



Understanding the COVID-19 Infodemic: Analyzing User-Generated Online Information During a COVID-19 Outbreak in Vietnam

Ha-Linh Quach^{1,2}, Thai Quang Pham^{1,3}, Ngoc-Anh Hoang^{1,2}, Dinh Cong Phung⁴, Viet-Cuong Nguyen⁵, Son Hong Le⁶, Thanh Cong Le⁷, Dang Hai Le¹, Anh Duc Dang⁸, Duong Nhu Tran⁸, Nghia Duy Ngu¹, Florian Vogt^{2,9,*}, Cong-Khanh Nguyen^{1,10,*}

¹Department of Communicable Diseases Control, National Institute of Hygiene and Epidemiology, Hanoi, Vietnam

²National Centre for Epidemiology and Population Health, Research School of Population Health, College of Health and Medicine, Australian National University, Canberra, Australia

³Department of Biostatistics and Medical Informatics, School of Preventive Medicine and Public Health, Hanoi Medical University, Hanoi, Vietnam

⁴National Agency for Science and Technology Information, Ministry of Science and Technology, Hanoi, Vietnam

⁵HPC Systems Inc., Tokyo, Japan

⁶CMetric JSC Inc., Hanoi, Vietnam

⁷INFORE Technology Inc., Hanoi, Vietnam

⁸National Institute of Hygiene and Epidemiology, Hanoi, Vietnam

⁹The Kirby Institute, University of New South Wales, Sydney, Australia

¹⁰Field Epidemiology Training Program, National Institute of Hygiene and Epidemiology, Hanoi, Vietnam

Objectives: Online misinformation has reached unprecedented levels during the coronavirus disease 2019 (COVID-19) pandemic. This study analyzed the magnitude and sentiment dynamics of misinformation and unverified information about public health interventions during a COVID-19 outbreak in Da Nang, Vietnam, between July and September 2020. **Methods:** We analyzed user-generated online information about five public health interventions during the Da Nang outbreak. We compared the volume, source, sentiment polarity, and engagements of online posts before, during, and after the outbreak using negative binomial and logistic regression, and assessed the content validity of the 500 most influential posts. **Results:** Most of the 54,528 online posts included were generated during the outbreak ($n = 46,035$; 84.42%) and by online newspapers ($n = 32,034$; 58.75%). Among the 500 most influential posts, 316 (63.20%) contained genuine information, 10 (2.00%) contained misinformation, 152 (30.40%) were non-factual opinions, and 22 (4.40%) contained unverifiable information. All misinformation posts were made during the outbreak, mostly on social media, and were predominantly negative. Higher levels of engagement were observed for information that was unverifiable (incidence relative risk [IRR] = 2.83; 95% confidence interval [CI], 1.33–0.62), posted during the outbreak (before: IRR = 0.15; 95% CI, 0.07–0.35; after: IRR = 0.46; 95% CI, 0.34–0.63), and with negative sentiment (IRR = 1.84; 95% CI, 1.23–2.75). Negatively toned posts were more likely to be misinformation (odds ratio [OR] = 9.59; 95% CI, 1.20–76.70) or unverified (OR = 5.03; 95% CI, 1.66–15.24). **Conclusions:** Misinformation and unverified information during the outbreak showed clustering, with social media being particularly affected. This in-depth assessment demonstrates the value of analyzing online “infodemics” to inform public health responses.

Keywords: Sentiment Analysis, Social Media, Infodemic, COVID-19, Vietnam

Submitted: September 23, 2021, **Revised:** 1st, March 7, 2022; 2nd, May 15, 2022; 3rd, July 8, 2022, **Accepted:** August 2, 2022

Corresponding Author

Ha-Linh Quach

Department of Communicable Disease Control, National Institute of Hygiene and Epidemiology, Hanoi, Vietnam. Tel: +84-966-001-080, E-mail: u7062716@alumni.anu.edu.au (<https://orcid.org/0000-0001-7160-8329>)

*These authors contributed equally to this work.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

I. Introduction

Since December 2019, the coronavirus disease 2019 (COVID-19) epidemic has incurred a significant health and economic burden worldwide. Even before severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was identified as its causal agent, COVID-19-related information was already spreading uninhibitedly over traditional and social media platforms at a strikingly rapid pace [1]. This phenomenon, called an “infodemic”—the over-abundance of information regarding an emerging event [2]—has been observed during prior public health emergencies [3-5]. During the COVID-19 pandemic, infodemics have reached unprecedented levels, and governments, and public health organizations around the world are now calling for measures to limit their effects [6].

