Vaccine 41 (2023) 3196–3203

Contents lists available at ScienceDirect

Vaccine

journal homepage: www.elsevier.com/locate/vaccine

Public discourse and debate about vaccines in the midst of the covid-19 pandemic: A qualitative content analysis of Twitter

Enrique Prada, Andrea Langbecker *, Daniel Catalan-Matamoros

UC3M Medialab, Department of Communication and Media Studies, Madrid University Carlos III, Spain

A R T I C L E I N F O

Article history:
Received 8 August 2022
Received in revised form 2 March 2023
Accepted 30 March 2023
Available online 4 April 2023

Keywords:
Covid-19
Social media
Vaccination
Discourse
Misinformation
Twitter

A B S T R A C T

Objectives: Characterize the public debate and discourse about vaccines during the covid-19 vaccination programmes.
Methods: We performed a manual content analysis of a sample of English-written Twitter posts that included the word vaccine and its derivatives. We categorized 7 variables pertaining to the content of the posts, and classified the type of user that published the post and the number of retweets. Then, the patterns of association between these variables were further explored.
Results: Among the tweets with negative tone towards vaccines, 33% display negationist discourses, 29% protest or defiance discourses, 13% discuss the pandemic management measures and yet another 13% of these tweets display a scientific discourse. Research results, vaccination data and practical information are more associated to positive tone towards vaccines, while news relate to neutral tone. The users that received more retweets were media accounts and journalists, followed by government accounts and scientific organizations related to the government. Tweets displaying preventive messages received more retweets in average. The discourses most associated with objective information are the preventive, institutional, medical-scientific, and those about the different measures to manage the pandemic. On the other hand, the most subjective tweets are those with negationist, antinegationist and protest discourses.
Conclusions: Although there is a non-negligible proportion of tweets that are directly opposed to vaccines, also an important part of vaccine-negative content takes the form of protest discourses, criticisms towards government actions as well as towards the measures to tackle the pandemic. Therefore, negative discourses during the pandemic included serious vaccine hesitancy cases. Moreover, they were not only fuelled by distrust in science, but also and very importantly they were connected to dissatisfaction towards the public management of the pandemic.

1. Introduction

The COVID-19 pandemic that originated in 2020 has had important economic, cultural, and political ramifications. More specifically, the spread of the virus amplified a debate about vaccines that was already taking place before the emergence of this disease [1] and has put the issue of vaccine hesitancy at the forefront of public debates. As Lentzen et al. [2] stated, "a fairly new and still partially known pandemic, a rapid vaccine development, a short follow up of clinical studies and vaccine side effects make people visualize more risks than benefits" (p. 44). On the other hand, since the development of the first COVID-19 vaccines, vaccination campaigns have been fairly successful, as a significant portion of the population have agreed to take part, although vaccination rates are still relatively small, especially in low-income countries [3].

This pattern reflects the ongoing debate about vaccines, in which defenders of the benefits of vaccines as well as its critics have expressed varied views and have put forth different arguments and discourses about the COVID-19 but also about governments' strategies to tackle the pandemic and vaccines themselves. These debates have taken place in a context where an important part of human communication takes place on the internet and, more specifically, on social media platforms, which have become increasingly important tools to disseminate health-related information [2]. Moreover, social media platforms also have the potential to be used for medical education purposes [4] and individuals' attitudes can be influenced by the continued interactions and exposure to information that takes place there [5], especially vaccine hesitant individuals, that is, those that have

* Corresponding author at: Department of Communication and Media Studies, Madrid University Carlos III, Calle Madrid, 133. 28903 Getafe, Madrid, Spain.
E-mail address: alangbec@hum.uc3m.es (A. Langbecker).

https://doi.org/10.1016/j.vaccine.2023.03.068
0264-410X/© 2023 The Author(s). Published by Elsevier Ltd.

some degree of reluctance to be vaccinated and even if they have access to vaccines, are slow to accept them or challenge them [6].

This exploratory study aims to understand the different features and characteristics of vaccine-related content generated in Twitter during the pandemic. Although this social media platform is not the most used [7], it has several characteristics that make it a relevant case study. Twitter is a micro-blogging service where users can write short messages of 280 characters maximum called "tweets" or repost other users' tweets ("retweet"). This social network is characterized by the immediacy of communication [8] and also it is an important arena for public debates: public figures (e.g., politicians) and institutions are very active on this platform, and the majority of its users regularly get news there [9]. Moreover, its short texts format and the attitudinal polarization that tends to take place there make this platform fertile ground for the spread of misleading information questioning vaccines [10].

Precisely, the abovementioned properties of this social media platform have led authors to also investigate the public debate about COVID-19 of Twitter. However, the previous literature has tended to focus on specific elements of the debate (e.g., tone about vaccines, or content about vaccine side effects, etc.) [11,12,13,14,15] but has not yet explored how the different features and dimensions of the public discourse about vaccines are related to each other. Precisely, the main goal of our study is to identify the different characteristics of vaccine-related discourses and explore its patterns of association. In order to do that, we have performed a manual content analysis, that allows us to categorize tweets according to 8 different variables of interest. We have also considered the number of retweets received by each tweet, as a measure of the tweet's engagement. We have done this for a sample of English-written tweets published in a 1-year period which coincides with the beginning and middle phases of vaccination campaigns in English-speaking countries. Therefore, our aim is to study vaccine-related discourses in this particular period. Following an exploratory and descriptive approach, our main research questions are:

RQ1: What are the main characteristics of the vaccine-related discourses in Twitter during this period?
RQ2: What are the main types of discourses associated with positive and negative vaccine-related contents?
RQ3: Did the engagement of Twitter posts vary according to their discursive characteristics?

