

Article

Misinformation messages shared via WhatsApp in Mexico during the COVID-19 pandemic: an exploratory study

Veronika J. Wirtz^{1,2}, Gabriel Millán-Garduño³, Jennifer Hegewisch-Taylor^{2,4}, Anahi Dreser², Andrea Anaya-Sanchez⁴, Tonatiuh T. González-Vázquez², Ricardo Escalera² and Pilar Torres-Pereda^{2,5}

¹Department of Global Health, Boston University School of Public Health, Boston, USA

²Center for Health Systems Research, National Institute of Public Health, Cuernavaca, Mexico

³Center for Information on Public Health Decisions, National Institute of Public Health, Cuernavaca, Mexico

⁴Graduate Group in Microbiology, University of California, Berkeley, Berkeley, California, USA

⁵PhD student in Anthropology and Communication, Universitat Rovira i Virgili, Tarragona, Spain

*Corresponding author. E-mail: jennifer.hegewisch@insp.mx

Abstract

Little is known about the role of WhatsApp in spreading misinformation during the start of the COVID-19 pandemic in Mexico. The aim of this study is to analyze the message content, format, authorship, time trends and social media distribution channels of misinformation in WhatsApp messages in Mexico. From March 18 to June 30, 2020 the authors collected all WhatsApp messages received via their personal contacts and their social networks that contained information about COVID-19. Descriptive and inferential statistics were used to analyze the scientifically inaccurate messages and the relationship between variables, respectively. Google image and video searches were carried out to identify sharing on other social media. Out of a total of 106 messages, the most frequently mentioned COVID-19 related message topics were prevention (20.0%), conspiracy (18.5%), therapy (15.4%) and origin of the virus (10.3%), changing throughout the pandemic according to users' concerns. Half of all WhatsApp messages were either images or videos. WhatsApp images were also shared on Facebook (80%) and YouTube (~50%). Our findings indicate that the design of information and health promotion campaigns requires to be proactive in adapting to the changes in message content and format of misinformation shared through encrypted social media.

Lay summary

As an encrypted social media platform with hardly accessible content, little is known about the role of WhatsApp in spreading misinformation messages (either false or misleading information) during the COVID-19 pandemic in Mexico. In this study, researchers studied the content, format, time and channel of distribution of WhatsApp messages containing information about COVID-19 collected via their personal contacts and their social networks from March 18 to June 30, 2020. Half of all messages were visually appealing and the content changed according to the population's concerns. WhatsApp messages were also distributed in several other social media platforms. Understanding the format and content of misinformation may help to design dynamic health information and promotion campaigns against it. Regulations of public social media such as Youtube can have a positive impact on WhatsApp.

Keywords: Mexico, social media, health information, COVID-19, public health

BACKGROUND

Accurate health information is relevant to protect populations from health risks (WHO, 2020). The provision of health information is part of the 10 essential public health services and responsibility of the government (CDC, 2021). It includes providing information to build knowledge, shape attitudes, support informed decision-making and develop skills and behavior for healthy living (CDC, 2021). Providing accurate health information has been described as an enabling factor to deliver public health services (WHO Europe, 2021).

The dissemination of misinformation in the last decade has threatened the provision of truthful and trusted information by government and multilateral organizations (WHO, 2020). Social media have accelerated the dissemination of information including misinformation (WHO, 2020). Artificial intelligence and the use of bots in social media are contributing factors to its dissemination (Ha et al., 2021). The number of publications studying the spread of misinformation in social media has increased over the past decade (Ha et al., 2021; van der Linden, 2022) and important lessons have been learned regarding the content and the potential drivers of health-related misinformation (Wang et al., 2019). For instance, potential drivers of misinformation are the use of unregulated social media as an information source (Allington et al., 2021), social media homophily (Schuetz et al., 2021), the low quality of health information on websites (Cuan-Baltazar et al., 2020) and users who do not sufficiently think about the content accuracy when sharing information (Pennycook et al., 2020).

The arrival of the COVID-19 pandemic heightened the awareness of the challenges of the infodemic which has been defined as ‘an overabundance of information, both online and offline’ including deliberate attempts to disseminate wrong information to undermine the public health response and advance alternative agendas of groups or individuals’ (WHO, 2021). In order to better protect the population from the continuous threat of the pandemic, it is important to study the content and the actors of spreading misinformation (WHO, 2021). A global survey studying the susceptibility to misinformation about COVID-19 found that COVID-19 related misinformation was perceived as the most reliable in Mexico and Spain compared to Ireland, USA and UK and that the susceptibility to misinformation is related to low adherence to the sanitary recommendations of the authorities (Roozenbeek et al., 2020). However, the study did not analyze the relationship between format and credibility. Also, misinformation belief was associated with lower COVID-19 knowledge scores (Lee et al., 2020) and conspiracy messages

were more widely spread when they expressed negative emotions and provided explanations (Rains et al., 2021).

For the spread of misinformation, it is necessary that the users of the messages find them important to share. The literature points out that aspects such as content, authorship and format may constitute cues for the frequency with which they are shared and their credibility (Karlova and Fisher, 2013; Wang et al., 2019). Although there are a few studies on the content, effects and motivation of misinformation during COVID-19 in Mexico in general (Fernández Poncela, 2020; Vázquez Luna, 2020), there is a gap in our understanding of the type of misinformation shared through social media, but particularly for encrypted social media such as WhatsApp, one of the largest social media platforms in Mexico.