Vietnam was one of the first countries that reported COVID-19 cases outside of mainland China. Very early into the epidemic, many interventions were put in place before any cases were reported, including border closures and travel restrictions, extensive case detection and contact tracing, and stringent quarantine measures. A series of public health interventions were gradually and promptly imposed to control the outbreak domestically [7], helping Vietnam to achieve 99 days without community transmission between April and late July 2020 (Figure 1). On July 25, 2020, a surge of unlinked COVID-19 cases without evidence of imported infection from abroad was spotted in Da Nang, a municipal city in Central Vietnam of high importance for foreign trade and tourism [8]. The outbreak affected mostly patients and staff linked to several hospitals, people in the community of Da Nang, and sporadic cases among people in other prov-

inces with a travel history to Da Nang. Overall, more than 500 cases were reported in relation to this outbreak, with 495 (89.84%) in Da Nang, and 56 (10.16%) in 10 other provinces and cities across Vietnam. It was the first outbreak in Vietnam with COVID-19-related deaths since the beginning of the pandemic, with a total of 35 related fatalities. By the end of August 2020, the Da Nang outbreak was declared under control.

A series of public health measures were implemented by Vietnam government in response to this outbreak to limit and prevent further spread [7]. While some of these measures had already been introduced before the outbreak, the scope and enforcement were greatly intensified due to the urgency of this outbreak. To keep the public informed, information about these policies was frequently broadcasted by governmental agencies and various other outlets, including online platforms. Many of the estimated 69 million internet users in Vietnam (more than 73% of the population) use online media and user-generated online information as their main source of information about current events [9]. It is therefore important to recognize online information as a crucial means of communicating about health and risk during a COVID-19 outbreak, and to understand its capacity to impact public adherence to public health interventions.

Most online platforms do not fact-check user-generated online information, which creates an opportune environment for misinformation, defined in the field of public health as a “claim of fact that is currently false due to a lack of scientific evidence” [10], to spread widely with no curation or verification. The viral ability of misinformation becomes amplified by the rapidly reciprocating nature of internet, and misinformation is easily content-tailored for

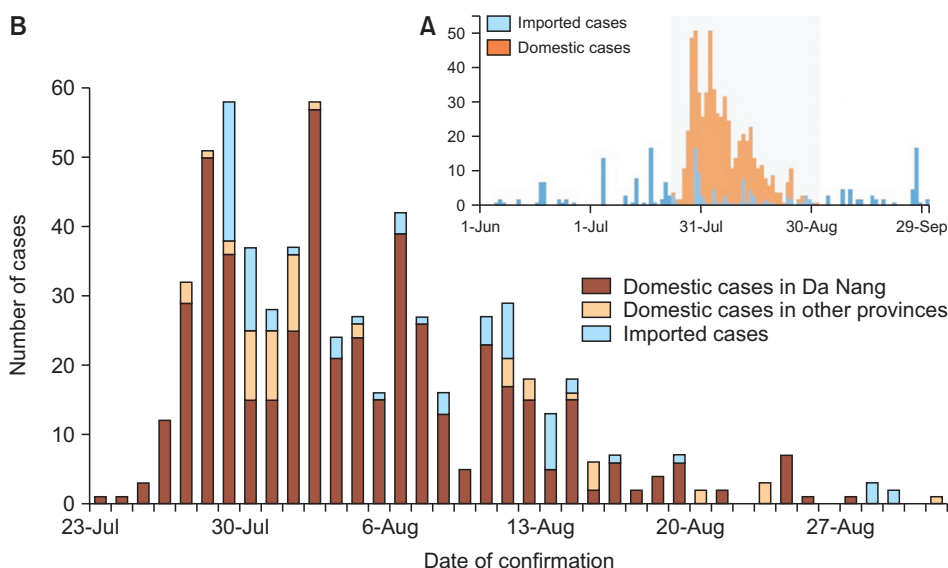


Figure 1. (A) Epidemic curve of the COVID-19 epidemic in Vietnam from June to August 2020. The shaded area indicates the outbreak period in Da Nang. (B) Epidemic curve of the COVID-19 outbreak in Da Nang, Vietnam, from July 25 to August 31, 2020. COVID-19: coronavirus disease 2019.