The rest of this paper is structured as follows. In the next section the data collection procedure, variables and statistical analyses are described. In Section 3, we present the results, and then discuss the main findings before concluding.

2. Materials and methods

2.1. Data collection

We conducted a qualitative content analysis performed by two researchers. A manual coding was used to provide a nuanced characterization of the Twitter content [16]. In order to get a sample of tweets, we run a query using the "advanced search" tool provided by Twitter. We downloaded tweets that included at least one of the following words: vaccine, vaccines, vaccinating, vaccinated, vaccinate, vaccinates, vaccination, immunization, immunizate. We retrieved tweets between May 2021 – April 2022 so we could focus on the COVID-19 vaccination campaigns among the English-speaking community. The data was extracted from the Twitter API using twarc2, a Python software tool.

We randomly selected 19 days inside this 1 year-period and we further randomly selected 500 tweets from these 19 days (after deleting the "retweets" and also 21 tweets that either did not refer to human vaccines, or that used the term vaccine in a metaphorical way, or that were not written in English).

In the manual content analysis, we focused on 8 variables (Table 1): type of user, type of content, whether the information displayed in the tweet is objective or subjective, tone about vaccines. We also coded the type of disease, whether the tweet mentions a pharmaceutical company, whether it cites scientific or non-scientific sources (or does not cite any source), the type of discourse. The categories of our Type of discourse variable have emerged from an inductive process in which statements with similar discursive characteristics were identified: we focused on the predominant discursive style of each tweet, and we delimited different categories so that they would capture different discursive formations, understood as discourses that share specific concerns, perspectives, concepts, or themes [17], and are related to specific fields of knowledge [18]. This allowed us to identify different discourses that could nonetheless share a similar tone about vaccines. This process generated 9 types of discourses, which are listed in Table 1. Finally, we also considered the number of retweets received by each tweet as a measure of the tweets' engagement.

A pretest to assess intercoder reliability was performed with a randomly selected 10 % of the sample [19] for the eight variables. We adopted Gwet's AC1 statistical coefficient of concordance

Table 1
Variables and categories from the analysis of English-written Twitter posts about vaccines.

Variable	Categories	Variable	Categories
Type of user	Anonymous	**Pharmaceutical company**	Mentions at least one company
	General Public		Unspecified
	Health professional	**Sources**	Cites a scientific source
	Media and journalists		Cites a non-scientific source
	Other		Does not provide source
Type of content	Personal opinion or experience	**Type of discourse**	Denialist
	Research results, vaccination data and practical information		Protest and defiance
	News		Political
	Other		Institutional
Type of information	Objective		Pandemic management measures
	Subjective		Medical/scientific
Tone	Positive		Preventive
	Neutral		Motivational
	Negative		Antidenialist
Disease	Covid-19		Other
	Other diseases		
	Unspecified / general		

Table 2
Results of the intercoder agreement test (Gwet's AC1).

User	Type of content	Type of information	Tone
0,827	0,651	0,889	0,794
Disease	**Pharmaceutical company**	**Sources**	**Type of discourse**
0,819	0,963	0,885	0,656

[20]. After the first intercoder agreement test, a further round of coding training was done for the variable 'Discourse type' because it scored below the threshold of 0.6. The results of the reliability scores between the evaluators were statistically significant, as can be seen in Table 2.

2.2. Statistical analysis

In order to assess to what extent our variables of interest are associated, we cross-tabulated all of them. We performed chi-squared tests to check if the variables are significantly associated ($p < 0.05$). In order to avoid inflated chi-squared values and therefore unreliable test results, we only selected those pairs of variables with less than 20 % of their cells with expected frequencies lower than 5 and no cell with expected frequency lower than 1 [21]. This yields a total of 12 comparisons out of 28 possible cross-tabulations.

Moreover, to assess which variables are more strongly associated, the Carmer's V statistic was computed. And, to explore which categories vary more, we display each cell's contribution to the chi-square value [22], that is, the squared standardized residual of each cell.

Finally, to explore if some characteristics of the tweets are associated with more circulation (via retweets) we tested whether there are differences in the number of retweets that each group of tweets received. They were grouped according to the categories of our variables. As the distribution of retweets is not normal in all the groups, instead of performing t-tests to compare the average number of retweets we performed Wilcoxon rank-sum tests to compare the distribution of retweets [21].

The present study is nested in the project "Multi-source and multi-method prediction to support COVID-19 policy decision-making", which was revised and approved by the Universidad Carlos III of Madrid Ethics Committee (code CEI22_05). In addition, the analysis plan has been published in AsPredicted under the code #100259.

