Mapping vaccine sentiment: analyzing spanish-language social media posts and survey-based public opinion

Agnes Huguet-Feixa, Wasim Ahmed, Eva Artigues-Barberà, Joaquim Sol, Xavier Gomez-Arbones, Pere Godoy, Marta Ortega Bravo

Table of Contents

Mapping vaccine sentiment: analyzing spanish-language social media posts and survey-based public opinion

Agnes Huguet-Feixa[1, 2, 3, 4] MD, Prof Dr; Wasim Ahmed[5] MsC, PhD; Eva Artigues-Barberà[2, 3, 6, 7] Nurse, PhD, Prof Dr; Joaquim Sol[3, 6, 8] PhD, Prof Dr; Xavier Gomez-Arbones[4] PhD, Prof Dr Med; Pere Godoy[4, 9, 10, 11] PhD, MSc, Prof Dr Med; Marta Ortega Bravo[3, 4, 12, 13] PhD, MD, Prof Dr Med

[1]CAP Rambla Ferran.Institut Català de la Salut Lleida ES
[2]Grup de Recerca multidisciplinari en Terapèutica i Intervencions en Atenció Primària. (Grup RETICAP) Lleida ES
[3]Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIA-PJGol) Barcelona ES
[4]Faculty of Medicine, University of Lleida Lleida ES
[5]Hull University Business School Hull GB
[6]Atenció Primària, Gerència Territorial de Lleida, Institut Català de la Salut Lleida ES
[7]Facultat d'Infermeria i Fisioteràpia, Universitat de Lleida Lleida ES
[8]Institut de Recerca Biomèdica de Lleida, Department of Experimental Medicine, University of Lleida Lleida ES
[9]Hospital Universitario Santa Maria. Gestió de Serveis Sanitaris., Departament de Salut Lleida ES
[10]Centro de Investigación Biomédica en Red de Epidemiologia y Salud Publica, Instituto de Salud Carlos III Madrid ES
[11]Institut de Recerca Biomèdica de Lleida, Applied Epidemiology Research Group, University of Lleida Lleida ES
[12]Grup de Recerca en Ecografia Clínica d'Atenció Primària (Grup GRECOCAP Lleida ES
[13]CAP Almacelles, Atenció Primària, Gerència Territorial de Lleida, Institut Català de la Salut Lleida ES

Corresponding Author:
Marta Ortega Bravo PhD, MD, Prof Dr Med
Fundació Institut Universitari per a la recerca a l'Atenció Primària de Salut Jordi Gol i Gurina (IDIA-PJGol)
Gran Via de les Corts Catalanes, 587
Barcelona
ES

Abstract

Background: The internet and social media have been considered useful platforms for obtaining health information. However, critical and erroneous content about vaccines on SM has been associated with vaccination delays and refusal.

Objective: To determine the profile of the population using social networks to obtain information about vaccines through a survey and to analyze the sentiments and opinions regarding vaccines in Spanish-language posts on the social network X, along with their geolocation.

Methods: A study was conducted using two methodologies: an observational study in the population of Spain aged 18 years or older, obtaining information through a self-completed electronic questionnaire in 2021, and a second study that analyzed Spanish posts gathered from 'X' between March and December 2021. Orange Data Mining was used to create a workflow for sentiment analysis of the posts. Location-based analysis was conducted by drawing upon self-defined user locations from X, entered into Microsoft PowerBI for analysis. Social network analysis was conducted to identify the nature of the five largest groups of users conversing about vaccinations in Spanish by drawing upon NodeXL Pro.

Results: Among the 1,312 respondents in the survey, 85.7% stated that they were regular social networks users, and 66% reported having encountered antivaccine information on social networks. Among these, 24.3% experienced doubts about receiving recommended vaccines, and of those with doubts, 13.3% refused at least one vaccine proposed by a healthcare professional. A total of 479,734 Spanish posts on X were analyzed, and 54.44% (261,183 posts) were negative, 28.18% were neutral, and 17.37% were positive. Sentiment varied across regions: more negative posts appeared to derive from South America, with a mix in Europe and more positive posts in North America. Analysis of the topic words and key themes allowed the grouping of the predominant themes of the five study groups, which were: vaccination efforts during the COVID-19 pandemic (1), issues of vaccine theft and struggles in managing and securing the vaccine supply (2), campaigns in the State of Mexico (3), vaccination efforts for older adults (4), and the vaccination campaign in Colombia to combat COVID-19 (5).

Conclusions: High proportions of exposure to antivaccine content were reported by the respondents. Sentiment analysis and

geolocation of posts on the social network X revealed a significant amount of negative Spanish posts, predominantly from South America. The thematic analysis of conversations on X proved to be a highly useful tool for understanding the population's opinions about vaccines.

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 v
  Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Mapping vaccine sentiment: analyzing spanish-language social media posts and survey-based public opinion

Abstract

Background: The internet and social media have been considered useful platforms for obtaining health information. However, critical and erroneous content about vaccines on SM has been associated with vaccination delays and refusal.

Objective: To determine the profile of the population using social networks to obtain information about vaccines through a survey and to analyze the sentiments and opinions regarding vaccines in Spanish-language posts on the social network X, along with their geolocation.

Methods: A study was conducted using two methodologies: an observational study in the population of Spain aged 18 years or older, obtaining information through a self-completed electronic questionnaire in 2021, and a second study that analyzed Spanish posts gathered from 'X' between March and December 2021. Orange Data Mining was used to create a workflow for sentiment analysis of the posts. Location-based analysis was conducted by drawing upon self-defined user locations from X, entered into Microsoft PowerBI for analysis. Social network analysis was conducted to identify the nature of the five largest groups of users conversing about vaccinations in Spanish by drawing upon NodeXL Pro.

