Taylor & Francis
Taylor & Francis Group

RESEARCH PAPER

🔓 OPEN ACCESS | Check for updates

Twitter as a sentinel tool to monitor public opinion on vaccination: an opinion mining analysis from September 2016 to August 2017 in Italy

Lara Tavoschi [a], Filippo Quattrone [a], Eleonora D'Andrea[b], Pietro Ducange [b], Marco Vabanesi [c], Francesco Marcelloni [b], and Pier Luigi Lopalco [a]

aDepartment of Translational Research and of New Surgical and Medical Technologies, University of Pisa, Pisa, Italy; bDipartimento di Ingegneria dell'Informazione, University of Pisa, Pisa, Italy; cDepartment of Neurology, San Raffaele Scientific Institute and University, Milan, Italy

ABSTRACT

Social media have become a common way for people to express their personal viewpoints, including sentiments about health topics. We present the results of an opinion mining analysis on vaccination performed on Twitter from September 2016 to August 2017 in Italy. Vaccine-related tweets were automatically classified as against, in favor or neutral in respect of the vaccination topic by means of supervised machine-learning techniques. During this period, we found an increasing trend in the number of tweets on this topic. According to the overall analysis by category, 60% of tweets were classified as neutral, 23% against vaccination, and 17% in favor of vaccination. Vaccine-related events appeared able to influence the number and the opinion polarity of tweets. In particular, the approval of the decree introducing mandatory immunization for selected childhood diseases produced a prominent effect in the social discussion in terms of number of tweets. Opinion mining analysis based on Twitter showed to be a potentially useful and timely sentinel system to assess the orientation of public opinion toward vaccination and, in future, it may effectively contribute to the development of appropriate communication and information strategies.

Introduction

In recent years, vaccination has become a controversial topic in public debate worldwide and Vaccine Hesitancy (VH), defined as "delay in acceptance or refusal of vaccination despite the availability of vaccination services"[1] is an increasingly important issue for country immunization programs. Diffusion of incomplete or wrong information by media about the effectiveness and safety of vaccines (e.g. the alleged connection between vaccines and autism) has been shown to be a determinant of this loss of trust in vaccination.[2]

In Italy, this phenomenon has led to an alarming drop in vaccination coverage since 2013.[3] Studies on traditional (e.g. newspapers) and social media (e.g. YouTube, Twitter) have found that in the last decade rumors, myths and disinformation regarding vaccines have been widely broadcasted, resulting in a negative impact on public opinion and people's willingness to be vaccinated.[4–6]

The drop in vaccination coverage, and the subsequent measles epidemic in 2017 with about 4885 cases and 4 deaths,[7] has attracted the interest of concerned experts, people, and media, stirring a heated political debate. In particular, two events happened in 2017 have dominated the scene in Italy:

- The publication of the National Immunization Prevention Plan (Piano Nazionale Prevenzione Vaccinale, PNPV) 2017–19 (January 19th, 2017)[8]

- The Legislative Decree n. 73 (June 7th, 2017) introducing compulsory vaccination for *Haemophilus influenzae* type b, measles, mumps, rubella, varicella and whooping cough (*pertussis*) for school-aged children in order to attend educational services, in addition to diphtheria, tetanus, polio and hepatitis B that were already mandatory (Vaccines decree).

Both events have been accompanied by strong public debate, also in the social media.[5,6] People likely share their viewpoints on social networks, including sentiments or behavior about health topics.[9] Among social network platforms, Twitter, counting in Italy about 6.4 million active users, has been widely used. Due to its specific features allowing instant posting of brief status update messages (*tweets*), Twitter is being explored more and more in the scientific literature as a source of health-related information, on a wide range of topics.[10–12] In particular, Twitter may be useful to capture real-time changes in public perception about vaccination, potentially providing a fast, low-cost, and easy alternative to traditional polls and surveys.

However, monitoring social media requires the ability to automatically analyze and interpret large amount of data in text format. This activity is known as text mining. Text mining refers to the process of automatic extraction of meaningful information and knowledge from unstructured natural language text.[13] The main difficulty in text mining is caused

by the vagueness of natural language.[13] More precisely, opinion mining refers to a special sub-field of text mining aimed at automatically determining the opinion polarity (positive, neutral, or negative, agree or disagree, etc.) associated with natural language texts.[14] This is challenged by ambiguity, the presence of sarcasm or irony in the text, or complex views on the same topic, e.g. one can be in favor of vaccinations but against the obligation of law. In addition, another task of opinion mining is to distinguish among objective and subjective texts. A subjective text, i.e., a single person's opinion, has a viewpoint, or a bias. An objective text, i.e., a fact, is meant to be completely unbiased, e.g., a news article, a neutral text.

Opinion mining may be performed with different approaches: machine learning, lexicon-based, and hybrid approaches.[15] Lexicon-based approaches perform better when used for general boundless contexts (i.e., without topic), with well-formed and grammatically correct texts, and are less suited for social networks where an informal language is used and context-related words are often missing or changes dynamically.[16] Instead, supervised machine-learning approaches overcome these problems.[17] Machine learning refers to algorithms and techniques able to automatically learn directly from data. Supervised learning is the dominant machine-learning approach. It consists of building, in an inductive way, a predictive model able to learn from a set of training data. The training data is a set of labeled examples, with each example being a pair consisting of an input object (described in terms of a set of features) and a desired output value, i.e., a class label in the case of a classification model. Once the training of the model is completed, the model is ready to be applied to new data.

The aim of this study was to monitor the public opinion on vaccination through Twitter using a machine-learning model to automatically assess opinion polarity, in relation to significant vaccine-related events occurred between September 2016 to August 2017 in Italy.