3. Results

3.1. Data description

In the anglophone Twitter, based on our sample of 500 units, we can observe that most of the users talking about vaccines are either anonymous (37,6%) or general public (33,0%), followed by 7,6% of media and journalists and a rather small portion of health professionals (3,2%).

Most of the content are personal opinions and experiences (65,4%). Another 13,8% of tweets shared news, and 12,2% research results including data about vaccination (mainly about coverage) and practical information (facilities or slots). Finally, we found a low proportion of tweets exploring topics such as testimonies of vaccines' side effects (1,2%) or products/apps/events related to vaccines (1,4%). These tweets where collapsed into the category 'Others' along with other tweets not possible to be classified (8,6% of the sample). In line with the big proportion of tweets sharing personal opinions, we find that 79,4% of the content provided subjective information (e.g. experiences, anecdotes, rumours), while the rest contained objective information (e.g. data from research studies focusing on vaccine effectiveness and safety).

The proportion of neutral and positive content regarding vaccines is similar (41 % and 40,6%, respectively), while there are 18,4% of negative tweets. Most of the tweets talk about the COVID-19 (84,6%) although we also find some tweets that either display general messages about vaccines or don't specify which vaccine they refer to (13,2%), as well as some few mentions to other vaccine-preventable diseases (2,2%). Moreover, only 4,2% of the tweets mentioned at least one specific pharmaceutical company. Among the ones who did it, the most cited ones are Pfizer (2,2% of cases) and Moderna (2 %).

78,4% of the content didn't mention any sources of information, and only 3,29 % of tweets included a scientific source (e.g., scientific organizations, researchers, or academic journals), while 18,3% included a non-scientific source (mainly webpages of newspapers or TV channels).

We identified different types of discourses, some more prevalent than others. The most used ones have to do with the measures to manage the COVID-19 pandemic (20,4%), followed by other discourses such as preventive (12,6%), medical/scientific (10 %) and of protest and defiance (10 %).

3.2. Bivariate relationships

In Table 3 we can explore what pairs of variables are associated in a statistically significant way (p-value < 0,05) as well as the strength of the relationship between variables (Cramer's V column).

It is remarkable to note that the type of information conveyed in the tweet is the variable that changes more depending on the other variables. Also, the tone towards vaccines varied relatively less than the other variables. In four out of six cases the tone did not vary depending on the author of the tweet, on the type of information, whether the tweet mentioned a pharmaceutical company or whether it cited any source. However, the tone regarding vaccines is associated with the discourse employed by users and also to the content of tweets.

In relation on how the selected variables are specifically associated with each other, firstly, most of the tweets that include scientific sources provide objective information, but there are also some cases where most of the information of the tweet is subjective (Table 4). This also happens with non-scientific sources: tweets that include them mostly display objective information. On the other hand, tweets that don't include sources tend to be disproportionately subjective.

Secondly, personal opinions and experiences tend to be mostly subjective (Table 5). Tweets that discussed research results, vaccination data or that provided practical information for people who wish to get vaccinated also tend to have objective information (59 % of those tweets). But most of all, tweets that included news are the ones more likely to be objective, although, as it was mentioned before, it is a small proportion of the overall sample.

Turning now to the association between the discourse of tweets and their tone regarding vaccines, we find a fairly strong association between some type of discourses and some tones. If we focus on the column percentages of Table 6 (in the third row of each category) we see that 33 % of negative tweets display negationist (more extreme) discourses, while up to 29 % have protest or defiance discourses. Moreover, 13 % of the negative tweets focused on discussing the pandemic management measures and yet another 13 % of these tweets displayed a scientific discourse.

On the other hand, regarding the tweets that are positive towards vaccines, 28,1% have a preventive discourse highlighting the benefits of vaccines (the most common discourse among those

Table 3
Relationship between variables.

Variables		Chi-square	p-value	Cramer's V
Type of information	Sources	243,70	0,000 **	0,71
Type of information	Content	246,01	0,000 **	0,70
Tone	**Discourse**	396,72	0,000 **	0,63
Type of information	User	77,59	0,000 **	0,42
Type of information	Discourse	49,72	0,000 **	0,32
Type of information	Disease	13,33	0,001 **	0,16
Tone	**Content**	55,76	0,009 **	0,13
User	Disease	17,21	0,028 *	0,13
Tone	User	14,36	0,07	0,12
Tone	Type of information	5,27	0,072	0,10
Tone	Pharmaceutical company	4,26	0,119	0,09
Tone	Sources	5,62	0,229	0,08

** denotes that the relationship is statistically significant with a level of confidence of 99 % (p-value < 0.01).
* denotes that the relationship is statistically significant with a level of confidence of 95 % (p-value < 0.05).

Table 4
Tabulation of Source and Type of information.

Source	Type of information		
	Objective	Subjective	Total
Non-scientific source	66	23	89
	74.16	25.84	100.00
	67.35	5.93	18.31
	128.7+	**32.5**+	
Scientific source	12	4	16
	75.00	25.00	100.00
	12.24	1.03	3.29
	23.9+	**6.0**+	
No source	20	361	381
	5.25	94.75	100.00
	20.41	93.04	78.40
	42.0+	**10.6**+	
Total	98	388	486
	20.16	79.84	100.00
	100.00	100.00	100.00

Pearson Chi2 = 243.70 Prob = 0.0000.
First row shows *frequencies*; second row shows *row percentages* and third row shows *column percentages*.
+ The cell's contribution to the chi square is greater than 2.