Results: Among the 1,312 respondents in the survey, 85.7% stated that they were regular social networks users, and 66% reported having encountered antivaccine information on social networks. Among these, 24.3% experienced doubts about receiving recommended vaccines, and of those with doubts, 13.3% refused at least one vaccine proposed by a healthcare professional. A total of 479,734 Spanish posts on X were analyzed, and 54.44% (261,183 posts) were negative, 28.18% were neutral, and 17.37% were positive. Sentiment varied across regions: more negative posts appeared to derive from South America, with a mix in Europe and more positive posts in North America. Analysis of the topic words and key themes allowed the grouping of the predominant themes of the five study groups, which were: vaccination efforts during the COVID-19 pandemic (1), issues of vaccine theft and struggles in managing and securing the vaccine supply (2), campaigns in the State of Mexico (3), vaccination efforts for older adults (4), and the vaccination campaign in Colombia to combat COVID-19 (5).

Conclusions: High proportions of exposure to antivaccine content were reported by the respondents. Sentiment analysis and geolocation of posts on the social network X revealed a significant amount of negative Spanish posts, predominantly from South America. The thematic analysis of conversations on X proved to be a highly useful tool for understanding the population's opinions about vaccines.

Keywords: Anti-Vaccination Movement; Vaccination Refusal; vacunacion; social media

Introduction

The critical and erroneous content about vaccines on social network (SN) has posed a significant obstacle to immunizing the population against vaccine-preventable diseases [1-3]. A decline in vaccination coverage represents a threat to the herd immunity acquired in recent years through the efforts of healthcare professionals, and could lead to outbreaks of diseases, such as the measles outbreak in Europe in 2023-24 [4].

The crisis in the vaccination system and the resurgence of antivaccine movements (AVM) are due to the increased accessibility of misinformation and the reduced credibility of healthcare personnel [5-7]. The internet and social media (SM) have changed the way the public accesses health information

[8, 9]. This information on SN frequently contains incorrect data about vaccine effectiveness or data from AVM, which can influence the decision to reject vaccination before individuals even set foot in primary care centres [9, 10]. A 2020 article stated that Facebook® pages that distrust established health guidelines are more effective at influencing hesitant individuals than government agencies [11]. Distrust in scientific positions could spread and dominate SN conversations over the next decade [11]. Exposure to antivaccine content on SM has been associated with delays and rejection of vaccination and is considered one of the major causes of vaccine hesitancy [3, 12-15].

Current literature has yet to explain how the antivaccine movement continues to engage and persuade the public to reject immunization despite the efforts of vaccine advocates [12]. Anti-vaccine proponents on SN have shown more active engagement patterns than pro-vaccine groups. In a 2020 study analyzing Facebook® users, it was shown that antivaccine groups on SN, though a minority, had the potential to be more influential than pro-vaccine groups. Moreover, undecided groups were not passive; they were the most active agents in the discussion. This fact favoured the interaction of the undecided with antivaccine groups [11], and for this reason, intervention strategies were proposed to identify central and influential antivaccine groups to reduce their growth and the formation of future antivaccine or undecided groups [11].

X (formerly known as "Twitter") has been an important source of information for studying vaccine hesitancy, as SM platforms have been considered effective tools for communication between individuals and organizations, but they have also been used as tools to spread false information and conspiracy theories about vaccines [16-18].

Several studies have been published analyzing public opinions or sentiments about vaccines on X using artificial intelligence. These publications suggested that the analysis of SN using artificial intelligence should be considered by institutions and governments alongside surveys and other conventional methods to assess public attitudes towards vaccines [12, 17, 19, 20]. Such analyses could enable real-time evaluation of the public's safety and confidence in vaccines, help address the concerns of vaccination skeptics, and develop more effective communication policies and strategies to maximize acceptance [20, 21].

The objectives of our study were to determine the profile of the population using SN to obtain information about vaccines through a population survey in Spain and to analyze the sentiments and opinions related to vaccines in Spanish-language messages on "X" and geolocate them using artificial intelligence.

Methods

Survey Data Collection and Analysis

Study Design

We conducted an observational, cross-sectional study on individuals who had access to an online survey from March to December 2021. The inclusion criteria required participants to be aged 18 years or older and have the authority to make vaccination decisions for themselves or others in Spain. We obtained the information through an electronic, self-administered questionnaire designed by the project research team. A pilot test was performed before the definitive questionnaire was obtained. To ensure the rigor and validity of the study, an experienced research team with expertise in conducting surveys will be needed, thereby guaranteeing the quality and reliability of the instrument. Both were registered on a Research Electronic Data Capture (REDCap) web platform on a centralized server where the data remain in the custody of the Institut Català de la Salut. Through the REDCap web platform, we also built a database of the participants. Anonymous information was exported to the statistical packages used for subsequent analysis. The REDCap platform generated a

link [22] for participation in the survey that was disseminated through scientific societies, SM, research institutes, pediatricians, and nurses in primary care.

Variables

The main variables of the questionnaire were the following: sociodemographic factors (sex, age, having children under 15 years of age, level of education), information on the use of SN (use of SN, duration of use, whether information about vaccines was sought on SN, which SN were used to search for vaccine information, the year of the first search for vaccines on SN, and whether the search was related to COVID-19), whether they had encountered antivaccine information on SN and whether such information had caused doubts, the type of doubts, whether the information on SN had led them to reject vaccines, whether vaccines had been rejected for a child, the age of the child, who initiated the doubt, gender, age, and level of education of the person rejecting vaccination, whether they had received opinions against vaccines on SN, whether they had received opinions in favor of vaccines on SN, whether they followed SN that generated doubts about vaccines or profiles against or in favor of vaccination, and whether they had made any comments against or in favor of vaccines on SN.