Methods

Selection of tweets and preprocessing

A dataset of tweets obtained from the Italian Twitter stream from September 2016 to August 2017 was identified and collected using keywords and hashtags related to vaccination, vaccine-preventable diseases and possible or alleged vaccine side effects. Examples of adopted keywords and hashtags are: "vaccini", "vaccino" (vaccine(s)); "controindicazioni vaccinali" (vaccine contraindications); "autismo" (autism); "malattie autoimmuni" (autoimmune diseases); #novaccino (hashtag for "no vaccine"); #iovaccino (hashtag for "I vaccinate"); #libertadiscelta (hashtag for "freedom of choice"). The complete set of keywords and extended methods have been published elsewhere.[18]

The extracted tweets were then pre-processed in preparation for the automatic classification by means of machine-learning techniques. Text preprocessing consisted of the elimination of useless information and the transformation of the tweets into numeric vectors, which can be processed by a machine-

learning algorithm. The first step of preprocessing is aimed to extract only the useful text from each tweet, e.g., links and mentions are discarded. The timestamp of each tweet is temporarily discarded for the purposes of text mining elaboration, but reconsidered for the analysis of temporal trends. Hashtags were reduced to single words eliminating the hash (#) symbol. Finally, a case-folding operation is applied to the texts, in order to convert all characters to lower case form.

Then, pre-defined text elaboration steps were applied to the tweets, with the aim of transforming the set of strings (i.e., the texts of tweets) in a structured form consisting in a set of numeric vectors (referred to as *features*). This approach is defined as *Bag-Of-Words* (BOW) text representation.[19] In particular, each tweet was first converted into the set of words contained in it (*tokenization*). Then, tokens providing little or no useful information to the text analysis, such as articles, conjunctions, prepositions, pronouns, were eliminated (*Stop-word filtering*). The remaining tokens were reduced to their *stems*, or root forms by removing suffixes, in order to group words having closely related semantics (*Stemming*). Then, stems not relevant for the analysis were eliminated (*Stem filtering*). The set of relevant stems were identified during the supervised learning stage (see below).

Eventually, for each tweet a corresponding vector of F numeric features was built (*Feature representation*). A numeric value was assigned to each feature, corresponding to a weight based on the importance of the stem in the training dataset and the frequency of the stem in the tweet. Indeed, we adopted the TF-IDF method[20] to determine the weights of each relevant stem which describes each tweet.

Supervised learning stage and classification model accuracy

In order to identify the set of relevant stems, the set of F numeric features and the parameters of the machine learning classification models, a supervised learning stage is needed. During this stage, a training set of labeled tweets must be used. In this work, we randomly selected and manually labeled 693 training tweets, consisting of 219 tweets *against vaccination*, 255 tweets *in favor of vaccination*, and 219 *neutral* tweets. Tweets of category *against vaccination* are those expressing a negative opinion about vaccination. Tweets of category *in favor of vaccination* are those expressing a positive opinion about vaccination. Tweets of category *neutral* may include news tweets about vaccines, neutral opinion tweets, and off-topic tweets containing the keywords selected (e.g., tweets related to the vaccination of pets). Tweets against and in favor of vaccines were considered subjective tweets. In Table 1, we show some examples of the extracted tweets of the training set and the corresponding identified labels. In Figure 1 we present a *word cloud* representation of the most common word in the training dataset for the tweets against or in favor of vaccinations. Word clouds were obtained using an online representation tool (wordart.com) and an automatic English translation service (Google Translate).

Several machine learning classification models (including also deep-learning models) were trained and compared by using a 10-fold cross validation analysis. The best performing

Table 1. Examples of tweets included in the training set.