Table 5
Tabulation of Content and Type of Information.

Content	Type of information		
	Objective	Subjective	Total
Personal opinion or experience	4	323	327
	1.22	98.78	100.00
	3.88	81.36	65.40
	59.6+	**15.5**+	
Research results, vaccination data and practical information	36	25	61
	59.02	40.98	100.00
	34.95	6.30	12.20
	43.7+	**11.3**+	
News	50	19	69
	72.46	27.54	100.00
	48.54	4.79	13.80
	90.1+	**23.4**+	
Other	13	30	43
	30.23	69.77	100.00
	12.62	7.56	8.60
	1.9	0.5	
Total	103	397	500
	20.60	79.40	100.00
	100.00	100.00	100.00

Pearson Chi2 = 246.01 Prob = 0.0000.
First row shows *frequencies*; second row shows *row percentages* and third row shows *column percentages*.
+ The cell's contribution to the chi square is greater than 2.

Table 6
Tabulation of Discourse and tone.

Discourse	Tone			
	Positive	Neutral	Negative	Total
Negationist	0	0	31	31
	0.00	0.00	100.00	100.00
	0.00	0.00	33.70	6.20
	12.6+	**12.7**+	**112.2**+	
Protest and defiance	4	19	27	50
	8.00	38.00	54.00	100.00
	1.97	9.27	29.35	10.00
	13.1+	0.1	**34.4**+	
Political	6	15	1	22
	27.27	68.18	4.55	100.00
	2.96	7.32	1.09	4.40
	1.0	**4.0**+	2.3	
Institutional	2	12	0	14
	14.29	85.71	0.00	100.00
	0.99	5.85	0.00	2.80
	2.4	**6.8**+	2.6	
Pandemic management measures	27	63	12	102
	26.47	61.76	11.76	100.00
	13.30	30.73	13.04	20.40
	5.0+	**10.7**+	2.4	
Medical/scientific	19	19	12	50
	38.00	38.00	24.00	100.00
	9.36	9.27	13.04	10.00
	0.1	0.1	0.9	
Preventive	57	6	0	63
	90.48	9.52	0.00	100.00
	28.08	2.93	0.00	12.60
	38.6+	**15.2**+	**11.6**+	
Motivational	38	2	0	40
	95.00	5.00	0.00	100.00
	18.72	0.98	0.00	8.00
	29.2+	**12.6**+	**7.4**+	
Antinegationist	27	3	0	30
	90.00	10.00	0.00	100.00
	13.30	1.46	0.00	6.00
	18.0+	**7.0**+	**5.5**+	
Other	23	66	9	98
	23.47	67.35	9.18	100.00
	11.33	32.20	9.78	19.60
	7.1+	**16.6**+	**4.5**+	
Total	203	205	92	500
	40.60	41.00	18.40	100.00
	100.00	100.00	100.00	100.00

Pearson Chi2 = 396.72 Prob = 0.0000.
First row shows *frequencies*; second row shows *row percentages* and third row shows *column percentages*.
+ The cell's contribution to the chi square is greater than 2.

who talk positively on vaccines), 18,7% try to motivate or encourage people to get vaccinated, while a non-negligible 13,3% are antinegationist, i.e., they overtly confront negationist people or people who don't want to get the vaccine.

As pointed out by each cells' contribution to the overall chi-square value (Table 7, values highlighted in bold), media and journalists, and health professionals are the users that most typically provide objective information, while anonymous users are the ones that provide less objective information.

In Table 8 we can see that the discourses most associated with objective information are the preventive, institutional, medical-scientific, and those about the different measures to manage the pandemic. On the other hand, the most subjective tweets are those with negationist, antinegationist and protest discourses.

Regarding the relationship between the type of information of a tweet and the disease that is mentioned in it, those tweets that don't specify any disease tend to provide objective information to a lesser extent than those that mention any disease (especially those that refer to COVID-19). Only 4,5% of the unspecific tweets had objective information, while up to 23,4% of the posts referring to COVID-19 provided objective information.

Regarding the relationship between content and tone of the tweets, research results, vaccination data and practical information are more associated to positive content towards vaccines, while news are more associated to neutral content, as shown by their cell's contribution to the overall chi-square value higher than 2 (Table 9).

Finally, different types of users differed slightly in the diseases that they tended to mention (Table 10). This is the less strong association that was statistically significant in our sample. Anonymous users didn't mention other specific diseases while the general public mentioned other diseases more frequently than the rest of the groups. Conversely, the media and journalists tended to specify more the disease they were talking about.

3.3. Which vaccine-related posts had more engagement?

We now turn to the analysis of which discursive features are associated with more engagement via retweets by means of Wilcoxon rank-sum tests. Retweets are important because they are the main tool for information circulation on Twitter and they are done primarily to show approval, support and to argue about topics [23]. Therefore, the number of retweets offers us an idea of the most supported ideas and arguments about vaccines, but also potentially the more controversial content.