Statistical Analysis

The data were gathered in an anonymized database using the REDCap platform. We conducted a descriptive analysis of the findings, categorizing qualitative or ordinal data using absolute and relative frequencies. The analysis was complemented within selected subgroups of interest to identify the factors independently associated with these results. Statistical analyses were performed with R software version 4.1.2. In the statistical analysis, only complete cases were considered, and missing values were assumed to be missing at random.

X Insights and Social Network Analysis

Data Retrieval
Posts were gathered from 'X' (formerly known as Twitter). The search string in order to retrieve posts was as followed: "vacunacion OR vacunacio OR vacunas OR vacunes OR antivacunas OR antivacunes OR antivacinacion OR antivacunacio". A total of 479,734 posts were captured (excluding reposts), and these were sent by 29,706 users. The data were randomly sampled between March 2021 and December 2021 (see Table 1).

Table 1. Overview Data (Tweet Volumes Total No. of X Users)

Month (2021)	Total No. of Posts	Total No. of X Users
March	54,837	40,093
April	53,793	39,426
May	53,613	39,114
June	54,499	38,590
July	54,074	40,170
August	54,002	39,303
September	47,141	33,402
October	35,071	25,545
November	35,417	27,944
December	37,287	29,706
Total	479,734	----------------------

Sentiment Analysis

Orange Data Mining was used to create a workflow for sentiment analysis of the 479,734 posts that were captured. Sentiment analysis aims to calculate the sentiment for each post within the dataset. The study drew upon the multilingual sentiment lexicon built into Orange Data Mining, developed from the Data Science Lab's SentiArt tool developed by Arthur Jacobs [23]. In Orange Data Mining, depending on the content of each post, a score is provided to each post, such as -10 (i.e., negative), 6 (which would be positive) and 0 (neutral). These scores are based on the positive/negative lexicons within the posts. An average sentiment score was also calculated by computing the average sentiment score of all 479,734 posts.

Location-Based Analysis

Location-based analysis was conducted by drawing upon self-defined user locations from X (formerly Twitter), entered into Microsoft PowerBI for analysis. A challenge with self-defined locations (i.e., user, reported) is that as they allow users to enter free text, they may provide a city or region rather than the country. These are, therefore, better analyzed visually. Data was collected from April 2021 to December 2021. Location analysis focused on examining a specific time period in our dataset: April 2021.

Social Network Analysis and Topic Insight

SN analysis was conducted to identify the nature of the top 5 groups conversing about vaccinations in Spanish by drawing upon NodeXL Pro. SN and topic analysis focused on examining April 2021. Each cluster of users is conversing about different topics, and our analysis has produced insight into the two words most often used together. This helps shed light on the issues users may have been conversing about. The networks were laid out using the Group in the Box format. The graphs were directed. The graph's vertices were grouped by cluster using the Clauset-Newman-Moore cluster algorithm [24]. The graph was laid out using the Harel-Koren Fast Multiscale layout algorithm [25].

Ethical Considerations

This study was approved by the ethics and clinical Research Committee of the Institut Universitari per a la recerca a l'Atenció Primària de Salut (IDIAP) Jordi Gol i Gurina, with code 20/221-P. The study was conducted in accordance with the principles of the Declaration of Helsinki. The variables collected were treated anonymously and to guarantee the confidentiality of the data, as established in Regulation (EU) 2016/679 of the European Parliament and the Council of April 27 on Data Protection (RGPD) and the organic law 3 /2018, of December 5, protection of personal data and guarantee of digital rights. The database is kept by the principal investigator and the research team in an Excel format, protected by password access. An anonymized database was used for the analysis. Before carrying out the survey, online informed consent had to be completed, accepted, and signed.

Results:

Survey results (Table 2)

A total of 1,312 respondents were analyzed, with 74.5% being female, 71.0% university graduates, 46.9% having children under 15 years of age, 14.0% aged 30 years or younger, 12.4% over 60 years old, and 73.7% between 31 and 59 years old. Among the respondents, 85.7% stated they were regular SN users and shared sociodemographic characteristics (age, children, gender, and education) similar to the total sample analyzed (Table 2), with a predominance of females and university education. Of the regular SN users, 76% had been using SN for more than five years, and 35.6% reported seeking information about vaccines on SN. Among these, 39.1% used X, 29.6% used Instagram, 33.1% used Facebook, and 46.6% used other SN. A total of 53.3% reported conducting their first vaccine search on SN in 2020, with 84.1% stating that their search was related to the COVID-19 pandemic.

Among the respondents, 66% reported having encountered antivaccine information on SN; of those

who received this information, 24.3% (205/850) had doubts about the administration of recommended vaccines, and among those who had doubts, 13.3% (27/205) rejected a vaccine recommended by a healthcare professional. Among those who rejected vaccines for children, 10 (50%) were women, 13 (59.1%) were aged between 40 and 59 years, and 17 (66%) had university degrees; vaccines were rejected for children with an average age of 9.1 years.

A total of 63.3% of the respondents reported having received opinions or comments against vaccines on SN, and 78.2% reported having received opinions or comments in favor. Additionally, 7.3% reported following SN accounts that generated doubts about vaccines, and 7.7% reported following antivaccine profiles on SN, while 37.1% followed pro-vaccine profiles on SN. A total of 2.4% stated they had made a comment against vaccines on SN, and 27.4% reported having made a comment in favor.