A large number of studies on health-related misinformation have focused on YouTube, Facebook and Twitter and comparatively little is known on the use of WhatsApp to disseminate health-related misinformation (Wang et al., 2019). WhatsApp Messenger, owned by the American company Meta Platforms, Inc, is a chat application focused on sending instantaneous messages allowing the exchange of audio, visual and text messages (Dean, 2021). Since it is end-to-end encrypted and private, studying content and dissemination is more challenging than in other social media where much content is public (Gupta and Taneja, 2018). WhatsApp is very popular in Latin America and Spain (Kemp, 2019). In the case of Mexico, 88.1% of the population has at least one cell phone with internet connection functionality (INEGI, 2020) and from those, 75.7% use WhatsApp (77 million of users in 2020 according to Saavedra (2020)), with the highest percentage among 18–24 age group (84.5%) followed by 25–34 age group (80.6%) (Instituto Federal de Telecomunicaciones, 2020). Recent studies done in Mexico on the use of WhatsApp and the circulation of COVID-19 information revealed that 83% of respondents received misinformation by means of social networks particularly Facebook (41%) and WhatsApp (39%) (Ojeda de la Torre, 2020). This underlines the relevance of understanding the characteristics COVID-19 messages shared by WhatsApp.

OBJECTIVES

The aim of this study is twofold: (i) to analyze the message content, format, authorship and time trends of WhatsApp messages regarding COVID-19 in Mexico between March and June, 2020; (ii) to study the extent of sharing of these WhatsApp images and videos from other social media.

METHODS

Study design, sample and data collection

In this exploratory, observational study, six authors of the research team (PT, AD, RE, JH, GT, TG) collected all WhatsApp messages related to COVID-19 received in their personal WhatsApps from March 18th, 2020 to June 15th, 2020. In order to capture messages from as many different people as possible, authors worked as message collection nodes asking all their contacts to share with them any message they received in group or individual chats about COVID-19; attention was paid to collect information from people from different regions in Mexico and across a variety of ages (teenagers, adults and the elderly). The authors chose the period March to June as it was the start of the pandemic in Mexico where people had to adapt to a new reality and rely on information in their networks, the government or news media.

Data analysis

The same six authors collecting the messages were involved in the verification of the veracity of the messages, and posteriorly, in the classification by type of characteristics. Collected messages were divided into three blocks of 35, and were reviewed in pairs. Inter-rater agreement between the two researchers and between the three pairs was assessed by the last author (PT). In case of disagreement in the classification, consensus was reached by group discussion with all authors.

Verification of the veracity of the messages was conducted using scientific peer-reviewed articles and websites of national and international health agencies such as the World Health Organization (WHO), the Centers for Disease Control and Prevention and the Ministry of Health of Mexico and fact-checker websites. If the message was indeed found incorrect according to these information sources mentioned above, we included it in the analysis. Excluded were jokes and messages containing correct information.

Descriptive statistics were used to analyze the frequency and type of messages received. All messages were analyzed as described before aiming to identify the following characteristics of the messages: content, format, and authorship presented. The specific categories in each of these characteristics can be found in the [Supplementary Material 1](#). In brief, message content was classified into topics derived from content analysis: prevention, conspiracy, therapy, how the authority handled the pandemic, the origin of the virus, the characteristics of the virus, vaccines, effects of COVID-19, diagnosis of COVID-19 and the transmission of the virus. One message could be classified into one or more topics.

To study the relationship between the message topic and the message format and time period, we

used the two-tailed z -test. The z -test studies whether the difference between two variables is by chance. To study the relationship between message topic and time period, we recorded the message topic as a function of the week in which they reached the inboxes of the study team.

With regard to the messages format, these were classified in: only text, text and images, videos, audio, TV or public broadcast, or link to a website. The text was categorized according to the elements that make them up, such as the use of upper- and lower-case letters; the image background, text color, type of illustration, whether it is part of a series and whether or not it was a screenshot. Videos were described in terms of their lengths, whether they had a narrator, type of presentation (report, interview, discussion, news) and categorized in terms of the complexity of their production (low, medium and high). Criteria to judge complexity included number of shots (more than one camera), frames and transitions, the type of animations, text superimposed to presentation, whether they had subtitles and curtains. For example, videos which were rated 'low complexity' were filmed without any type of editing to the video. The medium and high complexity videos were those that have more editing elements including audio, video, script and were edited using software. Video background music was also analyzed because it contributed to the information it conveyed; for instance, if 'dark' type of music was selected it conveyed a feeling of intrigue, fear, suspense. Images were described in terms of their color, size, objects shown, text characteristics (font size, emojis, typography, etc.).

Finally, we tracked whether the WhatsApp images and videos included in our study were shared on other social media. We restricted our search to Twitter, Facebook, Instagram and YouTube. For images, we used the Google function 'Search by Image'; and in case of not having results, we searched for words that were part of the text of the image or keywords. Google open search was used to identify videos via generated key words: for instance, if the video showed the President of Mexico promoting a cleaning agent as a treatment of COVID-19, we used the brand name of the promoted cleaning agent in the video and the words 'president' AND 'Mexico' to search for other social media outlets in which these videos had been shared. The transcription of the opening phrase or words of the video was also searched.

In the case of Instagram, keyword searches were carried out within the social network as the posts are not registered in Google. In contrast, Facebook posts were registered in Google and hence, could be identified via a keyword search. Since almost all messages did not have dates, we were unable to know where the message originated from. Finding at least

one Whatsapp message posted on Facebook, Twitter or Instagram was classified as shared on the respective social media.