First of all, the users that received more retweets in average were media accounts and journalists, followed by government accounts and government scientific organizations. Moreover, the median of retweets received by these three types of accounts were higher than those received by the anonymous or general public accounts, in a statistically significant way ($p < 0.01$ in all the Wilcoxon rank-sum tests of the pairwise comparisons and $p = 0.0143$ when comparing the retweets received by the general public and those received by government organizations). This shows that the media and government accounts are active, can effectively communicate and that there are asymmetries in the capabilities to disseminate information by different users. Also, interestingly, in our sample no health professional received retweets in their posts. This shows us that the content published by health professionals does not seem to interest general Twitter

Table 7
Tabulation of User and Type of information.

User	Type of information		
	Objective	Subjective	Total
Anonymous	19	169	188
	10.11	89.89	100.00
	18.45	42.57	37.60
	10.0+	**2.6**+	
General public	20	145	165
	12.12	87.88	100.00
	19.42	36.52	33.00
	5.8+	1.5	
Health professional	10	6	16
	62.50	37.50	100.00
	9.71	1.51	3.20
	13.6+	**3.5**+	
Media and journalists	21	17	38
	55.26	44.74	100.00
	20.39	4.28	7.60
	22.2+	**5.8**+	
Other	33	60	93
	35.48	64.52	100.00
	32.04	15.11	18.60
	10.0+	**2.6**+	
Total	103	397	500
	20.60	79.40	100.00
	100.00	100.00	100.00

Pearson Chi2 = 77.59 Prob = 0.0000.
First row shows *frequencies*; second row shows *row percentages* and third row shows *column percentages*.
+ The cells contribution to the chi square is greater than 2.

Table 8
Tabulation of Discourse and Type of information.

Discourse	Type of information		
	Objective	Subjective	Total
Negationist	1	30	31
	3.23	96.77	100.00
	0.97	7.56	6.20
	4.5+	1.2	
Protest and defiance	3	47	50
	6.00	94.00	100.00
	2.91	11.84	10.00
	5.2+	1.3	
Political	4	18	22
	18.18	81.82	100.00
	3.88	4.53	4.40
	0.1	0.0	
Institutional	5	9	14
	35.71	64.29	100.00
	4.85	2.27	2.80
	1.6	1.4	
Pandemic management measures	28	74	102
	27.45	72.55	100.00
	27.18	18.64	20.40
	2.3	0.6	
Medical/scientific	15	35	50
	30.00	70.00	100.00
	14.56	8.82	10.00
	2.1	0.6	
Preventive	27	36	63
	42.86	57.14	100.00
	26.21	9.07	12.60
	15.1+	**3.9**+	
Motivational	8	32	40
	20.00	80.00	100.00
	7.77	8.06	8.00
	0.0	0.0	
Antinegationist	1	29	30
	3.33	96.67	100.00
	0.97	7.30	6.00
	4.3+	1.1	
Other	11	87	98
	11.22	88.78	100.00
	10.68	21.91	19.60
	4.2+	1.1	
Total	103	397	500
	20.60	79.40	100.00
	100.00	100.00	100.00

Pearson Chi2 = 49.72 Prob = 0.0000.
First row shows *frequencies*; second row shows *row percentages* and third row shows *column percentages*.
+ The cells contribution to the chi square is greater than 2.

Table 9
Tabulation of Content and Tone.

Content	Tone			
	Positive	Neutral	Negative	Total
Personal opinion or experience	135	122	70	327
	41.28	37.31	21.41	100.00
	66.50	59.51	76.09	65.40
	0.0	1.1	1.6	
Research results, vaccination data and practical information	33	22	6	61
	54.10	36.07	9.84	100.00
	16.26	10.73	6.52	12.20
	2.7+	0.4	**2.4**+	
News	21	37	11	69
	30.43	53.62	15.94	100.00
	10.34	18.05	11.96	13.80
	1.8	**2.7**+	0.2	
Other	14	24	5	43
	32.56	55.81	11.63	100.00
	6.90	11.71	5.43	8.60
	0.7	**2.3**+	1.1	
Total	203	205	92	500
	40.60	41.00	18.40	100.00
	100.00	100.00	100.00	100.00

Pearson Chi2 = 16.98 Prob = 0.0093.
First row shows *frequencies*; second row shows *row percentages* and third row shows *column percentages*.
+ The cellś contribution to the chi square is greater than 2.

Table 10
Tabulation of User and Disease.