Text of tweet – [English translation]	Classification label
"#NoVaccini #LibertaDiScelta. Un fondo per i danni da vaccini" – ["#NoVaccines #FreedomOfChoice. A fund for vaccine drawbacks"]	*Against*
"Ci ammalavamo una volta e ottenevamo l'immunita. Altro che vaccino! La libertà ai tempi del morbillo" – ["We got sick once and got immunity. We do not need vaccines! Freedom at the time of measles"]	*Against*
"Esiste una relazione chiarissima tra vaccini e l'autismo. Più vaccini, più i bambini sviluppano l'autismo, oltre ad altre malattie!" – ["There is a very clear relationship between vaccines and autism. The more vaccines, the more children develop autism, in addition to other diseases"]	*Against*
"Non vaccinare i propri figli è come circolare con un auto senza freni: un pericolo per tutti" – ["Not vaccinating your children is like traveling with a car without brakes: a danger for everyone"]	*In favor*
"I vaccini hanno superato tutti i test di efficacia e sicurezza. Non lasciamoci insinuare paure ingiustificate" – ["Vaccines have passed all the efficacy and safety tests. Let us not allow unjustified fears"]	*In favor*
"Mi raccomando non vaccinate i vostri figli, cosi potranno morire di morbillo!" – ["I recommend you do not vaccinate your children, so they can die of measles!"]	*In favor*
"Ma se fingessi di stare male dopo il vaccino per non andare a scuola??" – ["But what if I pretended to be sick after the vaccine for not going to school?"]	*Neutral*
"Altri casi di meningite registrati oggi . Guardate io non sono razzista. Ma troppe coincidenze non possono essere nemmeno! #BastaImmigrati" – [Other cases of meningitis recorded today. Look, I'm not a racist. But too many coincidences can not even be! #StopImmigration]	*Neutral*
"In Sicilia vaccino gratuito contro la meningite per i giovani" – ["In Sicily free vaccine against meningitis for young people"]	*Neutral*
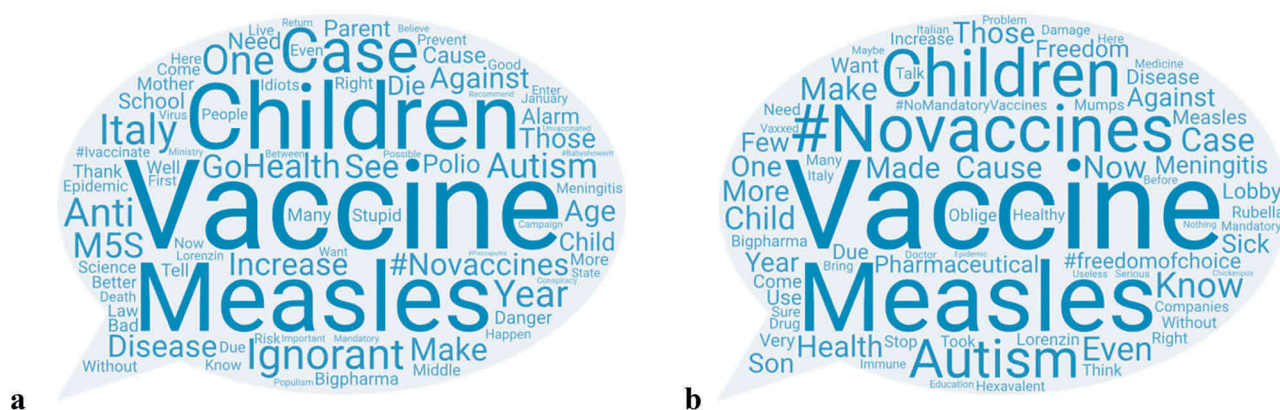
Figure 1. Word cloud representation of tweets in the training dataset by class (A. in favor, B. against).

models were based on the Support Vector Machine (SVM) classifiers.[21] Specifically, the selected model takes as input a text as a BOW with 2000 features and is characterized by an average accuracy (i.e. the number of tweets correctly labeled over the total number of tweets) of 64.8%. All the experiments were carried out using the Weka (Waikato Environment for Knowledge Analysis) Toolkit and its JAVA APIs.[22]

Additional details on the methods and on the achieved results can be found in a recent work[18] published by some of the authors. In particular, all the technical specifications regarding text representation and classification are discussed, including the complete statistical procedures for comparing the different machine learning-based classification models. The selected model was finally trained using the overall training set and employed for classifying all the collected tweets in three classes. We recall that the tweets analyzed during the monitoring campaign are represented using the BOW scheme with TF-IDF, considering a feature space formed by the 2000 relevant stems identified during the supervised learning stage.

In order to evaluate the generalization capability of the adopted classification system on future tweets, before the classification stage, for each event, we randomly read several tweets. Among them, we manually labeled around 60 tweets for each event, trying to identify 20 tweets for each class. Then, we automatically classified all the tweets of the event

and we used the labeled tweets to calculate the respective accuracy.

Data analysis

An analysis of the temporal distribution of tweets and trends by classes was then performed. Statistical analyses were performed with R statistical package (v3.6.1, R Statistical Foundation, Vienna, Austria), with the help of *decompose* function to separate time series (daily rates of tweet categories) into long-term trend, seasonal (weekly) fluctuations and random component. Univariate and multivariate linear regression models were built. The significance level was set at 0.05.

We checked how a set of pre-selected vaccine-related events influenced the number and distribution of tweets classes. In addition, peaks in number of tweets were assessed for correlation with additional vaccine-related events. Peaks of daily tweets were detected with a sampling algorithm, selecting the days with the highest daily vaccine-related tweet numerosity within a specified timeframe (10 days before and after). Significance of peaks was confirmed comparing the average daily tweet count during the peak with the average during the 10 days before the peak. Wilcoxon rank sum test was used to compare means.

Sentiment analysis around events and peaks was performed comparing Twitter data observed on days 0 to +4 ("peak") to the 5 days before the peak (days −5 to −1, "baseline"); in addition, comparison of days +5 to +9 ("aftermath") to the baseline was performed. 2-sample test for equality of proportions was used to compare rates; when applicable, p-values were adjusted with Bonferroni correction for multiple comparisons.

Results

We identified a total of 180,620 vaccine-related tweets during the period September 2016 – August 2017. A selection of analyzed tweets is presented in Table 1. The total number of tweets varies across the period from less than 50 to more than 3,500 per day.

Trend analysis

During the study period, the number of tweets showed an increasing trend ($p < .001$, $β = +2.42$ [SE: 0.21], $R^2 = 0.27$, linear model), peaking in the month of July 2017 (Figure 2). The day with the highest number of tweets during the study period was July 28th, 2017, with more than 3,500 tweets.

According to the overall analysis by category, 60% of tweets were classified as neutral, 23% against vaccination, and 17% in favor of vaccination. When considering the distribution over time, the rate of neutral tweets in total daily tweets ("neutrality rate") showed a decrease over time, with average rate of 75.0% (SD 8.5) in the first semester (monthly means between 68.2% and 80.6%) and average rate of 58.1% (SD 9.5) in the second semester (monthly means between 51.2% and 61.0%; Figure 3). Linear regression model on time series *trend* component for neutrality rate showed an

average decrease of 2.36% (SE 0.11) per month ($p < .001$, $R^2 = 0.55$). A multivariate model adjusted for tweet numerosity and rate of negative tweets produced analogous results (not shown, $R^2 = 0.59$). At the same time, the proportion of subjective tweets (e.g. non-neutral) showed a steady increase, indicating a progressive polarization of the opinions on vaccination. Tweets expressing opinions against vaccination became predominant over those in favor in the period April–August 2017, with a peak in July 2017 (Figure 3) for "negativity rate" (defined as the rate of negative tweets in non-neutral ones). Linear model on time series *trend* component for negativity rate showed an average increase of 0.27% (SE 0.08) per month ($p = .0012$, $R^2 = 0.03$), which was confirmed in a multivariate model adjusted for tweet numerosity and neutrality rate (not shown, $R^2 = 0.11$).