RESULTS

Tipification of message content, format and authorship

A total of 124 different messages were collected in the analysis, from which 18 messages were withdrawn after being classified as (i) jokes ($n = 4$) and (ii) identifying their claims as true content ($n = 14$), giving a final sample of 106 messages for analysis. In the 106 Whatsapp messages studied, we counted 198 times in which one of the 10 topics was mentioned. The most frequently mentioned topics in descending order were prevention, conspiracy, therapy, how the authority handled the pandemic and the origin of the virus (Figure 1). Only a low number of messages covered topics such as diagnosis of COVID-19 and the transmission of the virus.

The single largest category of message format was text, followed by video and image (Table 1). Interestingly, videos and images made up 50% of all messages. In terms of authorship, 81.1% had an author, the rest were anonymous. Anonymous messages made up about one fifth of all messages.

Time trends of topics of the WhatsApp messages

It is interesting to note that while prevention and conspiracy containing messages were received during

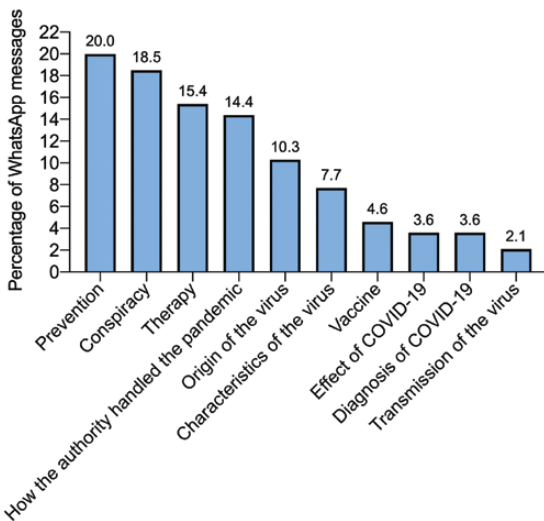


Fig. 1: Topics identified in WhatsApp messages.

the entire study time, messages regarding the origin of the virus were only received in the first 8 weeks of the study period (March 18–May 18) (Figure 2). The topic of therapy in the first 4 weeks of the study period was lower than in late April 2020 and particularly late May 2020 when the pandemic accelerated in Mexico. In Mexico, after the first month of the National Social Distancing Campaign (voluntary isolation of the population at home) in April 2020, the number of messages containing conspiracy content was particularly high. The topic of the management of the pandemic by authorities was relatively high between March 15 and April 15 and was relatively low in the month of May and the beginning of June before increasing in relative importance at the end of June during one of the peaks in terms of COVID-19 cases in Mexico. Other topics made up only a small share of the total, none being dominant. All WhatsApp messages except one did not have time, date, and geographic reference in the message.

Relationship between topic and format

With respect to the relationship between the topic and the format of the message, there were some topics that had a higher probability to be presented in a certain format. For instance, we found a higher probability of the topic of therapy to be presented as an unedited video compared to other topics, while the topic of the effects of COVID-19 and transmission had a higher probability to be represented as text (Table 2). Both results were statistically significant. Therapy and prevention as topics had a lower probability to be presented in the format of an image or a video respectively (see also Supplementary Material 1), but these were not statistically significant.

Characteristics of WhatsApp videos

The length of the videos had a mean length of 10 min and 30 s, a median length of 6 min and 49 s, a

Table 1: Format and authorship

	Frequency	Percentage %
Message format ($n = 106$)		
Text without image	37	34.9
Video	31	29.1
Image	24	22.6
Only audio	7	6.6
TV or public broadcast	4	3.8
Link to a website	3	2.8
Message authorship ($n = 106$)		
With author	86	81.13
Anonymous	20	18.7

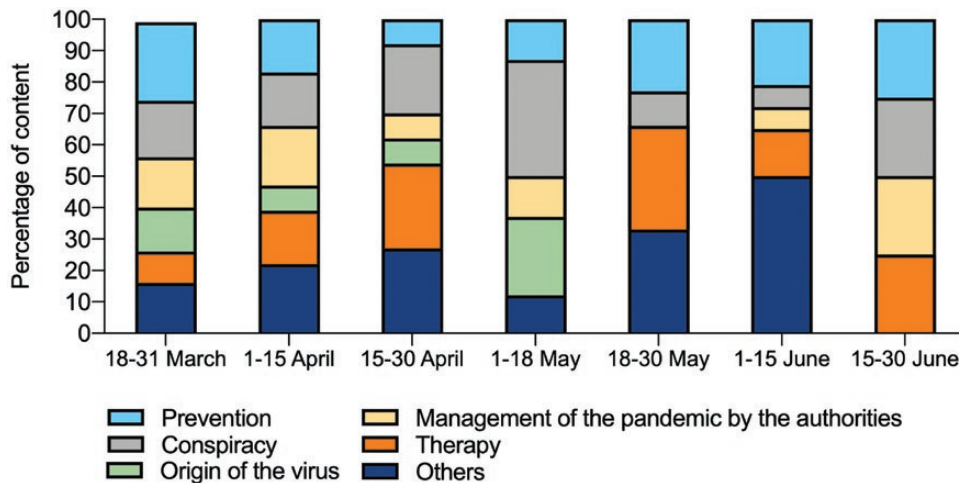


Fig. 2: Time trends of topics of the WhatsApp messages.