specific target audiences that are receptive to specific types of misinformation [11]. In addition, the interplay between sentiment and misinformation to spread certain agendas is a serious concern of researchers and policy-makers [12,13]. Existing evidence suggests that strong and polarized sentiments in online content play an important role in amplifying and driving the spread of misinformation [14,15]. Nonetheless, the sentiment profile of misinformation surrounding COVID-19 is still inconclusive. It has been recognized that online misinformation during the COVID-19 pandemic can create either a false sense of security or threat, and thus impact public health prevention and control efforts [16]. Existing evidence about online misinformation about COVID-19 is restricted to the evolution of the pandemic, conspiracy theories, or discriminative misinformation [17,18]. However, the infodemics surrounding specific public health interventions imposed by governments have not been studied. The World Health Organization recently published a research agenda to improve evidence-based tools, methods, and interventions for infodemics management [19]. This agenda highlighted the need to first measure and detect the spread and impact of misinformation in a localized context during the ongoing epidemic. In line with these priorities, we aimed to analyze the magnitude and sentiment dynamics of misinformation and unverified user-generated online information about five distinct public health interventions in response to the COVID-19 outbreak in Da Nang, Vietnam, between July and September 2020.

II. Methods

1. Study Design

This longitudinal study used publicly available online information about COVID-19-related public health interventions from July 1 to September 15, 2020 on popular online platforms in Vietnam. We divided the study period according to three phases: pre-outbreak (July 1–24 2020), during the outbreak (July 25 to August 31, 2020), and post-outbreak (September 1–15, 2020). The “during” period was defined as extending from the date of the first laboratory-confirmed case in Da Nang to the date the outbreak was declared over by the Vietnam Ministry of Health. The other periods were chosen conveniently relative to the “during” period as time intervals that also contained relevant discussions about the interventions.

The study topics included the five main public health interventions, which came as directives from national-level Ministry of Health in response to this outbreak and were

implemented at a national scale during the study period:

- (1) Cordon sanitaire of outbreak areas including the city of Da Nang and a nearby province;
- (2) Re-scheduling of the national high school examination nationwide and the exclusion from the examination of students from outbreak areas and/or students who were identified as COVID-19 cases or close contacts of COVID-19 cases;
- (3) Nationwide campaign to promote the use of Bluezone, Vietnam’s official contact tracing mobile phone application for COVID-19;
- (4) National tracking and prosecution of illegal border-crossing and individuals who breached COVID-19 quarantine requirements; and
- (5) National contact tracing, serologic testing, and quarantine for all people with a travel history involving the outbreak areas.

2. Data Collection

The inclusion criteria for data collected in the analysis were contents (1) made in public mode during the study period that remained public at time of data collection; (2) made and posted in the format of online newspapers, online forums or social media posts; and (3) had a verified postal area of operation in Vietnam. Our search was limited to online posts in Vietnamese. Data were provided by the Ministry of Science and Technology through an internet archive database named the Social Media Command Center. The research team developed keywords for data collection that were compiled into a list (Supplementary Table S1). We then used the keywords in the system to collect online information from several sources (including social media platforms, online forums, and online newspapers) operating in Vietnam (Supplementary Table S2). Based on the topics identified using the keywords, the following variables of online information were collected: source, period, number of engagements, influence score, and text (definitions in Supplementary Tables S3 and S4). The influence score was categorized into 10 categories according to the number of followers and/or views of the source of the post (Supplementary Table S4). The number of engagements was defined by the sum of all interactions with the posts by users/readers in response to the content of the post. The collected data were compiled and extracted into Microsoft Excel by personnel from the Ministry of Science and Technology, and no identifiable data were collected.

3. Data Processing

We selected the 100 posts with the highest number of en-

gements from each of the five topics, which yielded 500 posts for analysis. From these 500 selected posts, the following variables were manually collected/categorized by the research team and further processed. First, we conducted a content classification to categorize these posts based on textual content into the following categories: genuine information, misinformation, opinions, and unverified information. These categories were adapted from previous research by Kouzy et al. [20] on COVID-19 Twitter data and defined in Table 1. Two researchers independently followed the definition to categorize all selected posts separately by reading each post's textual content. The final results were cross-

checked between two sets, and any disputes were handled through consultation with a third researcher and an additional information search. No posts were excluded after the process due to an unresolved dispute.