Effect of single events

The analysis by event was performed on a set of pre-selected events and is presented in Figure 4. The first pre-specified event considered, the publication of the PNPV 2017–19 on the 19th of January 2017, did not produce a significant effect in the social discussion, and no peak was detected in correspondence of the event (Wilcoxon test, $p = .40$). On the contrary, the approval, on January 26th, 2017, of the Agreement between Italian Health Minister and Italian Regions about vaccinations requirement, shortly following the publication of PNPV 2017–2019, corresponded to a peak in tweet count (+282% vs. baseline, $p = .03$). The spike was associated with a marked decrease in tweet neutrality rate, lasting for the following 10 days (baseline: 0.80; peak: 0.54; aftermath: 0.72, $p < .001$ overall), with no significant change in negativity rate. The preliminary approval of the Legislative Decree n. 73, introducing the obligation for 12 vaccinations
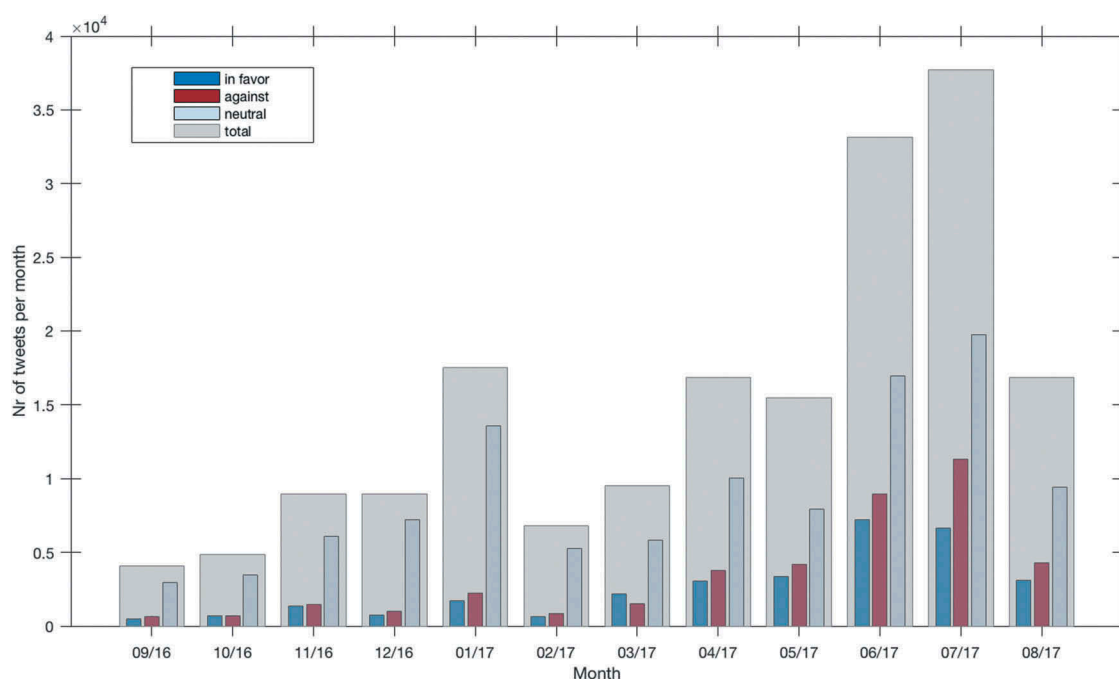
Figure 2. Number of tweets per month, total and by class (in favor, against, neutral), September 2016 – August 2017.
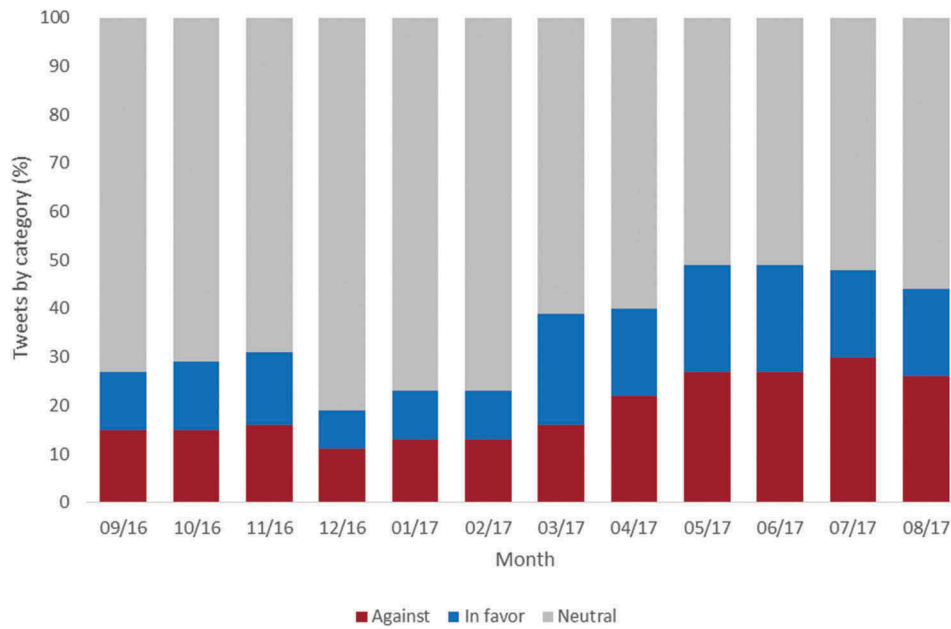
Figure 3. Proportion of tweets by category (in favor, against, neutral) by month, September 2016 – August 2017.
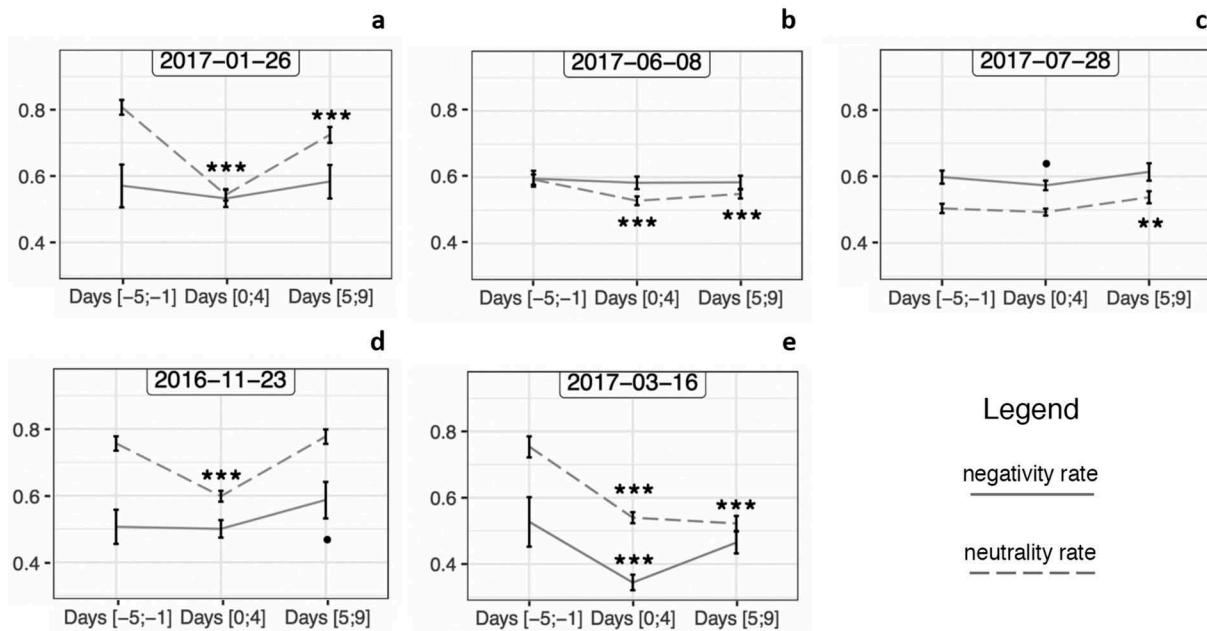