User	Disease			
	General	Other	Covid-19	Total
Anonymous	25	0	163	188
	13.30	0.00	86.70	100.0
	37.88	0.00	38.53	37.60
	0.0	**4.1**+	0.1	
General Public	20	7	138	165
	12.12	4.24	83.64	100.00
	30.30	63.64	32.62	33.00
	0.1	**3.1**+	0.0	
Health Professional	1	0	15	16
	6.25	0.00	93.75	100.00
	1.52	0.00	3.55	3.20
	0.6	0.4	0.2	
Media and journalists	1	1	36	38
	2.63	2.63	94.74	100.00
	1.52	9.09	8.51	7.6
	3.2+	0.0	0.5	
Other	19	3	71	93
	20.43	3.23	76.34	100.00
	28.79	27.27	16.78	18.60
	3.7+	0.4	0.7	
Total	66	11	423	500
	13.20	2.2	84.60	100.00
	100.00	100.00	100.00	100.00

Notes: Pearson Chi2 = 17.2098 Prob = 0.028.
+ The cellś contribution to the chi square is greater than 2. First row shows *frequencies*; second row shows *row percentages* and third row shows *column percentages*.

users and that these professionals could improve the communication potential to disseminate scientific content.

Secondly, although the positive content towards vaccines received slightly more retweets in average than neutral and especially negative content, the difference is not statistically significant (p = 0.849 and p = 0,610 respectively).

Regarding the type of discourse, tweets displaying preventive messages received more retweets in average although the difference with other discourse categories is not statistically significant. Regarding the type of content, news received more retweets than personal experiences (p = 0.0001). Regarding the type of information, tweets with objective information received more circulation in terms of retweets in average than those with subjective information (p = 0.0000). Finally, tweets that included non-scientific sources received more retweets than those that did not mention any source (p = 0.0000), but this did not happen with the posts that included scientific sources (p = 0.564).

4. Discussion

This paper set out to perform an exploratory content analysis of vaccine-related content in Twitter. Our goals were to characterize the different discursive elements associated to this kind of content in order to be able to understand these discourses in a more complete and nuanced way. Specifically, we wanted to explore the different types of discourses about vaccines. Although sometimes in our analysis the debate about vaccines has been simplified, we also identified that behind the prevalence of personal opinions in the discussion about vaccines, this debate was complex and comprised varied topics and discourses. Our focus on vaccine related content in general, beyond discourses about specific diseases or about specific elements (e.g., vaccination side effects) enables us to put in context some elements of the debate: For instance, we barely detect discussions about vaccine side effects. Given the potential of Twitter to amplify debates, the low proportion of tweets discussing this issue probably underscores the very low prevalence of serious vaccine side effects. Sv et al. [24], for example, found that a minority of tweets about COVID-19 vaccine side effects were negative.

Also, we see a proportion of denialist and anti-denialist messages, but these visible and radical posts are only a minor part of the overall discourse about vaccines. There are more negative posts questioning specific aspects about vaccines without denying in general their effectiveness. Our findings are somewhat similar to those found by Herrera-Peco et al [25], in the sense that they also detected a small proportion of more radical messages about COVID-19 vaccines (in their case, messages stating that vaccines would manipulate the human genetic code). On the other hand, most of the content that is positive towards vaccines in our analysis focuses on highlighting their advantages or on encouraging people to get vaccinated, instead of directly attacking negationists.

Contrary to what we would expect, the tone does not vary much depending on the rest of our variables of interest. For instance, tone and type of user. We would expect that negative messages about vaccines are disproportionally spread by anonymous users. However, there isn't a statistically significant relationship between these two variables, and neither there is a relationship between tone and type of information: both positive and negative messages about vaccines are mainly subjective. There is, however, a bigger association between the types of discourses used and the type of information conveyed by users: negationist and protest discourses tend to be disproportionally subjective, while preventive discourses comprise objective information more often.

For one, we can see that although there is a non-negligible proportion of tweets that are directly opposed to vaccines (i.e., expressing negationist discourses), also an important part of vaccine-negative content takes the form of protest discourses, criticisms towards government actions as well as towards the measures to tackle the pandemic. Therefore, we think that negative discourses during the pandemic included cases of serious vaccine hesitancy, not just plain denialism. Still, vaccine hesitancy constitutes a health challenge, both generally [26] and in the case of COVID-19 vaccination [27]. Moreover, negative discourses were not only fuelled by distrust in science, but also and very importantly they were connected to dissatisfaction towards the public management of the pandemic. The pattern found in our data supports the argument that vaccine hesitancy is shaped by state-society relations [28], both in its scope and in the forms it takes. We also consider that this is particularly the case because in the period that we cover vaccines were highly salient: not only because the pandemic, also because we cover the start and specially the middle-phases of the vaccination programmes in English-speaking countries.

Regarding the engagement received by tweets, it is remarkable to note that tweets coming from the media and journalists, with a preventive discourse, containing objective information and citing non-scientific sources tended to receive more retweets. This pattern contrasts with other descriptions of the type of vaccine-related content that gets more circulation in social media (i.e., misinformation/disinformation based on personal narratives). For instance, Das and Ahmed [29] consider that misinformation and disinformation have dominated social media since the advent of the pandemic, fuelling vaccine hesitancy [30]. Misinformation would be related to the dissemination of false content without the intention of causing harm, while disinformation would have the clear objective of circulating false content with the intention to cause harm to people or institutions. Although it is possible that the number of retweets is not only associated to support such messages but also to guide the debate and controversies regarding the content of those tweets [23]. Also, the lack of engagement of the content published by health professionals is somewhat in line with the results of Hsia and Kong [31], who found that the topics covered by doctors tend to differ from those expressed by the general public.