Table 2: Relationship between WhatsApp message topics and format

	Text	Image	Audio	Video	TV	Link to website	Total
Prevention	17	9	3	6	2	3	40
Conspiracy	12	6	2	12	3	1	36
Therapy	9	2*	2	15*	1	2	31
Management by authorities	8	7	4	8	1	0	28
Origin of the virus	5	3	0	9	2	2	21
Characteristics of the virus	6	2	0	6	1	0	15
Vaccine	2	3	0	3	1	0	9
Diagnosis	5	1	0	1	0	0	7
Effect of the virus	5*	0	0	1	0	1	7
Transmission	3*	1	0	0	0	0	4
Total	72	35	11	62	11	9	198

*Two-tailed, $p = 0.05$: there is less than a 5% probability of finding values that are farther from the population mean (population = total messages, sample = messages on the same topic).

minimum length of 30 s and a maximum length of 1 h. Testimonies ($n = 13$), interviews ($n = 5$), documentaries ($n = 5$), and reportage ($n = 4$) made up 75% of all the videos. Of the 31, the majority had a medium complexity ($n = 18$); only one video displayed a high complexity of edits. The majority of videos lacked background music.

The dominant topic of videos was therapy ($n = 11$ out of 37). Several videos showed the protagonists in their home or workspace advertising a treatment for COVID-19. Other videos showed international and national known conspiracy theorists disseminating false claims about the therapy of COVID-19. Similarly, other less well-known protagonists

appeared in white coats claiming to be medical doctors or scientists recommending harmful or ineffective therapy (e.g. blowing hot air with a hair dryer into the nose, or using bleach). Finally, there is a number of videos with imposter content where something appears to happen in the video which in fact never occurred: for instance, an edited video suggests that former President Trump promoted the use of chlorine dioxide as treatment for COVID-19. In reality, he did promote hydroxychloroquine but not chlorine dioxide. In another imposter video, the President of Mexico promotes as a cure Pinol®, a well-known house cleaning product in Mexico. In reality, he never promoted this product.

Characteristics of WhatsApp images sharing misinformation

Eight out of the 24 images comprised an image of a text; the rest included a photo or illustration. In terms of authorship, eleven of the images were anonymous; in the rest, the author's name was displayed in the message. Some of the authors were personalities such as a news anchor, and a pornstar. Regarding the date and place of origin of the images, none had a visible date and only 7 could be located somewhere. Background colors included: black (3), blue (1), green (1), gray (8) and white (10) and the texts were only presented in black (12) and white (13). In many of the images there was a text with a catchy title, capital letters or exclamation marks such as 'breaking', 'listen', 'important advice', and 'at last the truth about COVID-19 has come to light'. Photos were used to reaffirm the text: 'President of Russia Putin took 800 lions and tigers to the streets to force people to stay at home... maybe that's how they listen' (Imagen 15_Marzo). The photo showed some lions on a deserted street crossing at night. Some photos seemed to reinforce the message but were inaccurate, for example, the message 'Prime Minister of Italy with tears in his eyes tells all the people we have lost the battle against the coronavirus, we lost control, only GOD can help us PRAY FOR ITALY' (capital letters in the original message) (Imagen 17_Marzo) showed a photo of Brazilian President Bolsonaro and the Brazilian flag, instead of an Italian politician or flag.

Sharing of WhatsApp images and videos in other social media

Three images were chosen to illustrate the most common social media or source where they were shared (see [Supplementary material 2](#)): Instagram, Facebook and Twitter.

We found that the sharing of images and videos in other social media was very frequent. Twenty of the 25 WhatsApp images in our study had also been shared on Facebook at least once (see [Supplementary material 3](#)). Nearly half (46.7%) of all WhatsApp videos had also been shared on YouTube. One of the most frequently shared images was attributed to UNICEF: despite not having distinctive image characteristics of the organization (logo, range of colors, style, typography), it was shared in Facebook, Twitter, YouTube and Instagram (including in languages other than Spanish). The Spanish version was also taken up by three educational institutions in Mexico on their social media page and one Mexican municipal government because it was considered to be authentic.

Of note, in the case of images, we only found two messages marked by Facebook as false content. All others did not have a label of being false.

DISCUSSION

The COVID-19 pandemic has shown the vulnerability of information systems to misinformation and the harm that the spread of misinformation can cause. Over the past decade, technological advances have been made to detect misinformation including artificial intelligence (AI); however, their broad implementation for effective prevention from misinformation is still in its infancy ([Anderson and Rainie, 2017](#)). The WHO has called for research that helps understand the content and extent of the spread of misinformation, the main dissemination channels and effective measures to protect the population from misinformation ([WHO, 2021](#)). Our study is one of the few analyzing the misinformation shared on WhatsApp in Mexico, a country with one of the largest user communities after India, Brazil, USA, Indonesia and Russia ([Dean, 2021](#)). Our study findings support the development of actional recommendations to protect the population from harmful information.

First, our findings show that 50% of all WhatsApp messages shared were images and videos. Literature shows that particularly videos are thought to be more credible and catch more attention than images or just text ([Witze, 2021](#)). Videos are a format that has previously been widely used to spread false information about health information such as Zika, vaccines and influenza ([Basch et al., 2017](#); [Bora et al., 2018](#); [Chen et al., 2018](#)). Several studies show that fake or misleading videos for cancer and the West Nile Virus have more views and are more shared than videos based on science ([Sharma, 2017](#); [Shi, 2019](#)). It is also believed that videos are shared more frequently over text and image ([Changoiwala, 2019](#)). Furthermore, the study revealed that the video quality varied; while some videos are edited and have higher quality, the majority of videos had a low complexity. Previous Zika studies showed that videos made by common users have lower quality and are more misleading ([Bora et al., 2018](#)). Our findings also show that some topics such as therapy were much more likely to be addressed using the format of a video rather than any other message format. These findings highlight the importance of technical application to detect misinformation not only in text but equally important in images and videos.