Figure 4. Analysis of neutrality and negative rates on vaccine-related events.
Panel a. Publication of the National Plan for the Vaccine Prevention (PNPV) 2017–19 and agreement with Italian Regions for a vaccination-enforcing law (January 26th, 2017); Panel b. Approval of the Legislative Decree n. 73 introducing 12 compulsory vaccinations (June 7th, 2017); Panel c. Approval in the Italian Chamber of Deputies of the Vaccines Decree (July 28th, 2017); Panel d. Approval of the law establishing vaccination requirements for school children in Emilia Romagna Region (November 22nd, 2016); Panel e. News about the increase of 230% cases of measles in Italy (March 16th, 2017). A two-sample test for equality of proportions, with Bonferroni correction for multiple comparisons, was performed. Adjusted p-value significance is shown (• <0.10, * <0.05, ** <0.01, *** <0.001). Comparisons are made with baseline (days −5 to −1). Error bars show 95% binomial confidence intervals for proportions.

(Vaccines Decree) on June 7th, 2017 produced a prominent effect in the social discussion in terms of number of tweets (+98.3% vs. baseline, $p = .014$), with an increase of subjective tweets about vaccination (baseline: 0.41, peak: 0.48, aftermath: 0.46, $p < .001$ overall), but no effect on negativity rate (Figure 4b). The ratification of the Vaccines Decree by the Italian Chamber of Deputies on July 28th, 2017 resulted in the highest spike in the number of tweets (max tweet count 3662

on July 28th, +130% vs. baseline, $p = .03$), with moderate effects on neutrality and negative rates (Figure 4c).

An analysis of the distribution of the tweets over time identified two further major spikes during the study period. A review of the major media outlets identified the corresponding vaccine-related events: 1) the approval of the law establishing vaccination requirements for school children in Emilia Romagna Region, on November 22nd, 2016 (tweet

count +603%, $p = .014$); 2) the diffusion on March 16th, 2017 of the data on measles epidemic, reporting an increase of 230% cases compared with the previous year (tweet count +339%, $p = .007$). In the first case, the event determined a marked polarization of opinion and a tendency in the following 10 days toward an increase of negative tweets ($p < .10$, Figure 4d). In the second case, the opinion polarity was in favor of vaccination immediately after the event, but the negativity rate returned to basal condition in following days (negativity rate: baseline 0.53, peak 0.34 [$p < .001$], aftermath 0.46 [$p = .29$ vs. baseline]) (Figure 4e). Quality check performed on tweet classification of the five aforementioned events lead to an average accuracy of 62.1% on the selected and labeled tweets (a test set of around 300 tweets, see Table 2).

