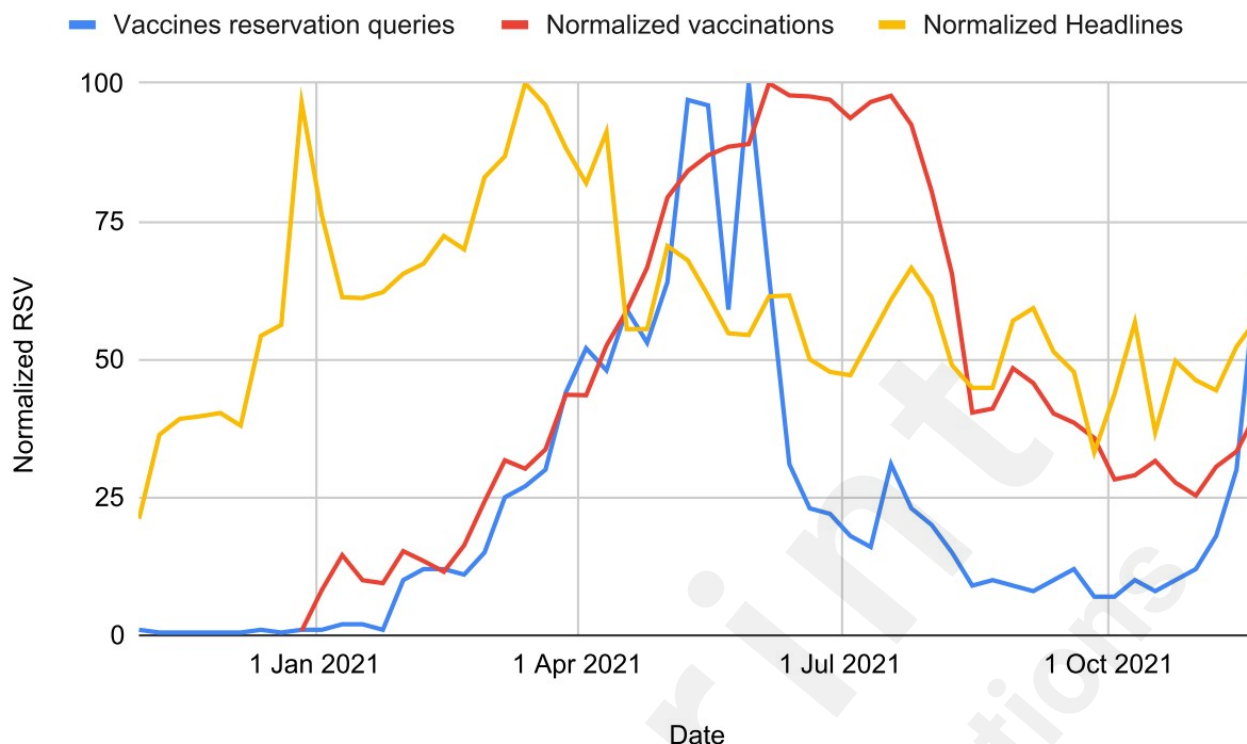


Figure 1. Comparison between the vaccines-related headlines of the newspaper "La Repubblica," national vaccinations, and national queries on vaccination reservations from November 2020 to November 2021.

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 (e.g., [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 very 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 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 anti-vaccination 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 intelligence tool by government agencies to monitor and predict vaccine adherence in this and future crises by following the methods proposed in this manuscript.

Ethical declaration

This study does not involve human subjects and/or animals. All Google Trends data is anonymized. Therefore, the research does not require approval from a committee. No funding was received. The author declares that he has no conflicts of interest.