This study has a number of limitations that could be addressed in future analyses. Crucially, the sample size is limited, which impedes us to assess and explore over-time and geographical differences in vaccine-related discourses. For instance, Batra et al. [32] have found that COVID-19 vaccine-related attitudes and sentiment in Twitter were similar in neighbouring countries, while the reactions to the coronavirus outbreak differed more across neighbouring countries. This study can't contribute to this line of research. However, we think that our exploratory analysis can serve as a preliminary study of the characteristics of vaccine-related discourses on average in the period in which the tweets were selected. That is, in the period that coincides with the initial and middle phases of vaccination campaigns in English-speaking countries. In that sense, results are related to this specific period and not with the initial stage of the pandemic, nor with the current situation. Moreover, the sample size has also hampered the possibility of doing precise estimations in some cases, so the patterns of association with lower expected cell frequencies were not explored. However, given the randomization process to select the sample and that manual content analysis is time consuming, this is a reasonable sample compared with other similar studies. Second, this analysis could be conducted in other social networks too in order to find out some other nuances in the public discourse and debate of vaccines during the pandemic. As there are different profiles of users and different news consumption patterns depending on the social network [5], we might find different discursive dynamics in other platforms (e.g., Instagram, Facebook, TikTok). Despite these limitations, our study offers a useful analysis of public discourse and the characteristics of the content that circulates on social media concerning vaccines by using a qualitative method that includes manual coding. As these results enable us to improve our understanding of the pandemic's public debate on vaccines, public health services and other health authorities could use these findings in the planning of communication strategies and in the implementation of vaccination campaigns.

5. Conclusion

In sum, our exploratory content analysis highlights the existence of nuances in the debate related to vaccines during the pandemic. These nuances deserve further scrutiny by means of analyses with larger sample size. Most of the positive debate focuses on the vaccine advantages or on encouraging people to get vaccinated instead of speaking out against those who oppose vaccination. On the other hand, the negative debate mostly shows strong cases of vaccine hesitancy, not only based on distrust in science, but also on dissatisfaction with the vaccine-related management during the coronavirus pandemic by governments and public organizations. This complexity, in turn, calls for a more complex communicational approach to disseminate scientific evidence [33].

Funding

Study financed through the Agreement signed between the Community of Madrid (Ministry of Education, Universities and Science) and the Madrid University Carlos III for the direct grant of 2,126,000 € to fund the development of research activities on SARS-COV 2 and COVID-19 disease financed with REACT-EU resources from the European Regional Development Fund "A way of making Europe".

Data availability

Data will be made available on request.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to express their gratitude to research fellow Mr. Felipe Nuñez-Sanchez for his kind support in the inter-coder screening and analysis process.

References

[1] Dubé E, Laberge C, Guay M, Bramadat P, Roy R, Bettinger JA. Vaccine hesitancy: An overview. Hum Vaccin Immunother 2013;9(8):1763–73. https://doi.org/10.4161/hv.24657.

[2] Lentzen MP, Huebenthal V, Kaiser R, Kreppel M, Zoeller JE, Zirk M. A retrospective analysis of social media posts pertaining to COVID-19 vaccination side effects. Vaccine 2022;40(1):43–51. https://doi.org/10.1016/j.vaccine.2021.11.052.

[3] World Health Organization (WHO). COVID-19 vaccines; 2022. Available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/covid-19-vaccines.

[4] Katz M, Nandi N. Social Media and Medical Education in the Context of the COVID-19 Pandemic: Scoping Review. JMIR Med Educ 2021;7(2):e25892.

[5] Park S, Massey PM, Stimpson JP. Primary Source of Information About COVID-19 as a Determinant of Perception of COVID-19 Severity and Vaccine Uptake: Source of Information and COVID-19. J Gen Intern Med 2021;36(10):3088–95. https://doi.org/10.1007/s11606-021-07080-1.

[6] Centro Europeo para la Prevención y Control de Enfermedades (ECDC). Vaccine hesitancy; n.d. Available at: https://www.ecdc.europa.eu/en/immunisation-vaccines/vaccine-hesitancy.

[7] Statista. Number of social media users worldwide from 2018 to 2027; 2022. Available at: https://www.statista.com/statistics/278414/number-of-worldwide-social-network-users.

[8] Manco VA. Análisis de los discursos de médicos vertidos en Twitter sobre la validez de posibles tratamientos contra la COVID-19. Rev Española Comun Salud 2021;12(1):75–88. https://doi.org/10.20318/recs.2021.5593.

[9] Pew Research Center. News Consumption Across Social Media in 2021; 2021. Available at: https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021/.

[10] Garimella, VRK, Weber I. A long-term analysis of polarization on Twitter. Proceedings of the International AAAI Conference on Web and Social Media 2017; 11(1): 528–531. https://ojs.aaai.org/index.php/ICWSM/article/view/14918.

[11] Hu T, Wang S, Luo W, Zhang M, Huang X, Yan Y, et al. Revealing Public Opinion Towards COVID-19 Vaccines With Twitter Data in the United States: Spatiotemporal Perspective. J Medical Internet Res 2021;23(9):e30854.