A qualitative analysis of Word cloud representations of the training datasets highlighted a higher occurrence of hashtags (in particular #novaccines) and of the world autism in the tweets against vaccination than in the ones in favor of vaccination. In the tweets in favor of vaccinations, instead, we found a higher occurrence of insults to anti-vaccination activists and references to the political world. In both the datasets the main vaccine-preventable disease discussed was measles.

Discussion

To our knowledge, our study represents the first attempt to use Twitter as a monitoring system to gauge public opinion propensity toward vaccination in the Italian context. Similar approaches have been already applied, especially to understand HPV vaccination acceptance and the variation of public opinion in presence of outbreaks,[23–26] but we have not found examples of this sort of analysis to monitor public opinion during vaccination policy changes. We believe this analysis is important in the context of a progressive politicization of the vaccination topic, as seen during 2016 American election.[27]

Our study is the result of a multi-sectorial approach, applying text mining and machine-learning techniques to tweets' opinion mining in the frame of a substantial public health issue such as vaccine hesitancy.[28–30] The obtained monitoring tool had accuracy performance in line with another recent work on Twitter opinion mining.[31]

In particular, Twitter proved useful as a sentinel tool to monitor: a) the interest of the public on vaccinations by observing the trends of numbers of tweets on the topic; b) the polarization of public opinion observing the variations of the percentage of tweets against or in favor of vaccination; c) to monitor the effect of selected or unselected vaccine-related events on the polarization of public opinion.

Table 2. Accuracy of the monitoring tool for single events.

Event*	Accuracy (%)
A	61.9
B	61.6
C	62.4
D	62.1
E	64.7
Average Accuracy	62.1

*Letters refer to the same events represented in Figure 4

According to our findings, vaccination, as a topic, has received growing attention in the social media in Italy between September 2016 and August 2017. While this trend has been steady over the study period, a number of spikes have been identified, in correspondence with the occurrence of vaccine-related events. These data suggest that the number of people talking about vaccination increased, as a consequence of vaccine-related events occurred during the year, which have attracted the interest of people and media. Other analyses of Italian media outlets on vaccination in this period confirm this trend.[4] From a qualitative analysis of the contents of the training datasets, we found that measles polarized the attention of Twitter users, while other VPD were scarcely mentioned.

Yet our analysis showed a growing polarization of the public opinion on vaccination. While overall the majority of identified tweets were neutral toward vaccination, the proportion of subjective tweets increased over time. The relative share of positive and negative tweets varied during the period and it appears to be influenced by the occurrence of vaccine-related events and the publication of data and relevant information on vaccine-preventable diseases. For example, the release of the epidemiological data on measles cases in Italy was associated with an upsurge of pro-vaccination tweets. This phenomenon has already been described in other outbreaks or cases of fatal vaccine-preventable diseases;[32,33] the endorsement of the Vaccine Decree with the introduction of mandatory vaccination, instead, generated the highest peak of tweets about vaccination. The publication of the PNPV, considered one of the most modern and updated immunization schedules on the European scene,[34] failed to gain the attention of the public, highlighting the difficulty to effectively communicate an innovative health policy in Italy.

Still, according to our findings, the share of tweets against vaccination showed an increasing trend during the study period, superseding the quota of pro-vaccination tweets. This observation is particularly concerning, even more so as it coincided with reports of an expanding volume of web material classifiable as negative toward vaccination.[5,6] This situation is in accordance with other national and international surveys on VH that found that Italy is ranked among the WHO European Region countries with the highest levels of skepticism related to the importance, effectiveness and safety of vaccinations.[30,35]

Despite this situation, since 2016 an increase in vaccine coverage rates, especially for measles,[36] has been detected, even before the introduction of mandatory immunizations. We believe that the increase of public debate on vaccinations and the diffusion of data on the ongoing measles epidemic have already had a positive effect on vaccine perception. The introduction of mandatory vaccinations, despite being generally not well accepted by public, further consolidated this trend leading to an increase in polio and measles vaccines uptake.[37]

Our study has some limitations. Despite the popularity of Twitter, its users are a selected population and may not be representative of the Italian general population. The identification of tweets may have been incomplete, for example, due to lack of inclusion of additional relevant keywords, which may have skewed the distribution of subjective and neutral tweets. The

classification of the tweets may have been subject to errors due to the ambiguity of some entries and to the unavoidably limited accuracy of the model used. In particular, opinion mining is considered a challenging topic with respect to other text mining applications. In fact, whereas humans can easily detect irony or sarcasm in a text, automatic irony detection is a challenging task, given that the presence of irony may completely reverse the text polarity.[38] Ambiguous tweets, i.e., those containing discording opinions, are more challenging to classify, as in this case even humans may not able to decide for the correct category label. In addition, we may have missed relevant fluctuations in the public opinion in correspondence of vaccine-related events we were not aware of or we failed to identify through our analyses. We did not explore possible variations in the distribution of tweets categories by different vaccine products or target populations (e.g. children, adults). The analysis was meant to be an example of a prospective monitoring tool. This approach could prove a challenging task for AI-based monitoring systems and lead to overestimation of monitored phenomena, as happened to Google Flu Trends.[39] Finally, our study period ended shortly after the endorsement of the Vaccine Decree and we failed to monitor longer-term effects of this policy on the public opinion.