Table 2. Summary descriptives

	Total *N=1312*	*Responders*
Sex		1280
Male	326 (25.5%)	
Female	954 (74.5%)	
Age:		1287
30 years or less	180 (14.0%)	
31-59 years	948 (73.7%)	
60 years or more	159 (12.4%)	
University studies	915 (71.0%)	1289
Do you have children aged 14 or younger?	604 (46.9%)	1289
Do you regularly use social networks?	1124 (85.7%)	1312
Since when?:		1121
Less than 1 year ago	13 (1.16%)	
Since Covid-19 lockdown	18 (1.61%)	
1 to 5 years	238 (21.2%)	
More than 5 years	852 (76.0%)	

Have you searched for vaccine information on social networks?	399 (35.6%)	1121
On which social networks? (choice=Twitter)	156 (39.1%)	399
On which social networks? (choice=Instagram)	118 (29.6%)	399
On which social networks? (choice=Others)	186 (46.6%)	399
On which social networks? (choice=Facebook)	132 (33.1%)	399
In which year did you first search for vaccine information on social networks?		369
2020	196 (53.1%)	
Before 2020	154 (41.7%)	
After 2020	19 (5.15%)	
Was it related to Covid-19?	329 (84.1%)	391
Have you found information AGAINST vaccines on social networks?	850 (66.0%)	1287
Did it generate any doubts about the recommended vaccination?	205 (24.3%)	845
Has information from social networks led you to reject any vaccine proposed by a healthcare professional?	27 (13.3%)	203
If you have rejected vaccinating a child, what were their ages?:	9.11 (7.41)	9
If you have rejected vaccinating a child, who initially had the doubt?:		15
Myself	13 (86.7%)	
Another person	2 (13.3%)	
Sex:		17
Male	5 (29.4%)	
Female	10 (58.8%)	
Other	2 (11.8%)	
Age:		22
Less than 30	5 (22.7%)	
30 – 39	2 (9.09%)	
40 – 49	6 (27.3%)	
50 – 59	7 (31.8%)	
60 – 69	1 (4.55%)	
More than 70	1 (4.55%)	

Highest educational level completed by the person who rejected vaccination:		25
No education or incomplete primary studies	1 (4.00%)	
Secondary	3 (12.0%)	
High School	3 (12.0%)	
Vocational training programes	1 (4.00%)	
University	17 (68.0%)	
Have you received any type of opinion or comment AGAINST vaccines through social networks?	815 (63.3%)	1287
Have you received any type of opinion or comment IN FAVOR of vaccines through social networks?	1007 (78.2%)	1287
Do you follow social networks that generate doubts about vaccines?	94 (7.28%)	1291
Do you follow any social media profile that is AGAINST vaccines?	99 (7.67%)	1291
Do you follow any social media profile IN FAVOR of vaccines?	479 (37.1%)	1290
Have you made any comments on social networks AGAINST vaccines?	31 (2.40%)	1289
Have you made any comments on social networks IN FAVOR of vaccines?	351 (27.4%)	1281

Social Network results:

Sentiment Analysis Results

The analysis revealed the following distribution: 17.37% (n= 83,357), were classified as 'Positive'. A significant portion, 54.44% (n=261,183) of posts, were categorized as 'Negative'. Neutral sentiments accounted for 28.18% (n=135,194). These findings suggest that the content in posts was likely to be more negative in nature (Table 3). Furthermore, the overall average sentiment was negative, with a result of -1.96.

Table 3. Classification of Sentiment

Sentiment	No.	Per cent
Positive	83,357	17.37
Negative	261,183	54.44
Neutral	135,194	28.18
Total	479,734	100

Location-Based Analysis

Location analysis focused on examining a specific time period in our dataset: April 2021, as shown in Supplementary Figures 1 and 2; Most posts are from South America (such as Argentina, Venezuela, Columbia, and Mexico) and Europe (mainly Spain). Some posts are across other European Countries, such as the UK and Germany).

We can observe a significant concentration of users in regions where Spanish is predominantly spoken. As highlighted, these posts originate from South American countries like Argentina, Venezuela, Colombia, and Mexico. This trend highlights the Spanish-speaking demographic within

these nations. Additionally, a notable volume of posts is traced back to Europe, particularly Spain, further highlighting the Spanish language's influence when retrieving the data, as Spanish keywords were used to retrieve posts.