Next, we conducted sentiment classification based on the textual content of selected posts. All selected posts underwent for sentiment analysis using a Vietnamese sentiment lexicon [21] and VnCoreNLP packages [22] for Vietnamese-language word and sentiment processing in Python 3.6. For each post, the number of positive and negative words that appeared was calculated. Each post was further classified into one of three sentiment categories (positive, neutral, or

Table 1. Definitions of content categories

Category	Definition
Genuine information	Posts expressed information that cross-matched with the information presented by Official Guideline of Vietnam Ministry of Health, official news outlet from Vietnam government, World Health Organization, and/or at least two peer-reviewed scientific journals.
Misinformation	Posts expressed information that was easily refuted by at least one of abovementioned references.
Opinions	Posts expressed an opinion and did not relay any novel information.
Unverified information	Posts expressed information that could not be proven correct or incorrect by the references.

Table 2. Descriptive characteristics of all collected online posts on non-pharmacological interventions during the COVID-19 outbreak in Vietnam from July to September 2020

Variable	Topic				
	Cordon sanitaire (n = 10,099)	National high school examination (n = 3,529)	Bluezone application (n = 16,769)	COVID-19 quarantine breach (n = 22,170)	National contact tracing (n = 1,961)
Number of posts per day	131.16	45.83	254.07	287.92	33.61
Number of engagements per posts	101.38	36.82	54.70	180.81	86.17
Influence score per post	5.00 ± 3.38	4.59 ± 3.21	3.91 ± 3.02	4.31 ± 3.28	4.71 ± 3.27
Number of posts per source					
Social media	1,647 (16.31)	438 (12.41)	5,322 (31.74)	6,435 (28.62)	191 (9.97)
Online forum	318 (3.15)	49 (1.39)	3,511 (20.94)	4,588 (20.69)	40 (2.09)
Online newspaper	8,134 (80.54)	3,042 (86.02)	7,936 (47.33)	11,237 (50.69)	1,685 (87.94)
Number of posts per period					
Pre-outbreak	1,216 (12.04)	703 (19.92)	47 (0.28)	2,351 (10.60)	8 (0.42)
During outbreak	8,224 (81.43)	2,471 (70.02)	14,684 (87.57)	18,874 (85.13)	1,782 (93.01)
Post-outbreak	659 (6.53)	355 (10.06)	2,038 (12.15)	945 (4.26)	126 (6.58)
Number of posts per day per period					
Pre-outbreak	50.67	29.29	1.96	97.96	0.33
During outbreak	216.42	65.03	386.42	496.68	46.89
Post-outbreak	43.93	23.67	135.87	63.00	8.4

Values are presented as mean ± standard deviation or number (%).

COVID-19: coronavirus disease 2019.

negative) based on an automatic calculation of the sentiment score of each post.

4. Data Analysis

The variables were summarized and plotted chronologically by date of the posts according to the timeline of the outbreak and/or interventions implemented and differentiated between content categories by appropriate statistical tests. We used negative binomial regression to explore the relationship between the number of engagements and selected posts' characteristics; the incidence relative risk (IRR) and 95%

confidence interval (CI) were calculated. Univariate and multivariable logistic regression were used to explore the association between posts' characteristics and posts' categories, focusing on misinformation and unverified information versus other post categories; for this analysis, the odds ratio (OR) and 95% CI were reported. The median number of negative and positive words stratified by posts' characteristics were summarized and differentiated using analysis of variance. All analyses were performed using Stata version 16 (StataCorp, College Station, TX, USA).