In conclusion, opinion mining analysis based on Twitter may be a useful and timely tool to assess the orientation of public opinion toward vaccination, as well as other public health interventions. The information derived from this analysis can complement traditional surveys (e.g. State of Vaccine Confidence initiative[40]) potentially allowing a more prompt response to emerging concerns and inform public health initiatives. This approach may be particularly beneficial when implemented in correspondence of key events, such as the adoption of a new health policy (e.g. Vaccine Decree), as a sentinel system to rapidly gather signals from the public. Therefore, opinion mining may become a useful tool for public health institutions and may effectively contribute to the development of appropriate communication and information strategies.

Disclosure of potential conflicts of interest

Funding

ORCID

Lara Tavoschi 🔟 http://orcid.org/0000-0002-2217-4664
Filippo Quattrone 🔟 http://orcid.org/0000-0002-7721-9559
Pietro Ducange 🔟 http://orcid.org/0000-0003-4510-1350
Marco Vabanesi 🔟 http://orcid.org/0000-0002-8283-7079
Francesco Marcelloni 🔟 http://orcid.org/0000-0002-5895-876X
Pier Luigi Lopalco 🔟 http://orcid.org/0000-0002-3181-8054

References

1. MacDonald NE. Vaccine hesitancy: definition, scope and determinants. Vaccine. 2015;33(34):4161–64. doi:10.1016/J.VACCINE.2015.04.036.

2. Stahl J-P, Cohen R, Denis F, Gaudelus J, Martinot A, Lery T, Lepetit H. The impact of the web and social networks on vaccination. New challenges and opportunities offered to fight against vaccine hesitancy. Médecine Mal Infect. 2016;46(3):117–22. doi:10.1016/J.MEDMAL.2016.02.002.

3. Signorelli C, Odone A, Cella P, Iannazzo S, d'Ancona F, Guerra R. Infant immunization coverage in Italy (2000–2016). Ann Ist Super Sanità. 2017. doi:10.4415/ANN_17_03_09.

4. Odone A, Tramutola V, Morgado M, Signorelli C. Immunization and media coverage in Italy: an eleven-year analysis (2007–17). Hum Vaccin Immunother. July 2018;1–4. doi:10.1080/21645515.2018.1486156.

5. Aquino F, Donzelli G, De Franco E, Privitera G, Lopalco PL, Carducci A. The web and public confidence in MMR vaccination in Italy. Vaccine. 2017;35(35):4494–98. doi:10.1016/j.vaccine.2017.07.029.

6. Donzelli G, Palomba G, Federigi I, Aquino F, Cioni L, Verani M, Carducci A, Lopalco P. Misinformation on vaccination: a quantitative analysis of YouTube videos. Hum Vaccines Immunother. 2018;14(7):1654–59. doi:10.1080/21645515.2018.1454572.

7. National Integrated Measles-Rubella Surveillance System. *Measles in Italy: weekly* Bulletin. *Week: 4–10 December 2017 (W49)*. Rome; 2017. http://www.epicentro.iss.it/problemi/morbillo/bollettino/Measles_WeeklyReport_N35eng.pdf.

8. Ministero della Salute. Piano Nazionale Prevenzione Vaccinale PNPV 2016–2018. 2017. [Accessed 2018 July 13]. http://www.salute.gov.it/imgs/C_17_pubblicazioni_2571_allegato.pdf.

9. 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(1):7–33. doi:10.1111/1468-0009.12038.

10. Klein AZ, Sarker A, Cai H, Weissenbacher D, Gonzalez-Hernandez G. Social media mining for birth defects research: a rule-based, bootstrapping approach to collecting data for rare health-related events on Twitter. J Biomed Inform. 2018 October;87:68–78. doi:10.1016/j.jbi.2018.10.001.

11. Lohmann S, White BX, Zuo Z, Chan MPS, Morales A, Li B, Zhai C, Albarracín D. HIV messaging on Twitter: an analysis of current practice and data-driven recommendations. AIDS. 2018 October;32:2799–805. doi:10.1097/QAD.0000000000002018.

12. Wakamiya S, Kawai Y, Aramaki E. Twitter-based influenza detection after flu peak via Tweets with indirect information: text mining study. JMIR Public Heal Surveill. 2018;4(3):e65. doi:10.2196/publichealth.8627.

13. Talib R, Hanif MK, Ayesha S, Fatima F. Text mining: techniques, applications and issues. Vol. 7. 2016. Accessed 2018 Oct 8. www.ijacsa.thesai.org.

14. Liu B. Sentiment Analysis. Cambridge: Cambridge University Press; 2015. doi:10.1017/CBO9781139084789.

15. Hailong Z, Wenyan G, Bo J. Machine learning and lexicon based methods for sentiment classification: a survey. 2014 11th web information system and application conference; 2014; Tianjin. IEEE; 2014. p. 262–65. doi:10.1109/WISA.2014.55.

16. Taboada M, Brooke J, Tofiloski M, Voll K, Stede M. Lexicon-based methods for sentiment analysis. Comput Linguist. 2011;37(2):267–307. doi:10.1162/COLI_a_00049.

17. Agarwal B, Mittal N, editors. Machine learning approach for sentiment analysis. In: Prominent feature extraction for sentiment analysis. Cham: Springer; 2016. p. 21–45.

18. D'Andrea E, Ducange P, Bechini A, Renda A, Marcelloni F. Monitoring the public opinion about the vaccination topic from tweets analysis. Expert Syst Appl. 2019;116:209–26. doi:10.1016/j.eswa.2018.09.009.