To gain an insight into the distribution of positive, neutral, and negative posts further analysis was conducted using data captured during April 2021. The positive, neutral, and negative groups were identified during the SN analysis. Figure 1 and 2 (April 2021) provides a visual overview of the locations. The sentiment varies across regions, for instance, more negative posts appear to derive from South America, with a mix in Europe and more positive posts in North America. This would reflect the conversations taking place at the time and reflects only those users that disclose their location voluntarily.
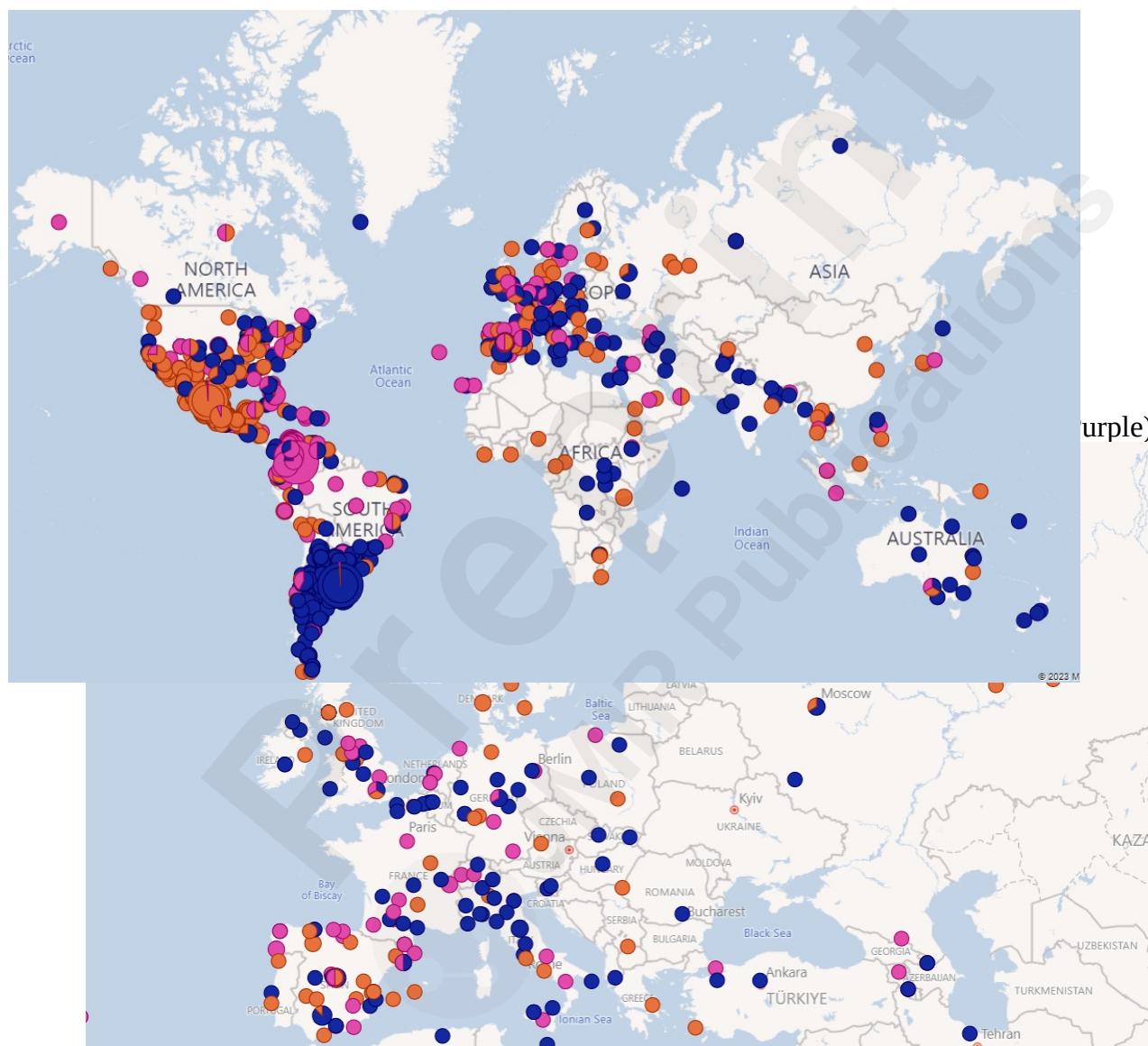

Figure 2. Zooming into Europe (Mixed Sentiment)

Social Network Analysis and Topic Insight (Figure 3 and Table 4)

The most prominent Group identified was the 'isolates' Group (Group 1). This Group is characterized by individuals who sent original posts without engaging in mentions or reposts. The predominance of this Group suggests that the topics were widely popular and managed to draw a variety of unique and original opinions from a diverse audience.

The groups numbered 3 to 5 exhibited characteristics of community network shapes intertwined with elements of broadcast. These groups demonstrated high interconnectedness, indicative of frequent and active back-and-forth conversations. At the same time, there was a noticeable amplification of

certain accounts within these groups, which points to a blend of personal interaction and wider broadcast of information or opinions.
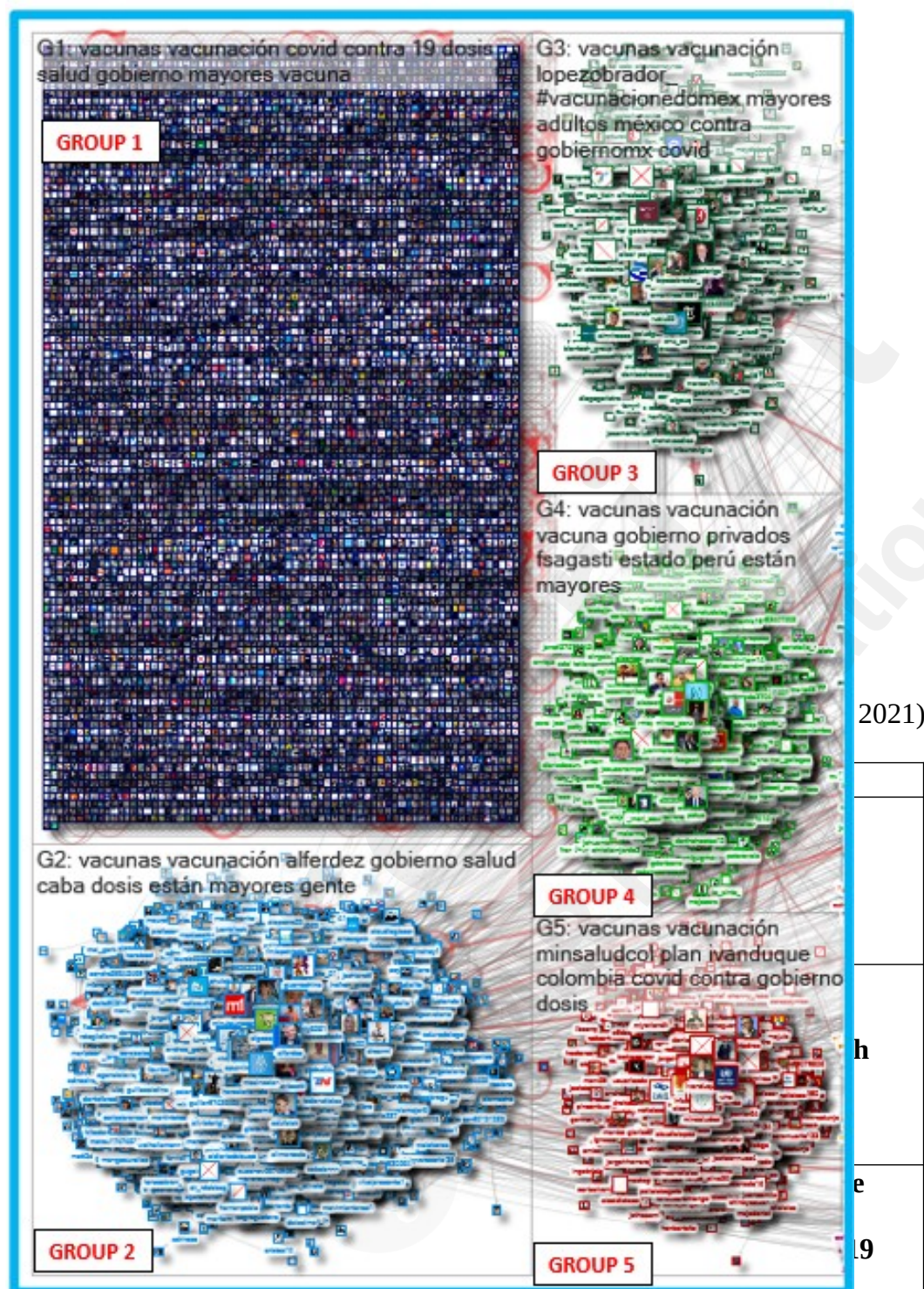
	Community Focus in Vaccination
Group 4	Chinese Vaccines and Procurement
	Vaccination Process and Challenges
	Healthcare Capacity and Response