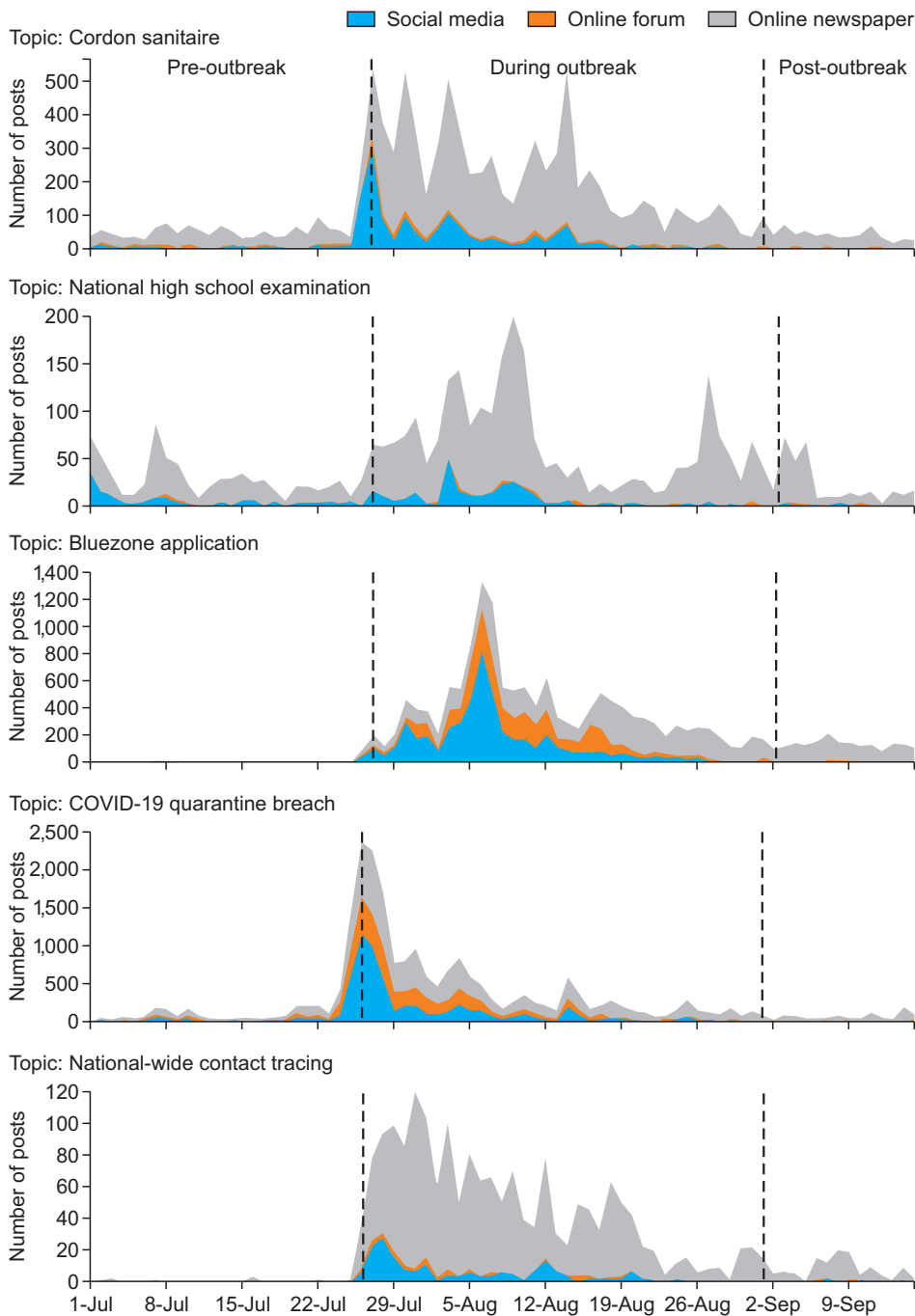


Figure 2. Distribution of online information across the time periods of the outbreak stratified by the five NPI-related topics. NPI: non-pharmacological intervention, COVID-19: coronavirus disease 2019.

5. Ethics

We obtained approval from the Australian National University's human research ethics committee (Protocol No. 2020/605) and the Vietnam National Institute of Hygiene and Epidemiology's Institutional Review Board (No. NIHE IRB-29/2020) for this research.

III. Results

1. Descriptive Characteristics of All Collected Online Posts

Table 2 and Figure 2 display the distribution of online posts' characteristics stratified by search topics for five distinct non-pharmacological interventions implemented in Vietnam during the study outbreak. Across the five topics, the "COVID-19 quarantine breach" discussion had the highest number of posts ($n = 22,170$), the highest number of posts made per day (287.92 posts per day), and the highest median of engagements (180.81 engagements per post). Meanwhile, posts concerning the "national high school examination schedule" had the fewest engagements (36.82 engagements per post), and posts mentioning "nationwide contact tracing" were posted least frequently during the study period (33.61 posts per day). While "cordon sanitaire" information was posted by sources with the highest influence scores

compared to the other topics (5.00 ± 3.38), the opposite was true for information about the "Bluezone application" (3.91 ± 3.02). Figure 2 shows that while the highest numbers of posts were made during outbreak for all five topics, online newspapers were the most consistent source reporting the highest number of posts about these topics. The number of posts about the "Bluezone application," although it had the lowest traffic before the outbreak, increased drastically during the outbreak and even reached the highest rank among the five topics after the outbreak (Table 2). A similar trend was observed for posts on "nationwide contact tracing." In contrast, the remaining three topics saw a considerable decline in the number of posts after the outbreak, reaching a volume that was even lower than before the outbreak.