Second, the study results show the dynamic nature of the topics changing over time in response to the trajectory of the pandemic: whereas the topic of prevention dominated the first weeks of the pandemic, therapy and how authorities handled the pandemic grew over time. Our findings are in line with studies from other geographical regions including China ([Tsao et al., 2021](#)) and Spain ([López-García et al., 2021](#)). People gradually shifting the topic from 'origin' to other relevant ones such as how authorities handled the pandemic

or the economic implications (Zhu *et al.*, 2020). Our study findings in conjunction with others point to the need for being responsive to the changing nature of the content of the misinformation dissemination over time, as well as the real-time government response needed to deal with the emergence of epidemics.

Third, our study identified a variety of techniques that messengers used to increase the credibility of their messages. For instance, many of the messages in our study were found to assimilate an authoritative tone similar to news reports and used accurate orthography. Studies have shown that such an authoritative tone is thought to be used to conceal the lack of verifiable information; the correct use of writing rules is aimed at enhancing the credibility of the message (Vereshchaka *et al.*, 2020). There were some exceptions to the rule in which messages used titles in capital letters or exclamation marks which are aimed to keep the reader engaged but are less likely to be used in credible news outlets (Vereshchaka *et al.*, 2020). Similar to our study, a study in Spain exploring the spread of misinformation via WhatsApp found that the messages, mostly in non-technical language, aimed at the general public while a quarter of the messages used technical jargon to increase their credibility (Moreno-Castro *et al.*, 2020).

Another way to increase credibility was authorship by either a scientist or health professional, or featured a protagonist with authority or pretending authority in the subject matter. Several studies have shown that social, cultural and economic factors in Mexico have contributed to medical hegemony where health professionals, particularly MDs, have a high credibility (Villanueva and Castro, 2020). Furthermore, many of the protagonists of the WhatsApp messages shared in Mexico were well-known outside of Mexico for spreading conspiracy theories which makes it evident that mitigating the risk of misinformation should be done internationally. For example, a dissident scientist and promoter of chlorine dioxide, a French microbiologist, and a medical doctor from Argentina all propagate similar conspiracies of the WHO and big pharma against humanity and against vaccines and conventional medical approved treatments.

Furthermore, some videos were very sophisticated in the way that they were edited so that the imposter content were difficult to discern. While some images and videos may appear as jokes or satire to those who know the real event (e.g. Mexican President Lopez Obrador's press conference), people unfamiliar with the truth or context may consider the content valid. Some scientists argue that even though the initial intent was satire, it should be classified as part of misinformation because of the probability of accepting it as true (Molina *et al.*, 2019). Moreover, humor is not always understood in cultural contexts different from those in

which it originated, and its messages can be interpreted as true (WHO, 2021). Knowledge on how messengers try to increase the credibility of their messages can help receivers to detect misinformation and make them more critical consumers of information.

Fourth, the fact that almost all WhatsApp messages studied had no date, time and geographical reference in them means that they can be shared even when it does not correspond to the country in which it is distributed and can be sent again at different points in time without losing temporal validity. Considering that plain text messages consume few storage bits for forwarding, they can be highly disseminated, unlike images and videos. Plain text messages do not require special skills or software to compose them. Authorities addressing misinformation need to be aware of the fact that the absence of date, time and geographical reference will make it more difficult to identify misinformation and its origin.

Fifth, our study found the misinformation content shared as image or video on WhatsApp was also widely shared on other social media platforms. This has important implications for the detection and prevention of misinformation as interventions should be focused on a series of social media channels simultaneously and not one in isolation. The study of messages in other social media channels may help us to identify the otherwise hardly accessible content, of WhatsApp message content. Moreover, preventing the spread of misinformation needs to take a system approach where social media is considered an ecosystem where one social media channel is interrelated with others. Studies have shown that the tighter regulation of Facebook posts have had an effect on the number of misinformation spread (Chiou and Tucker, 2018). Given the study findings of how the misinformation in WhatsApp is also shared via other social media channels, we would expect that regulation of other social media would also have an effect on the sharing of misinformation via WhatsApp.

Sixth, while conspiracy theories were subjacent to other message content, in almost a fifth of our sample the main content of the messages was conspiracy theories themselves. Many of them coincide with those identified in other countries, for example, 5G (Meese *et al.*, 2020). However, we found conspiracy theories that appear to be specific to the Mexican context such as the alleged order of the president to hide COVID-19 cases (Latinus, 2021) or health care personnel making people sick on purpose (Palos, 2020). Some of these messages, as well as the ones related with how the authority handled the pandemic, could be linked to political interests against the current government within a context of a particularly polarized political climate in Mexico.