Group 5	**Comprehensive National Vaccination Plan**
	Special Attention to Vulnerable Populations
	Ministry of Health's Involvement

Examining the keywords and network insights, **in Group 1**, conversations center around the global response to the COVID-19 pandemic, and they emphasize vaccination efforts. Users discussed the vaccination campaign targeting various groups, particularly older adults, who are among the most vulnerable to the virus.

In **Group 2**, discussions indicated challenges pointing to vaccine theft issues and highlighting the struggles in managing and securing the vaccine supply. There were also discussions on organized strategies and public health initiatives to combat the pandemic.

In **Group 3**, discussions were based on a focused campaign in the State of Mexico (Edomex), and users discussed the scale and nature of the initiative, which is the large-scale distribution of vaccines. There was also a personal, community-centric dimension to the campaign, which emphasized the societal importance of protecting older adults against COVID-19.

In **Group 4**, there was a focused narrative on the COVID-19 pandemic, particularly highlighting the vaccination efforts for older adults. Users also referred to the "Chinese vaccines" which implies a specific type of vaccine being discussed. Central themes included the process and challenges of vaccination, the procurement of vaccines, and the scale of the effort.

In Group 5, the keywords suggest a comprehensive national plan in Colombia to combat COVID-19 through a vaccination campaign. The discussions also noted the well-defined procedure for administering the vaccines. There were also discussions of a targeted approach towards vaccinating older adults, particularly those over 80 years old, recognizing their vulnerability to the virus. There are also discussions around the involvement of the Colombian Ministry of Health and its officials, including Fernando Ruiz Gómez, in orchestrating and overseeing the public health initiative.

In terms of the sentiment with the groups, Group 1 appears to be more neutral without much inherent negativity in the terms. However, in Group 2, as it is based on the controversy surrounding vaccine distribution, there is a more critical tone. Group 3 contains some positivity. Group 4, is mostly neutral, with some positive words. Group 5 is primarily neutral.

Discussion

This study aimed to determine the profile of the population using SN to obtain information about vaccines through a population survey, as well as to analyze the sentiments and opinions about vaccines present on SM platforms. It is well-known that SM platforms are sources of health information and can have considerable influence on health decision-making [26]. In our study, we confirmed that SN was frequently used as a source of information regarding vaccines for the surveyed population. However, 66% of respondents reported encountering antivaccine information on SN, and among the analyzed messages on X, 54.44% or 261,183 posts (formerly called tweets) were categorized as 'Negative'. In a publication analyzing 9,352,509 English posts in 2020, only 232,975 (2.49%) posts were identified as anti-vaccination, although negative posts increased after the implementation of the first COVID-19 vaccination in November 2020 [17]. Another study analyzing English tweets in the United Kingdom and the United States from March to November 2020 found negative tweet percentages of 27.95% and 30.57%, respectively [20]. These data reinforce the well-known fact that although X is a relevant tool for obtaining verified information, conspiracy theories with incorrect information also emerge [27, 28]. These negative messages

observed on SN generate hesitancy in the population, as shown in our survey analysis and previous publications, potentially leading to delays or refusals of vaccination proposed by healthcare professionals [3, 13-15].

In our analysis of SN, we identified relational patterns within the studied groups on the topic of vaccination. Specifically, groups numbered 3 to 5 exhibited characteristics of community network shapes intertwined with elements of broadcast. These groups demonstrated high interconnectedness, indicative of frequent and active back-and-forth conversations, mixing positive, negative, and neutral sentiments. A previous publication shows that interaction between antivaccine groups and undecided groups could facilitate vaccination hesitancy in the undecided due to exposure to antivaccine messages [11].