2. Descriptive and Analyzed Characteristics of Selected Online Posts

Table 3 presents the characteristics of all selected posts stratified by the posts' categories. Among the selected 500 posts with the highest number of engagements, there were 316 (63.20%) genuine information posts, 10 (2.00%) misinformation posts, 152 (30.40%) opinions, and 22 (4.40%) unverified posts. The highest number of engagements was observed for unverified information (median, 13,415; interquartile range,

Table 3. Descriptive characteristics of selected online posts stratified according to posts' categories

Variable	Posts' category				Total
	Genuine information	Misinformation posts	Opinion	Unverified information	
Number of posts	316 (63.20)	10 (2.00)	152 (30.40)	22 (4.40)	500 (100)
Number of engagements	2,004 (200.5–11,230)	1,964.5 (43–5,052)	2,474.5 (924–11,160.5)	13,415 (8,507–22,869)	2,474.5 (407–11,777.5)
Influence score	4.43 ± 2.09	4.50 ± 1.58	4.60 ± 2.18	4.59 ± 2.11	4.49 ± 2.10
Source					
Social media	99 (31.33)	8 (80)	41 (26.79)	4 (18.18)	152 (30.40)
Online forum	80 (25.32)	2 (20)	51 (33.55)	11 (50.00)	144 (28.80)
Online newspaper	137 (43.35)	0 (0)	60 (39.47)	7 (31.82)	204 (40.80)
Periods					
Pre-outbreak	19 (6.01)	0 (0)	6 (3.95)	3 (13.64)	28 (5.60)
During outbreak	257 (81.33)	10 (100)	139 (91.45)	19 (86.36)	425 (85.00)
Post-outbreak	40 (12.66)	0 (0)	7 (4.61)	0 (0)	47 (9.40)
Sentiment polarity					
Positive	61 (19.30)	1 (10)	81 (53.29)	4 (18.18)	147 (29.40)
Neutral	196 (62.03)	0 (0)	11 (7.24)	0 (0)	207 (41.40)
Negative	59 (18.67)	9 (90)	60 (39.47)	18 (81.82)	146 (29.20)

Values are presented as number (%) or median (interquartile range) or mean \pm standard deviation.

8,507–22,869). The highest number of genuine information and opinion pieces were made in newspapers and during the outbreak. Most posts identified as neutral in terms of sentiment (196/207) were categorized as genuine information, while most posts classified as having positive sentiments (81/147) were opinion-expressing posts. Meanwhile, most misinformation posts were made on social media (8/10), and half of the posts with unverified information (11/22) were posted on online forums. While all identified misinformation posts were made during the outbreak, none of these posts were made with a neutral sentiment or on an online newspaper platform. Similarly, the identified unverified information posts did not display neutral sentiments and were not posted after the outbreak.

Negative binomial regression was conducted, as shown in Table 4, to demonstrate the association of identified posts' categories and number of engagements, adjusted for several characteristics. The number of engagements for unverified information was significantly higher than that for genuine

information, with an IRR of 2.83 (95% CI, 1.33–6.02), and this relationship remained significant after adjusting for the source, time periods, and sentiment of the posts. The adjusted model of posts' categories showed that online information published during the outbreak received significantly higher numbers of engagements than online information published before or after the outbreak—IRR = 0.15; 95% CI, 0.07–0.35 and IRR = 0.46; 95% CI, 0.34–0.63, respectively. After controlling for sentiment, the number of engagements of neutral posts was significantly higher than that of positive posts (IRR = 0.60; 95% CI, 0.37–0.97) and significantly lower than that of negative posts (IRR = 1.84; 95% CI, 1.23–2.75).

3. Analysis of Identified Misinformation and Unverified Information

Table 5 shows multivariable logistic regression for the distribution of posts' characteristics (source, sentiment polarity, and time period) between identified misinformation and the other post categories (model 1) and between unverified

Table 4. Negative binomial regression for the association between the number of engagements and posts' categories

Variable	Univariate analysis		Model 1		Model 2		Model 3	
	IRR	95% CI	IRR	95% CI	IRR	95% CI	IRR	95% CI
Posts' categories								
Genuine information	Ref		Ref		Ref		Ref	
Misinformation	0.91	0.28–2.98	0.84	0.26–2.71	0.80	0.24–2.62	0.58	0.18–1.83
Opinion	1.17	0.81–1.71	1.19	1.83–1.70	1.09	0.75–1.58	1.29	0.81–2.04
Unverified information	2.83***	1.33–6.02	3.15***	1.40–7.10	2.81**	1.42–5.56	1.98***	0.99–3.98
Source								
Social media	Ref		Ref					
Online forum	0.88	0.56–1.37	0.74	0.51–1.06				
Online newspaper	0.97	0.67–1.41	0.94	0.65–1.35				
Periods								
During outbreak	Ref				Ref			
Pre-outbreak	0.17*	0.08–0.35			0.15*	0.07–0.35		
Post-outbreak	0.42	0.31–0.57			0.46*	0.34–0.63		
Sentiment polarity								
Neutral	Ref						Ref	
Positive	0.70	0.47–1.05					0.60**	0.37–0.97
Negative	2.12*	1.45–3.09					1.84**	1.23–2.75