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;68(4):2610-2615. doi: 10.1111/tbed.13887. Epub 2020 Nov 17. PMID: 33085851.
- [2] Springer S, Zieger M, Strzelecki A. The rise of infodemiology and infoveillance during COVID-19 crisis. *One Health*. 2021 Jul 3;13:100288. doi: 10.1016/j.onehlt.2021.100288. PMID: 34277922; PMCID: PMC8271150.
- [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. doi: 10.2196/jmir.1157. PMID: 19329408; PMCID: PMC2762766.
- [4] Zitting KM, 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 1;17(2):177-184. doi: 10.5664/jcsm.8810. PMID: 32975191; PMCID: PMC7853234.
- [5] Brodeur A, Clark AE, Fleche S, Powdthavee N. COVID-19, lockdowns and well-being: Evidence from Google Trends. *J Public Econ*. 2021 Jan;193:104346. doi: 10.1016/j.jpubeco.2020.104346. Epub 2020 Nov 30. PMID: 33281237; PMCID: PMC7703221.
- [6] Ho J, Hussain S, Sparagano O. Did the COVID-19 Pandemic Spark a Public Interest in Pet Adoption? *Front Vet Sci*. 2021 May 7;8:647308. doi: 10.3389/fvets.2021.647308. PMID: 34046443; PMCID: PMC8145284.
- [7] Hanna A, Hanna LA. 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. Epub 2018 Mar 30. PMID: 29602180.
- [8] Nuti SV, Wayda B, Ranasinghe I, Wang S, Dreyer RP, Chen SI, Murugiah K. The use of google trends in health care research: a systematic review. *PLoS One*. 2014 Oct 22;9(10):e109583. doi: 10.1371/journal.pone.0109583. PMID: 25337815; PMCID: PMC4215636.
- [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. doi: 10.1016/j.jegh.2017.06.001. Epub 2017 Jun 9. PMID: 28756828; PMCID: PMC7320449.
- [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 May 25;6:670226. doi: 10.3389/frma.2021.670226. PMID: 34113751; PMCID: PMC8186442.
- [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. doi: 10.1186/s12874-021-01338-2. PMID: 34275447; PMCID: PMC8286439.
- [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. doi: 10.1186/s12874-022-01523-x. PMID: 35094682; PMCID: PMC8801192.
- [13] Harder T, Külper-Schiek W, Reda S, Treskova-Schwarzbach M, Koch J, Vygen-Bonnet S, Wichmann O. 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. doi: 10.2807/1560-7917.ES.2021.26.41.2100920. PMID: 34651577; PMCID: PMC8518304.
- [14] Rapporto annuale sulla sicurezza dei vaccini anti-COVID-19 27/12/2020 - 26/12/2021. Agenzia Italiana del Farmaco (AIFA). 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: Feb 11, 2022].

[15] Istituto Superiore di Sanità. COVID-19: sorveglianza, impatto delle infezioni ed efficacia vaccinale. URL: <https://www.facebook.com/ISS.social/posts/400586958532877> [Accessed: Feb 11, 2022].

[16] Cadeddu C, Sapienza M, Castagna C, Regazzi L, Paladini A, Ricciardi W, Rosano A. Vaccine Hesitancy and Trust in the Scientific Community in Italy: Comparative Analysis from Two Recent Surveys. *Vaccines* (Basel). 2021 Oct 19;9(10):1206. doi: 10.3390/vaccines9101206. PMID: 34696314; PMCID: PMC8540200.

[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: Nov 27, 2021]

[18] Vaccine hesitancy query. Google Trends. URL: <https://trends.google.com/trends/explore?date=2020-11-01%202021-11-27&geo=IT&q=disdire%20vaccino%20%2B%20cancellare%20vaccino%20%2B%20evitare%20vaccino%20%2B%20non%20vaccinarsi%20%2B%20green%20pass%20falso,comprare%20green%20pass.prenotazione%20vaccino,%22prenotazione%20vaccino%22>

[19] Github, COVID-19 Data. URL: https://github.com/owid/covid-19-data/blob/master/public/data/vaccinations/country_data/Italy.csv [Accessed: Nov 27, 2021].

[20] La Repubblica, Archivio 2021. URL: <https://ricerca.repubblica.it/repubblica/archivio/repubblica/2021> [Accessed: Nov 28, 2021].