Our study results and subsequent recommendations mentioned above can help design policies to counteract the harmful effects of misinformation. For instance, it has been found that prebunking misinformation can protect people: if individuals are warned beforehand about a falsehood there are more resistant to believe it when it appears in social media (Witze, 2021). Since the same misinformation is widely circulated on different social media platforms, a surveillance system can provide early warnings and disseminate prebunking of those falsehoods, thus contributing to the creation of health and information literacy (Al Khalid *et al.*, 2018). Some countries have created non-governmental organizations specialized to counteract misinformation spread on social media including WhatsApp. For instance, the organization Metafact's in India is using nature language process, (also called NLP) to help identify and counteract misinformation (Changoiwala, 2019). However, checking misinformation spread on video or images maintains an ongoing challenge for AI to identify. Other groups have created a software called *WhatsApp Monitor*, web-based system that helps researchers and journalists explore the nature of content shared on WhatsApp public groups and measure the extent of misinformation spread (Melo *et al.*, 2019). Some initiatives in Latin America and Spain (Latam Chequea Coronavirus) have created repositories where users can find misinformation (Animal Político, 2020), however, these pages require a minimum of information literacy, which older populations may not possess.

Apart from non-governmental organizations, governments have the key responsibility to protect citizens from harmful information that can threaten their health and wellbeing and prevent misinformation from being disseminated. The Mexican Government launched an initiative 'Susana Distancia' which was intended to inform citizens about the pandemic and how they can protect them against harm. Citizens were able to submit queries in the form of text messages to receive information about COVID-19 and preventative measures. However, the authors felt that tool is not flexible enough to respond to more nuanced questions and left users frustrated with the lack of useful responses to their queries. Furthermore, since our study show that some misinformation messages not shared in other social media besides WhatsApp, the Mexican authorities must conduct surveillance of all social media including WhatsApp to design messages to counteract misinformation and perform appropriate health promotion.

When discussing the study findings, it is important to consider the study limitations. This is an exploratory, observational study with the intent to generate hypotheses which limits the generalizability of the

findings. A convenient sample of WhatsApp messages in one country in a specific time frame from a network of public health researchers and their social contacts was used. Since WhatsApp messages are encrypted and usually not shared on public platforms, messages can only be studied within a certain network of users. As Javed *et al.* (2022) noted, since WhatsApp is a closed network without any official access for analyzing its content, most studies use a self-selection sampling where individuals voluntarily share messages for research purposes which may limit external validity. The sender of a message is not necessarily the author as messages are often forwarded which prevents studying the frequencies with which the messages are shared and socio-demographic data of the author of the message. We did not study the effect that the messages had on the receivers or the intent that the author had when writing the message, nor the relationship between the messages and the person receiving and resending them. As many messages do not have a date or author, we cannot know the origin of all or the temporality of all.

CONCLUSIONS

This study fills a gap in misinformation dissemination through WhatsApp in Mexico and can support the development of prevention policies in Mexico and possibly other countries. People need to be able to access trustable sources and distinguish truthful information from misinformation. This study shows that WhatsApp messages play a significant role in fueling misinformation about COVID-19 in the early phase of the pandemic and contributed to the mistrust in the authorities. Future research should focus on the development of culturally appropriate communication strategies that address misinformation and proactively adapting to the changes in message content, format of the misinformation shared as well as distribution channels. Strengthening health communication in Mexico is critical to enhance the country's capacity to deal with the current and future pandemics.

Supplementary Material

Supplementary material is available at *Health Promotion International* online.

Supplementary material 1: Table with classification of message content, format, and authorship

Supplementary material 2: Examples of images shared on WhatsApp during the first months of the COVID-19 pandemic in Mexico

Supplementary material 3: Number of times any message was identified as posted on a particular social media at least once

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CONFLICT OF INTEREST

The authors do not declare any conflict of interest.

ETHICS APPROVAL

The study was approved by the Institutional Review Board of the National Institute of Public Health in Mexico (project No. 1697).

AUTHOR CONTRIBUTIONS

VW conceived the idea of the paper. Acquisition of data was executed by PT, AD, RE, TG, JH and GM. GM collected information on social media channels with the help of AA. VW, GM, AA, AD, PT, TG and JH analyzed, organized and interpreted the data. Statistical analysis was executed by VW. VW drafted the first version of the manuscript. Critical revision of the manuscript for intellectual content was made by VW, JH, AD, PT, GM. Everyone provided technical support and contributed with subsequent manuscript versions.

REFERENCES

Al Khalid, K.A., Al Khaja, A.K., and Sequeira R.P. (2018) Drug information, misinformation, and disinformation on social media: a content analysis study. *Journal of Public Health Policy*, 39, 343–357, doi: [10.1057/s41271-018-0131-2](https://doi.org/10.1057/s41271-018-0131-2).

Allington, D., Duffy, B., Wessely, S., Dhavan, N. and Rubin, J. (2021) Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. *Psychological Medicine*, 51, 1763–1769.

Anderson, J. and Rainie, L (2017) *The Future of Truth and Misinformation Online*. Pew Research Center. Available at: <https://www.pewresearch.org/internet/2017/10/19/the-future-of-truth-and-misinformation-online/> (last accessed 29 May 2021).

Basch, C. H., Zyburt, P., Reeves, R. and Basch, C. E. (2017) What do popular YouTube videos say about vaccines? *Child: Care, Health and Development*, 43, 499–503, doi:[10.1111/cch.12442](https://doi.org/10.1111/cch.12442).

Bora, K., Das, D., Barman, B. and Borah, B. (2018) Are internet videos useful sources of information during global public health emergencies? A case study of YouTube videos during the 2015–16 Zika virus pandemic. *Pathogens and Global Health*, 112(6), 320–328, doi:[10.1080/20477724.2018.1507784](https://doi.org/10.1080/20477724.2018.1507784).

Center for Disease Control and Prevention. (2021) *The 10 Essential Public Health Services. An Overview*. Available at: <https://www.cdc.gov/publichealthgateway/publichealth-services/pdf/essential-phs.pdf> (last accessed 6 March 2021).