Model 1: Negative binomial regression for the association between number of engagements and posts' categories adjusted for source of posts. Model 2: Negative binomial regression for the association between number of engagements and posts' categories adjusted for time periods. Model 3: Negative binomial regression for the association between number of engagements and posts' categories adjusted for sentiment polarities.

IRR: incidence relative risk, CI: confidence interval.

* $p < 0.001$, ** $p < 0.01$, *** $p < 0.05$.

Table 5. Multivariable logistic regression for posts' characteristics between misinformation and verified information versus other post categories

Variable	Model 1: Misinformation					Model 2: Unverified information				
	n	Univariate		Multivariable		n	Univariate		Adjusted	
		OR	95% CI	OR	95% CI		OR	95% CI	OR	95% CI
Source										
Social media (n = 152)	8	Ref		Ref		4	Ref		Ref	
Online forum (n = 144)	2	0.25	0.05–1.21	0.23	0.05–1.13	11	3.06	0.95–9.84	3.04	0.91–10.24
Online newspaper (n = 204)	0					7	1.31	0.38–4.57	1.84	0.50–6.73
Periods										
During outbreak (n = 425)	10					19	Ref		Ref	
Pre-outbreak (n = 28)	0					3	2.56	0.71–9.25	2.77	0.68–11.32
Post-outbreak (n = 47)	0					0				
Sentiment polarity										
Positive (n = 147)	1	Ref		Ref		4	Ref		Ref	
Negative (n = 146)	9	9.59***	1.20–76.70	7.62	0.93–62.05	18	5.03**	1.66–15.24	5.05**	1.63–15.65
Neutral (n = 207)	0					0				

Model 1: Multivariable logistic regression for source and sentiment polarity between misinformation versus other post categories.

Model 2: Multivariable logistic regression for source, sentiment polarity, and time periods between unverified information versus other post categories.

OR: odds ratio, CI: confidence interval.

** $p < 0.01$, *** $p < 0.05$.

information and other post categories (model 2). Negative posts were significantly more likely to be misinformation (OR = 9.59; 95% CI, 1.20–6.70) or unverified information (OR = 5.03; 95% CI, 1.66–15.24) than positive posts. There were no significant differences in the probabilities of posts made in different time periods or made by different sources being misinformation or unverified information.

4. Sentiment Analysis of Selected Online Posts

There were 2,660 positive words and 4,748 negative words used in the 500 selected posts. More common use of negative words than positive words was observed across all post categories, sources, and time periods (Figure 3). Significantly higher numbers of negative words were used in online newspapers, after the outbreak, and in misinformation and unverified information (Supplementary Table S5). A significantly higher number of positive words was also found in posts reporting misinformation and unverified information than in the other categories.

IV. Discussion

We found a low volume of misinformation and unverified information among online information about public health

interventions during the Da Nang COVID-19 outbreak. Online posts containing unverified information and misinformation were more likely to have negative sentiments and contain a higher number of negative or positive words, and unverified information received higher engagement than other post categories.

Our study reported lower rates of misinformation than previous studies of COVID-19 or other recent epidemics [23,24]. This might be explained by the previous research of Gallotti et al. [25], who found a lower risk of infodemics in countries with stable political contexts and consistent public health interventions and messages throughout the epidemic. In 2020, Vietnam was regularly praised for its strict measures in response to COVID-19 and achieved one of the lowest COVID-19 infection and fatality rates globally. This success might have mitigated infodemic spread in Vietnam.