[21] Rovetta A, Castaldo L. Influence of Mass Media on Italian Web Users During the COVID-19 Pandemic: Infodemiological Analysis. *JMIRx Med*. 2021 Oct 18;2(4):e32233. doi: 10.2196/32233. PMID: 34842858; PMCID: PMC8601032.

[22] Strömberg, D. Mass Media Competition, Political Competition, and Public Policy. *The Review of Economic Studies*. 2004;71(1), 265–284.

[23] Multiple Linear Regression Calculator. Statistics Kingdom. URL: [https://www.statskingdom.com/410multi linear regression.html](https://www.statskingdom.com/410multi+linear+regression.html) [Accessed: Feb 11, 2022].

[24] Ming WK, Huang F, Chen Q, Liang B, Jiao A, Liu T, Wu H, Akinwunmi B, Li J, Liu G, Zhang CJP, Huang J, Liu Q. 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. doi: 10.2196/26644. PMID: 34591781; PMCID: PMC8691414.

[25] 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. doi: 10.2196/20673. PMID: 32748790; PMCID: PMC7458585.

[26] Skelly AC, Dettori JR, Brodt ED. Assessing bias: the importance of considering confounding. *Evid Based Spine Care J*. 2012 Feb;3(1):9-12. doi: 10.1055/s-0031-1298595. PMID: 23236300; PMCID: PMC3503514.

[27] 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. doi: 10.7759/cureus.11794. PMID: 33409040; PMCID: PMC7779167.

[28] Kwak SG, Kim JH. Central limit theorem: the cornerstone of modern statistics. *Korean J Anesthesiol*. 2017 Apr;70(2):144-156. doi: 10.4097/kjae.2017.70.2.144. Epub 2017 Feb 21. PMID: 28367284; PMCID: PMC5370305.

[29] Fagerland MW. t-tests, non-parametric tests, and large studies--a paradox of statistical practice? *BMC Med Res Methodol*. 2012 Jun 14;12:78. doi: 10.1186/1471-2288-12-78. PMID: 22697476; PMCID: PMC3445820.

[30] Sky tg24, Vaccino Covid: dati e grafici sulle somministrazioni in Italia, regione per regione.

URL: <https://tg24.sky.it/cronaca/approfondimenti/dati-vaccini-covid-italia> [Accessed: Nov 29, 2021].

[31] Regione Lombardia, Campagna vaccinazione anti COVID-19. URL: <https://prenotazionevaccinacovid.regione.lombardia.it/> [Accessed: Nov 29, 2021].