Changoiwala, P. (2019) *Can an AI Fact-Checker Solve India's Fake News Problem?* *Science. The Wire*. <https://science.the-wire.in/economy/tech/can-an-ai-fact-checker-solve-indias-fake-news-problem/> (last accessed 18 July 2020).

Chen, B., Shao, J., Liu, K., Cai, G., Jiang, Z., Huang, Y. et al. (2018) Does eating chicken feet with pickled peppers cause avian influenza? Observational Case Study on Chinese Social Media During the Avian Influenza A (H7N9) outbreak. *JMIR Public Health and Surveillance*, 4, e32, doi:[10.2196/publichealth.8198](https://doi.org/10.2196/publichealth.8198).

Chiou, L. and Tucker, C. (2018) *Fake News and Advertising on Social Media: A Study of the Anti-vaccination Movement*. NBER Working Paper Series. National Bureau of Economic Research. Available at: <https://www.nber.org/papers/w25223> (last accessed 18 July 2020).

Cuan-Baltazar, J. Y., Muñoz-Perez, M. J., Robledo-Vega, C., Pérez-Zepeda, M. F. and Soto-Vega, E. (2020) Misinformation of COVID-19 on the internet: Infodemiology study. *JMIR Public Health and Surveillance*, 6, e18444.

Dean, B. (2021) *WhatsApp 2021 User Statistics: How Many People Use WhatsApp?* Available at: <https://backlinko.com/whatsapp-users> (last accessed 29 May 2021).

Fernández Poncela, A. M. (2020) Infodemia: rumors, fake news, myths. *Sintaxis*, (1), 35–55, doi:[10.36105/stx.2020edespcovid-19.02](https://doi.org/10.36105/stx.2020edespcovid-19.02).

Gupta, H., Taneja, H (2018). WhatsApp has a fake news problem—that can be fixed without breaking encryption. *Columbian Journalism Review*, August 23. Available at: https://www.cjr.org/tow_center/whatsapp-doesnt-have-to-break-encryption-to-beat-fake-news.php (last accessed 18 July 2020).

Ha, L., Perez, L. A. and Ray, R. (2021) Mapping recent development in scholarship on fake news and misinformation, 2008 to 2017: disciplinary contribution, topics and impact. *American Behavioral Scientist*, 65, 290–315, doi:[10.1177/0002764219869402](https://doi.org/10.1177/0002764219869402).

INEGI. (2020) Estadística a propósito del día mundial del internet (17 de mayo) datos nacionales. *Comun prensa núm 216/20*, 1–8, https://www.inegi.org.mx/contenidos/salade-prensa/aproposito/2020/eap_internet20.pdf (accessed 23 June 2021).

Instituto Federal de Telecomunicaciones. (2020) *Primera Encuesta 2020, Usuarios de Servicios de Telecomunicaciones*. Available at: <http://www.ift.org.mx/sites/default/files/contenidogeneral/usuarios-y-audiencias/primerencuesta2020.pdf> (accessed 23 June 2021).

Javed, R. T., Usama, M., Iqbal, W., Qadir, J., Tyson, G., Catsro, I. et al. (2022) A deep dive into COVID-19-related messages on WhatsApp in Pakistan. *Social Network Analysis and Mining*, 12, 1318.

Karlova, N. A. and Fisher, K. E. (2013) A social diffusion model of misinformation and disinformation for understanding human information behaviour. *Information Resesarch*, 18(1), 573 (Available at <http://InformationR.net/ir/18-1/paper573.html>)

Kemp S. (2019) *Digital 2019: Global Digital Overview*. <https://datareportal.com/reports/digital-2019-global-digital-overview> (last accessed 23 June 2021).

Latinus. (2021) *Gobierno de AMLO 'es criminal al ocultar número real de muertes por COVID': Zambrano*. (Last accessed 18 July 2021).