The analysis of each group allows us to see the key themes of the conversations, which can be very useful for detecting the population's concerns and creating vaccination campaigns aimed at combating misinformation or doubts about vaccination, potentially generating new evidence in community interventions as described in previous publications [9, 12, 18, 20, 29]. Some of these publications propose a method to counteract fake news, which involves quickly detecting and directly addressing them as they arise [27, 28]. The analysis of topic words and key themes in our study allowed us to group the predominant themes of the five study groups, which were: vaccination efforts for the COVID-19 pandemic (1), vaccine theft issues and struggles in managing and securing the vaccine supply (2), a campaign in the State of Mexico (3), vaccination efforts for older adults (4), and a vaccination campaign in Colombia to combat COVID-19 (5). We hypothesize that the topic words and themes may be influenced by the demographic distribution of posts from Spanish-speaking countries, primarily in South America. Another article also analyzed the themes of vaccine-related tweets on X in the Australian population, primarily detecting three themes: attitudes and actions toward COVID-19 and its vaccination, infection control measures, and confidence in COVID-19 vaccine trials, alongside baseless claims, conspiracy theories, complaints, and misconceptions about various measures against COVID [19]. Another article develops algorithms for automatically classifying a large number of vaccine-related tweets into three classes: pro-vaccine, anti-vaccine, and neutral [12].

The geographical location of the messages and the analysis of positive, negative, and neutral sentiments are very useful tools for generating vaccination campaigns and detecting groups with higher hesitancy [20]. In our study, we found that more negative posts appeared to derive from South America, with a mix in Europe, and more positive posts in North America. In Spain, positive and neutral posts predominated, which aligns with Spain generally showing good vaccination coverage and acceptance among the population [30]. However, given the high level of technological connectivity in today's society, a tweet or post generated in North or South America could influence the Spanish population.

A limitation of our study is that the SN Analysis and Topic Insight focused specifically on the X network in April 2021, so our findings may not be applicable to other time periods. Future research could aim to extend the time periods and examine X discussions based on other locations. One of the limitations of the survey was the inherent recruitment bias associated with the online survey method, which restricted participation to individuals with internet access. However, given that 96.1% of Spanish households have internet access, and 85% of Spaniards are users of SN [31, 32], the reach of our survey remains substantial. Another possible limitation is potential respondent repetition, although we anticipate minimal impact on the final results because of expected low rate of repetition. Additionally, our sample skewed towards women and people with a university education, this could limit the interpretation and generalization of our study to the broader population. Finally, reaching the anti-vaccine population posed challenges, as interactions with these groups is difficult. To mitigate eventual bias deriving from this, in our sample size calculation, we considered that the proportion of antivaccine responses would be much lower than the proportion of pro-vaccine responses. Despite these limitations, our study provides valuable insights into vaccine refusal and

hesitancy, but caution is warranted in extrapolating findings to the entire population.

Conclusions:

Our study analyzed the use of SN to obtain information about vaccines, detecting high proportions of exposure to anti-vaccine content reported by surveyed users. The sentiment analysis on the SN X and its geolocation revealed a significant amount of negative Spanish-language posts circulating on X, with a predominance of negative content mainly in South America. The analysis of the themes of conversations on the SN X using artificial intelligence proved to be a highly useful tool for understanding public opinions about vaccines. It would enable targeted campaigns to address skeptical content about vaccination and scientific positions in specific populations at particular moments in time.

Acknowledgments

Data Availability Statement

The data are hosted on the research team's internal servers and will be provided under reasonable request.

Authors' Contributions

AHF, EAB, WA and MOB contributed to the study concept and research design and selected and reviewed the literature on vaccination.

WA, JS, XA participated in the acquisition, analysis, and interpretation of data. AHF, EAB, MOB, JS, and PG participated in the writing and critical review of the manuscript and approved the final version. AHF and MC-MUVA contributed to recruitment.

Conflicts of Interest

None declared.

Abbreviations

Antivaccine movements (AVM)
Coronavirus disease (COVID-19)
Research Electronic Data Capture (REDCap)
Social Media (SM)
Social Network (SN)

Multimedia Appendix 1

MC-MUVA work group.

References

1. Lafnitzegger A, Gaviria-Agudelo C. Vaccine Hesitancy in Pediatrics. Adv Pediatr. 2022;69(1):163-76.

2. Larson HJ, Jarrett C, Eckersberger E, Smith DM, Paterson P. Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: a systematic review of published literature, 2007-2012. Vaccine. 2014;32(19):2150-9.

3. Ortiz RR, Smith A, Coyne-Beasley T. A systematic literature review to examine the potential for social media to impact HPV vaccine uptake and awareness, knowledge, and attitudes about HPV and HPV vaccination. Hum Vaccin Immunother. 2019;15(7-8):1465-75.

4. Europa registra un aumento alarmante de casos de sarampión: Naciones Unidas. Centro Regional de Información.; 25/01/2024 [19/05/2024]. Available from: https://unric.org/es/europa-registra-un-aumento-alarmante-de-casos-de-sarampion/.

5. Hussain A, Ali S, Ahmed M, Hussain S. The Anti-vaccination Movement: A Regression in Modern Medicine. Cureus. 2018;10(7):e2919.

6. Gualano MR, Bert F, Voglino G, Buttinelli E, D'Errico MM, De Waure C, et al. Attitudes towards compulsory vaccination in Italy: Results from the NAVIDAD multicentre study. Vaccine. 2018;36(23):3368-74.

7. Germinario C, Gallone MS, Tafuri S. [How to fight anti-vaccinists prejudices: the viewpoint of public health]. Epidemiol Prev. 2014;38(6 Suppl 2):120-3.

8. Lee K, Hoti K, Hughes JD, Emmerton L. Dr Google and the consumer: a qualitative study exploring the navigational needs and online health information-seeking behaviors of consumers with chronic health conditions. J Med Internet Res. 2014;16(12):e262.