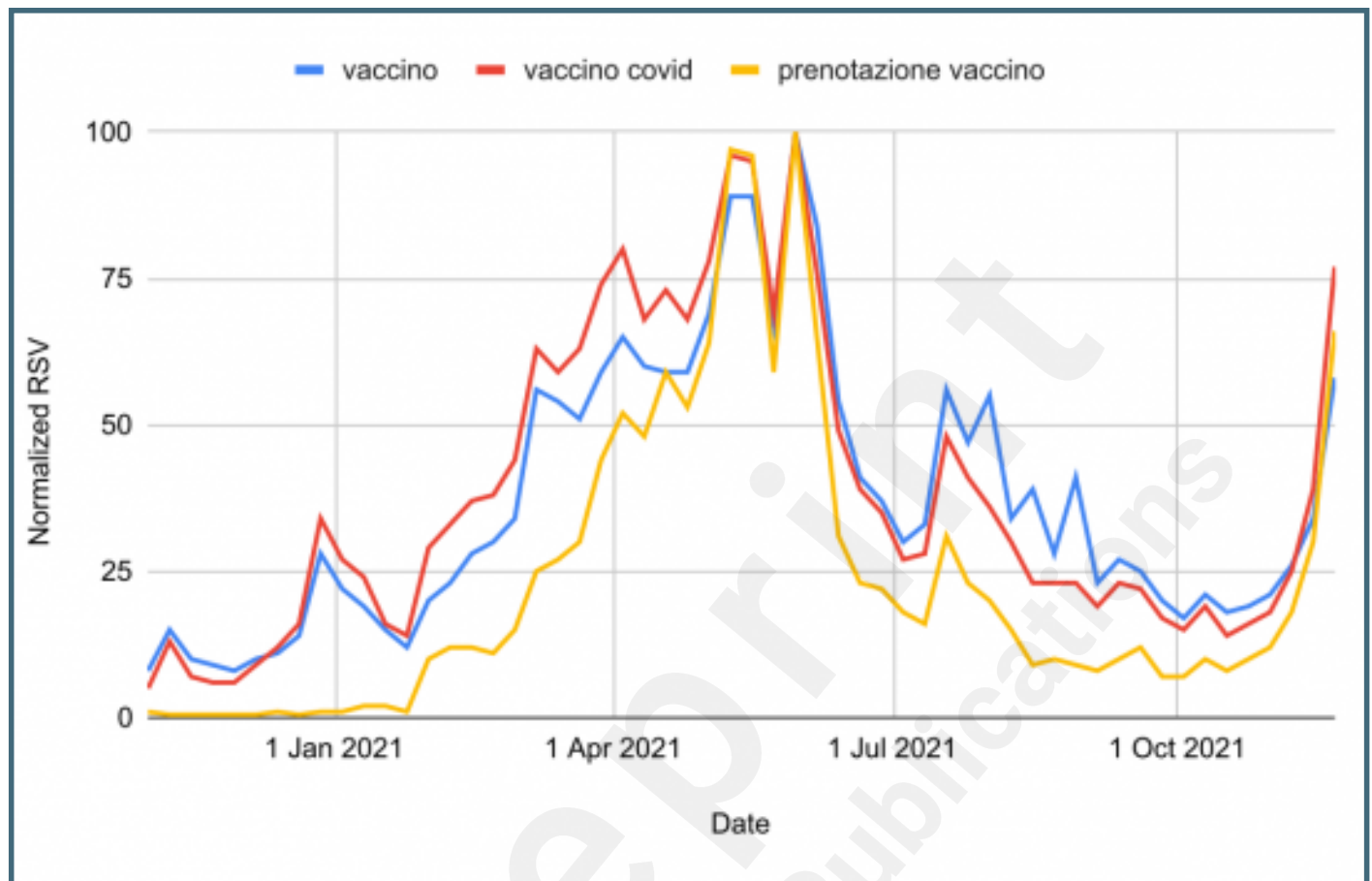
[32] Rovetta A. The Impact of COVID-19 on Conspiracy Hypotheses and Risk Perception in Italy: Infodemiological Survey Study Using Google Trends. JMIR Infodemiology. 2021 Aug 6;1(1):e29929. doi: 10.2196/29929. PMID: 34447925; PMCID: PMC8363126.

[33] Bian L, Gao Q, Gao F, Wang Q, He Q, Wu X, Mao Q, Xu M, Liang Z. Impact of the Delta variant on vaccine efficacy and response strategies. Expert Rev Vaccines. 2021 Oct;20(10):1201-1209. doi: 10.1080/14760584.2021.1976153. Epub 2021 Sep 9. PMID: 34488546; PMCID: PMC8442750.