19. Zhang Y, Jin R, Zhou Z-H. Understanding bag-of-words model: a statistical framework. Int J Mach Learn Cybern. 2010;1(1–4):43–52. doi:10.1007/s13042-010-0001-0.

20. Wu HC, Luk RWP, Wong KF, Kwok KL. Interpreting TF-IDF term weights as making relevance decisions. ACM Trans Inf Syst. 2008;26:3. doi:10.1145/1361684.1361686.

21. Platt JC 12 fast training of support vector machines using sequential minimal optimization. *Adv Kernel Methods*. 1999. doi:10.1046/j.1469-1809.1999.6320101.x

22. Witten IH, Frank E, Hall MA, Pal CJ. Data mining: practical machine learning tools and techniques. Burlington (MA): Morgan Kaufmann; 2016.

23. Shapiro GK, Surian D, Dunn AG, Perry R, Kelaher M. Comparing human papillomavirus vaccine concerns on Twitter: a cross-sectional study of users in Australia, Canada and the UK. BMJ Open. 2017;7(10):e016869. doi:10.1136/bmjopen-2017-016869.

24. Keim-Malpass J, Mitchell EM, Sun E, Kennedy C. Using Twitter to understand public perceptions regarding the #HPV vaccine: opportunities for public health nurses to engage in social marketing. Public Health Nurs. 2017;34(4):316–23. doi:10.1111/phn.12318.

25. Du J, Xu J, Song H-Y, Tao C. Leveraging machine learning-based approaches to assess human papillomavirus vaccination sentiment trends with Twitter data. BMC Med Inform Decis Mak. 2017;17 (S2):69. doi:10.1186/s12911-017-0469-6.

26. Luo X, Zimet G, Shah S. A natural language processing framework to analyse the opinions on HPV vaccination reflected in twitter over 10 years (2008–2017). Hum Vaccines Immunother. 2019;15(7–8):1496–504. doi:10.1080/21645515.2019.1627821.

27. Dredze M, Wood-Doughty Z, Quinn SC, Broniatowski DA. Vaccine opponents' use of Twitter during the 2016 US presidential election: implications for practice and policy. Vaccine. 2017;35 (36):4670–72. doi:10.1016/j.vaccine.2017.06.066.

28. Vrdelja M, Kraigher A, Vercic D, Kropivnik S. The growing vaccine hesitancy: exploring the influence of the internet. Eur J Public Health. 2018;28(5):934–39. doi:10.1093/eurpub/cky114.

29. Salmon DA, Dudley MZ, Glanz JM, Omer SB. Vaccine hesitancy: causes, consequences, and a call to action. Vaccine. 2015;33(Suppl 4):D66–71. doi:10.1016/j.vaccine.2015.09.035.

30. Giambi C, Fabiani M, D'Ancona F, Ferrara L, Fiacchini D, Gallo T, Martinelli D, Pascucci MG, Prato R, Filia A, et al. Parental vaccine hesitancy in Italy – results from a national survey. Vaccine. 2018;36 (6):779–87. doi:10.1016/j.vaccine.2017.12.074.

31. Dey K, Shrivastava R, Kaushik S. Topical stance detection for twitter: A two-phase LSTM model using attention. In: Pasi G, Piwowarski B, Azzopardi L, Hanbury A, editors. Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics). Vol. 10772 LNCS. Cham: Springer Verlag; 2018. p. 529–36. doi:10.1007/978-3-319-76941-7_40.

32. Deiner MS, Fathy C, Kim J, Niemeyer K, Ramirez D, Ackley SF, Liu F, Lietman TM, Porco TC. Facebook and Twitter vaccine sentiment in response to measles outbreaks. Health Informatics J. 2019;25(3):1116–32. doi:10.1177/1460458217740723.

33. Porat T, Garaizar P, Ferrero M, Jones H, Ashworth M, Vadillo MA. Content and source analysis of popular tweets following a recent case of diphtheria in Spain. Eur J Public Health. 2018 July. doi:10.1093/eurpub/cky144.

34. Signorelli C, Guerra R, Siliquini R, Ricciardi W. Italy's response to vaccine hesitancy: an innovative and cost effective national immunization plan based on scientific evidence. Vaccine. 2017;35 (33):4057–59. doi:10.1016/J.VACCINE.2017.06.011.

35. Larson HJ, de Figueiredo A, Xiahong Z, Schulz WS, Verger P, Johnston IG, Cook AR, Jones NS. The state of vaccine confidence 2016: global insights through a 67-country survey. EBioMedicine. 2016;12:295–301. doi:10.1016/J.EBIOM.2016.08.042.

36. Istituto Superiore di Sanità. Vaccinations in Italy. [Accessed 2018 Oct 10]. http://www.epicentro.iss.it/temi/vaccinazioni/dati_Ita.asp#morbillo.

37. Burioni R, Odone A, Signorelli C. Lessons from Italy's policy shift on immunization. Nature. 2018;555(7694):30–30. doi:10.1038/d41586-018-02267-9.

38. Giachanou A, Crestani F. Like it or not: a survey of Twitter sentiment analysis methods. ACM Comput Surv. 2016;49 (2):1–28. doi:10.1145/2966278.

39. Lazer D, Kennedy R, King G, Vespignani A. The parable of Google flu: traps in big data analysis. Science (80-). 2014;343 (6176):1203–05. doi:10.1126/science.1248506.

40. Larson HJ, Schulz WS, Tucker JD, Smith DMD. Measuring vaccine confidence: introducing a global Vaccine Confidence Index. PLoS Curr. 2015;7(OUTBREAKS). doi:10.1371/currents.outbreaks.ce0f6177bc97332602a8e3fe7d7f7cc4.