9. Grant L, Hausman BL, Cashion M, Lucchesi N, Patel K, Roberts J. Vaccination persuasion online: a qualitative study of two provaccine and two vaccine-skeptical websites. J Med Internet Res. 2015;17(5):e133.

10. Tafuri S, Gallone MS, Cappelli MG, Martinelli D, Prato R, Germinario C. Addressing the anti-vaccination movement and the role of HCWs. Vaccine. 2014;32(38):4860-5.

11. Neil F. Johnson NV, Nicholas Johnson Restrepo, Rhys Leahy, Nicholas Gabriel, Sara El Oud, Minzhang Zheng, Pedro Manrique, Stefan Wuchty,Yonatan Lupu. The online competition between pro- andanti-vaccination views https://doi.org/10.1038/s41586-020-2281-1: Nature. www.nature.com; 2020 [

12. Argyris YA, Monu K, Tan PN, Aarts C, Jiang F, Wiseley KA. Using Machine Learning to Compare Provaccine and Antivaccine Discourse Among the Public on Social Media: Algorithm Development Study. JMIR Public Health Surveill. 2021;7(6):e23105.

13. Margolis MA, Brewer NT, Shah PD, Calo WA, Gilkey MB. Stories about HPV vaccine in social media, traditional media, and conversations. Prev Med. 2019;118:251-6.

14. Dunn AG, Surian D, Leask J, Dey A, Mandl KD, Coiera E. Mapping information exposure on social media to explain differences in HPV vaccine coverage in the United States. Vaccine. 2017;35(23):3033-40.

15. Burki T. The online anti-vaccine movement in the age of COVID-19. Lancet Digit Health. 2020;2(10):e504-e5.

16. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccin Immunother. 2020;16(11):2586-93.

17. To QG, To KG, Huynh VN, Nguyen NT, Ngo DT, Alley S, et al. Anti-vaccination attitude trends during the COVID-19 pandemic: A machine learning-based analysis of tweets. Digit Health.

2023;9:20552076231158033.

18.   Fuster-Casanovas A, Das R, Vidal-Alaball J, Lopez Segui F, Ahmed W. The #VaccinesWork Hashtag on Twitter in the Context of the COVID-19 Pandemic: Network Analysis. JMIR Public Health Surveill. 2022;8(10):e38153.

19.   Kwok SWH, Vadde SK, Wang G. Tweet Topics and Sentiments Relating to COVID-19 Vaccination Among Australian Twitter Users: Machine Learning Analysis. J Med Internet Res. 2021;23(5):e26953.

20.   Hussain A, Tahir A, Hussain Z, Sheikh Z, Gogate M, Dashtipour K, et al. Artificial Intelligence-Enabled Analysis of Public Attitudes on Facebook and Twitter Toward COVID-19 Vaccines in the United Kingdom and the United States: Observational Study. J Med Internet Res. 2021;23(4):e26627.

21.   Badur S, Ota M, Öztürk S, Adegbola R, Dutta A. Vaccine confidence: the keys to restoring trust. Hum Vaccin Immunother. 2020;16(5):1007-17.

22.   IDIAP JordiGol.   REDCap.   [accessed 2023-02-06].   Available   from: https://redcap.link/uj4jozhy.

23.   Jacobs AM. Sentiment Analysis for Words and Fiction Characters From the Perspective of Computational (Neuro-)Poetics. Front Robot AI. 2019;6:53.

24.   Clauset A, Newman ME, Moore C. Finding community structure in very large networks. Phys Rev E Stat Nonlin Soft Matter Phys. 2004;70(6 Pt 2):066111.

25.   Harel D, Koren Y. A Fast Multi-scale Method for Drawing Large Graphs.

26.   Ahmed W, Vidal-Alaball J, Vilaseca JM. A Social Network Analysis of Twitter Data Related to Blood Clots and Vaccines. Int J Environ Res Public Health. 2022;19(8).

27.   Ahmed W, López Seguí F, Vidal-Alaball J, Katz MS. COVID-19 and the "Film Your Hospital" Conspiracy Theory: Social Network Analysis of Twitter Data. J Med Internet Res. 2020;22(10):e22374.

28.   Ahmed W, Vidal-Alaball J, Downing J, López Seguí F. COVID-19 and the 5G Conspiracy Theory: Social Network Analysis of Twitter Data. J Med Internet Res. 2020;22(5):e19458.

29.   Ahmed W, Vidal-Alaball J, Lopez Segui F, Moreno-Sánchez PA. A Social Network Analysis of Tweets Related to Masks during the COVID-19 Pandemic. Int J Environ Res Public Health. 2020;17(21).

30.   Vacunas y Programa de Vacunación. Coberturas de Vacunación. Datos Estadísticos. Ministerio de Sanidad. Gobierno de España.2023 [accessed 2024-01-07]. Available from: https://www.mscbs.gob.es/profesionales/saludPublica/prevPromocion/vacunaciones/calendario-y-coberturas/coberturas/home.htm.

31.   Idescat. Llars amb accés a Internet: Institut d'Estadística de Catalunya; 2022 [Accessed 28 Nov 2023]. Available from: https://www.idescat.cat/indicadors/?id=ue&n=10144.

32.   Statista. Penetración de las redes sociales en España de 2010 a 2023 2023 [Accessed 28 Nov 2023]. Available from: https://es.statista.com/estadisticas/910163/redes-sociales-porcentaje-de-usuarios-por-comunidad-autonoma-en-espana/.

Supplementary Files

Multimedia Appendixes

Untitled.
URL: http://asset.jmir.pub/assets/bb747272d4bf599c14f268fc3048fdce.docx