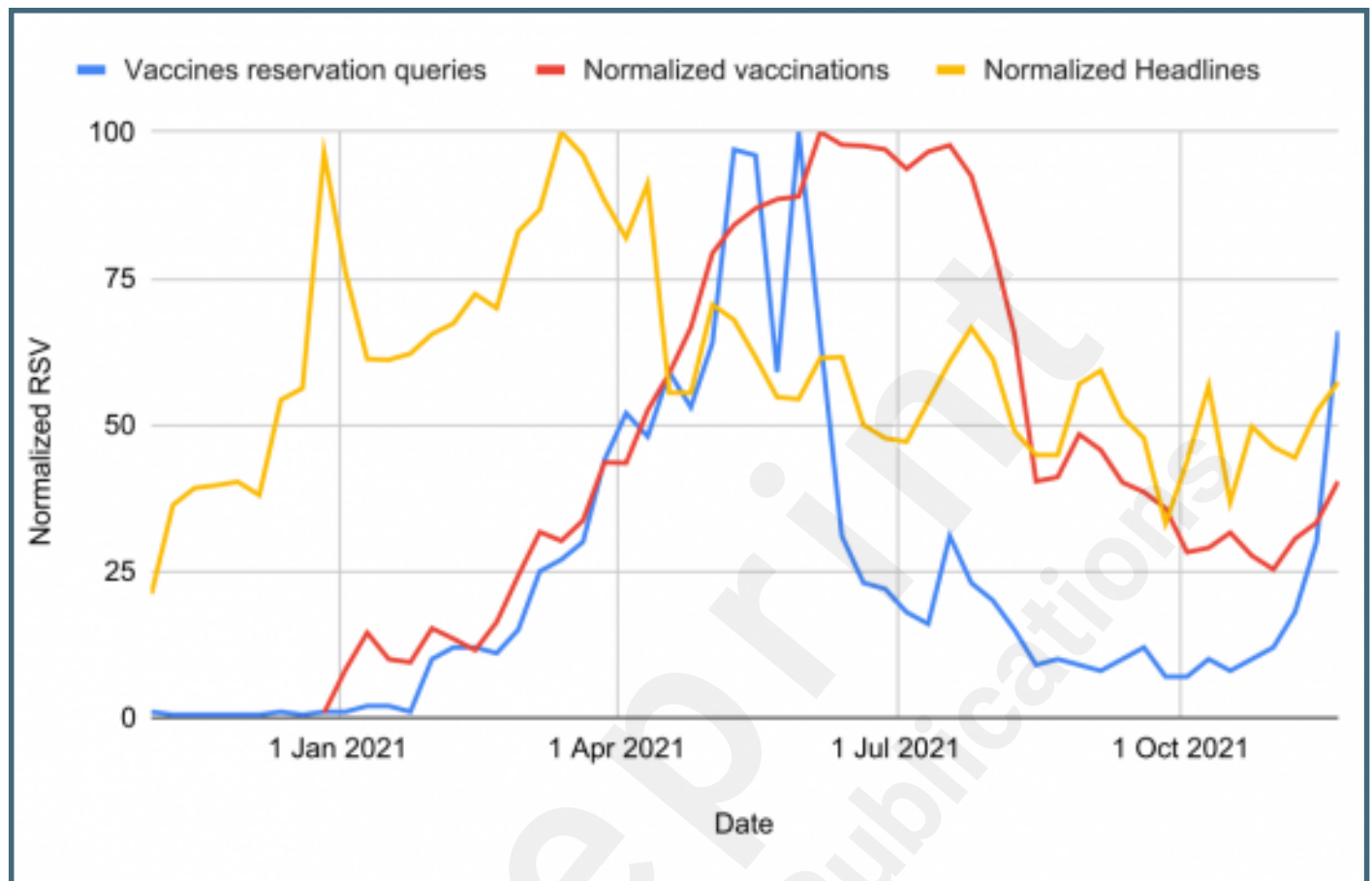
Supplementary Files

Figures

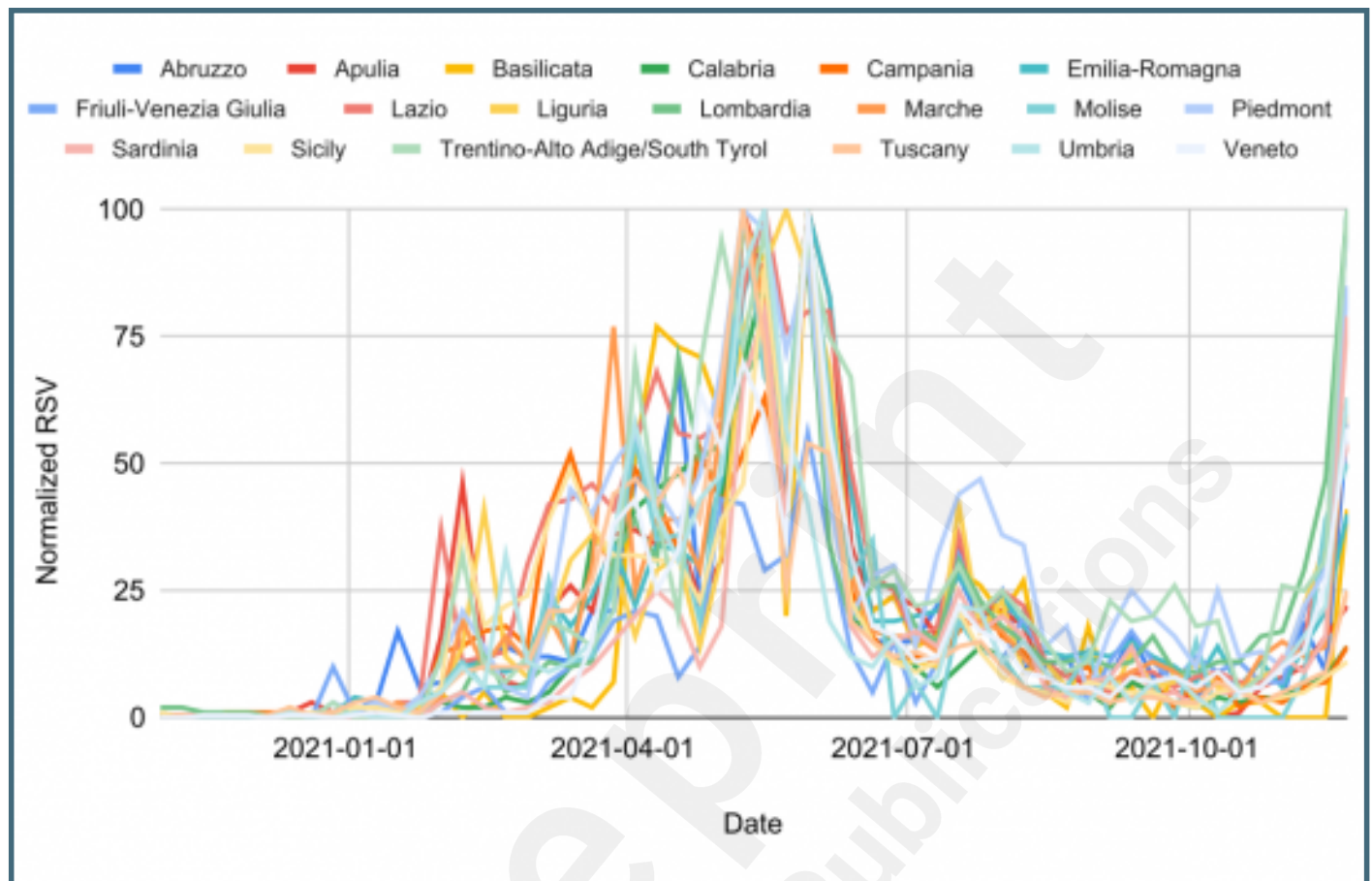
Normalized relative search volumes of the queries: "vaccino" (vaccine), "vaccino covid" (covid vaccine), and "prenotazione vaccino" (vaccine reservation).



Comparison between the vaccines-related headlines of the newspaper "La Repubblica," national vaccinations, and national queries on vaccination reservations from November 2020 to November 2021.



Comparison between the "prenotazione vaccino" (vaccine reservation) queries of all Italian regions from November 2020 to November 2021.



Multimedia Appendixes

Supplementary figures and tables.

URL: <http://asset.jmir.pub/assets/ffd0d065d98fd21fb9fe8d2fa1149b5b.docx>

Manuscript 35356 JMIR R1 - Tracked Changes.

URL: <http://asset.jmir.pub/assets/f2becc20ffc4e5e4a7fbf518c1adcb3e.docx>