- Lee, J. J., Kang, K. A., Wang, M. P., Zhao, S. Z., Wong, J. Y. H., O'Connor, S. et al. (2020) Associations between COVID-19 misinformation exposure and belief with COVID-19 knowledge and preventive behaviors: Cross-sectional online study. *Journal of Medical Internet Research*, 22, e22205.
- López-García, X., Costa-Sánchez, C. and Vizoso, A. (2021) Journalistic fact-checking of information in pandemic: stakeholders, hoaxes, and strategies to fight disinformation during the COVID-19 Crisis in Spain. *International Journal of Environmental Research and Public Health*, 2918, 1227, doi:10.3390/ijerph18031227.
- Meese, J., Frith, J. and Wilken, R. (2020) COVID-19, 5G conspiracies and infrastructural futures. *Media International Australia*, 177, 30–46, doi:10.1177/1329878X20952165.
- Melo, P., Messias, J., Resende, G., Garimella, K., Almeida, J., and Benevenuto, F. (2019). WhatsApp monitor: a fact-checking system for WhatsApp. Proceedings of the International AAAI Conference on Web and Social Media, 13, 676–677. Retrieved from <https://ojs.aaai.org/index.php/ICWSM/article/view/3271> (accessed 23 June 2021).
- Molina, M. D., Sundar, S. S., Le, T. and Lee, D. (2019) 'Fake news' is not simply false information: a concept explication and taxonomy of online content. *American Behavioral Scientist*, 65(2), 180–212, doi:10.1177/0002764219878224.
- Moreno-Castro, C., Vengut-Climent, E., Cano-Orón, L. and Mendoza-Poudereux, I. (2020) Exploratory study of the hoaxes spread via WhatsApp in Spain to prevent and/or cure COVID-19. *Gaceta Sanitaria*, 31, S021330195–S021391113, doi:10.1016/j.gaceta.2020.07.008.
- Ojeda de la Torre, I. (2020). López-Gatell es el más creíble en tema del COVID-19; la oposición a AMLO gana en fake news: estudio. *SimEmbargo*. <https://www.sinembargo.mx/28-03-2020/3756101> (published March 28, 2020).
- Palos, M. (2020). 'Planes Spray the City at Night': COVID-19 Conspiracy Theories in Mexico's Motor Town. *Guardian*. <https://www.theguardian.com/global-development/2020/may/22/planes-spray-the-city-at-night-covid-19-conspiracy-theories-in-mexicos-motor-town> (last accessed 22 May 2020).
- Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G. and Rand, D. G. (2020) Fighting COVID-19 misinformation on social media: experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science*, 31, 770–780.
- Animal Político, A. (2020). *LatamChequea Coronavirus, la plataforma para verificar información en 15 países*. <https://www.animalpolitico.com/2020/04/latamchequea-coronavirus-una-plataforma-para-verificar-informacion-en-15-paises> (accessed 23 June 2021).
- Rains, S. A., Leroy, G., Warner, E. L. and Harber, P. (2021) Psycholinguistic markers of COVID-19 conspiracy tweets and predictors of tweet dissemination. *Health Communication*, 38(1), 21–30.
- Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L. J., Recchia, G. et al. (2020) Susceptibility to misinformation about COVID-19 around the world. *Royal Society of Open Science*, 147, 201199, doi:10.1098/rsos.201199.
- Saavedra, D. (2020). *Utilizan WhatsApp 77 millones de mexicanos—Gaceta UNAM*. *Gaceta UNAM*. <https://www.gaceta.unam.mx/utilizan-whatsapp-77-millones-de-mexicanos/> (published 2020) (accessed 23 June 2021)
- Schuetz, S. W., Sykes, T. A. and Venkatesh, V. (2021) Combating COVID-19 fake news on social media through fact checking: antecedents and consequences. *European Journal of Information Systems*, 30, 1–13.
- Sharma, M., Yadav, K., Yadav, M. and Ferdinand, K. C. (2017) Zika virus pandemic—analysis of Facebook as a social media health information platform. *American Journal of Infection Control*, 45, 301–302, doi:10.1016/j.ajic.2016.08.022.
- Shi, S., Brant, A. R., Sabolch, A. and Pollom, E. (2019) False news of a cannabis cancer cure. *Cureus*, 11, e3918, doi:10.7759/cureus.3918.
- Tsao, S. F., Chen, H., Tisseverasinghe, T., Yang, Y., Li, L. and Butt, Z. A. (2021) What social media told us in the time of COVID-19: a scoping review. *The Lancet Digital Health*, 3, e175–e194, doi:10.1016/S2589-7500(20)30315-0.
- van der Linden, S. (2022) Misinformation: susceptibility, spread, and interventions to immunized the public. *Nature Medicine*, 28, 460–467.
- Vázquez Luna, J. L. (2020) Las noticias falsas (fake news), la desinformación y la infodemia durante la pandemia de la COVID-19. *Sintaxis*, (1), 189–202, doi:10.36105/stx.2020edespcovid-19.09.
- Vereshchaka, A., Cosimini, S. and Dong, W. (2020) Analyzing and distinguishing fake and real news to mitigate the problem of disinformation. *Computational and Mathematical Organization Theory*, 26, 350–364.
- Villanueva, M. and Castro, R. (2020) Hierarchy systems of the medical field in Mexico: a sociological analysis. *Ciencia Saúde Coletiva*, 25(6), 2377–2386, doi:10.1590/1413-81232020256.28142019.
- Wang, Y., McKee, M., Torbica, A. and Stuckler, D. (2019) Systematic literature review on the spread of health-related misinformation on social media. *Social Science and Medicine*, 240, 112552, doi:10.1016/j.socscimed.2019.112552.
- Witze A (2021). How to detect, resist and counter the flood of fake news. *Science News*. Available at: <https://www.sciencenews.org/article/fake-news-misinformation-covid-vaccines-conspiracy> (last accessed 6 May 2021).
- World Health Organization. (2020) *Managing the COVID-19 Infodemic: Promoting Healthy Behaviours and Mitigating the Harm From Misinformation and Disinformation*. Available at: <https://www.who.int/news/item/23-09-2020-managing-the-covid-19-infodemic-promoting-healthy-behaviours-and-mitigating-the-harm-from-misinformation-and-disinformation> (last accessed 6 March 2021).
- World Health Organization. (2021) *WHO Public Health Research Agenda for Managing Infodemics*. Geneva: World Health Organization; 2021. Licence: CC BY-NC-SA 3.0 IGO. Available at: <https://www.who.int/publications/item/9789240019508> (last accessed 29 May 2021).
- World Health Organization Europe. (2021) *The 10 Essential Public Health Operations*. Available at: <https://www.euro.who.int/en/health-topics/Health-systems/public-health-services/policy/the-10-essential-public-health-operations> (accessed 23 June 2021).
- Zhu, B., Zheng, X., Liu, H., Li, J. and Wang, P. (2020) Analysis of spatiotemporal characteristics of big data on social media sentiment with COVID-19 epidemic topics. *Chaos, Solitons and Fractals*, 140, 110123, doi:10.1016/j.chaos.2020.110123.