[12] Jayasurya GG, Kumar S, Singh SK, Kumar V. Analysis of Public Sentiment on COVID-19 Vaccination Using Twitter. IEEE Trans Comput Soc Syst 2021;9(4):1–11. https://doi.org/10.1109/TCSS.2021.3122439.

[13] Jemielniak D, Krempovych Y. An Analysis of AstraZeneca COVID-19 Vaccine Misinformation and Fear Mongering on Twitter. Public Health 2021;200:4–6. https://doi.org/10.1016/j.puhe.2021.08.019.

[14] Jiang X, Su MH, Hwang J, Lian R, Brauer M, Kim S, et al. Polarization Over Vaccination: Ideological Differences in Twitter Expression About COVID-19 Vaccine Favorability and Specific Hesitancy Concerns 205630512110484. Soc Media Soc 2021;7(3). https://doi.org/10.1177/20563051211048413.

[15] Malova E. Understanding online conversations about COVID-19 vaccine on Twitter: vaccine hesitancy amid the public health crisis. Commun Res 2021;38(5):346–56.

[16] Saldaña J. The coding manual for qualitative researchers, 3E [Third edition]. Los Angeles; London: SAGE; 2016. 339 p.

[17] Chandler D, Munday R. A dictionary of media and communication, 1st ed. Oxford; New York: Oxford University Press; 2011. 472 p. (Oxford paperback reference).

[18] Foucault M. A arqueologia do saber. 7. ed. Rio de Janeiro: Forense Universitária; 2008.

[19] O'Connor C, Joffe H. Intercoder Reliability in Qualitative Research: Debates and Practical Guidelines 160940691989922. Int J Qual Methods 2020;19. https://doi.org/10.1177/1609406919899220.

[20] Scannell D, Desens L, Guadagno M, Tra Y, Acker E, Sheridan K, et al. COVID-19 Vaccine Discourse on Twitter: A Content Analysis of Persuasion Techniques. Sentiment and Mis/Disinformation J Health Commun 2021;26(7):443–59. https://doi.org/10.1080/10810730.2021.1955050.

[21] Blaikie N. Analysing Quantitative Data. From Description to Explanation. London: SAGE Publications Ltd; 2003.

[22] Sharpe D. Chi-Square Test is Statistically Significant: Now What? Pract Assess Res 2015;20(8):1–10. https://doi.org/10.7275/tbfa-x148.

[23] Majmundar A, Allem J, Cruz TB, Unger JB. The Why We Retweet scale. PLoS One 2018;13(10):e0206076.

[24] Sv P, Tandon J, Vikas, Hinduja H. Indian citizen's perspective about side effects of COVID-19 vaccine – A machine learning study. Diabetes Metabolic Syndrome: Clin Res Rev 2021; 15(4): 102172. doi: 10.1016/j.dsx.2021.06.009.

[25] Herrera-Peco I, Jiménez-Gómez B, Peña Deudero JJ, Benitez De Gracia E, Ruiz-Núñez C. Healthcare Professionals' Role in Social Media Public Health Campaigns: Analysis of Spanish Pro Vaccination Campaign on Twitter. Healthcare 2021;9(6):662. https://doi.org/10.3390/healthcare9060662.

[26] World Health Organization. Ten threats to global health in 2019. [Internet]. 2019. Disponible en: https://www.who.int/news-room/spotlight/ten-threats-to-global-health-in-2019.

[27] Sallam M. COVID-19 Vaccine Hesitancy Worldwide: A Concise Systematic Review of Vaccine Acceptance Rates. Vaccines 2021;9(2):160. https://doi.org/10.3390/vaccines9020160.

[28] Kieslich K. Addressing vaccination hesitancy in Europe: a case study in state-society relations. Eur J Public Health 2018;28(Suppl 3):30–3. https://doi.org/10.1093/eurpub/cky155.

[29] Das R, Ahmed W. Rethinking Fake News: Disinformation and Ideology during the time of COVID-19 Global Pandemic. IIM Kozhikode Soc Manage Rev 2022;11(1):146–59. https://doi.org/10.1177/22779752211027382.

[30] Roozenbeek J, Schneider CR, Dryhurst S, Kerr J, Freeman ALJ, Recchia G, et al. Susceptibility to misinformation about COVID-19 around the world. R Soc Open Sci octubre de 2020;7(10):. https://doi.org/10.1098/rsos.201199201199.

[31] Hsia K, Kong E. Characterizing the Reaction of Doctors to COVID-19 on Twitter. J Quant Descr Digit Media 2022;2(1–18). , https://journalqd.org/article/view/3245/2578. https://doi.org/10.51685/jqd.2022.012.

[32] Batra R, Imran AS, Kastrati Z, Ghafoor A, Daudpota SM, Shaikh S. Evaluating Polarity Trend Amidst the Coronavirus Crisis in Peoples' Attitudes toward the Vaccination Drive. Sustainability 2021;13(10):5344. https://doi.org/10.3390/su13105344.

[33] Holford D, Fasce A, Tapper K, Demko M, Lewandowsky S, Hahn UA, et al. A Manifesto for Science Communication as Collective Intelligence. 2022; 1–12. 10.17605/OSF.IO/TZUFW