Many other studies on infodemics and misinformation in the COVID-19 context focused solely on Twitter [3,26]. Meanwhile, our study extended to all publicly available in-country online information, and beyond social media platforms, it included online newspapers and online forums, both of which are powerful information dissemination outlets in Vietnam. This thereby helped reflect online information flow outside the main international platforms that

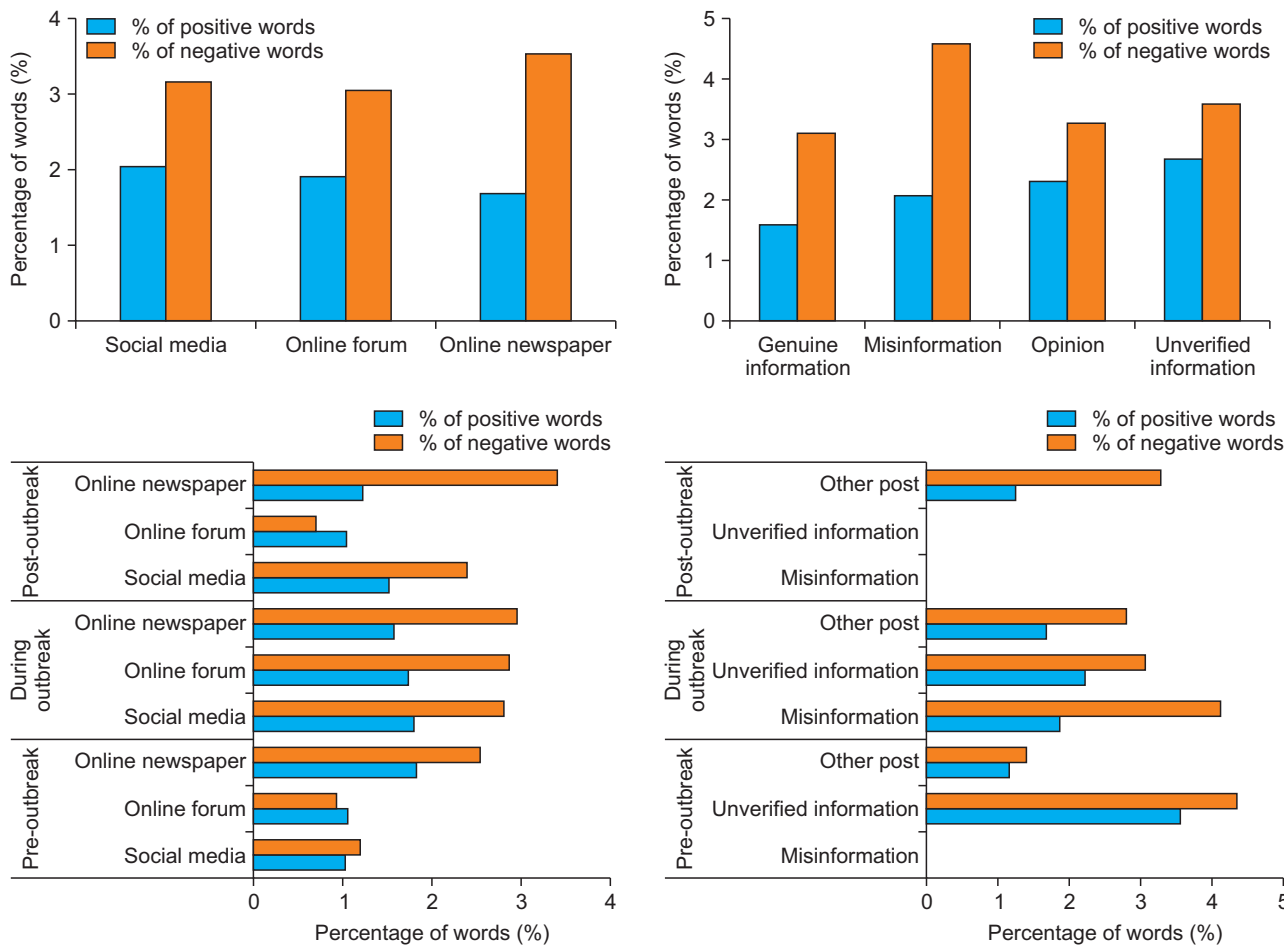


Figure 3. Distribution of positive and negative words used in 500 selected online posts stratified by posts' characteristics.

were used in previous research. Through this, we found that online forums and online newspapers were also a source of misinformation and unverified information. This highlights the gap in current research on COVID-19 misinformation, and also the need for infodemic control by public health agencies.

We found a strong relationship of misinformation and unverified information with sentiment polarity. This finding is in line with the study of Shahi et al. [27] on COVID-19 misinformation on Twitter, which showed that false claims featured more negative emotions than other news. In this study, more polarizing sentiments in misinformation were also closely related to the number of engagements of posts, which was also observed in previous studies on the number of retweets of COVID-19 information [28,29]. Our study showed that both unverified information and posts with negative sentiment received higher engagements than other posts. Similar conclusions were drawn from other studies on COVID-19, showing that negative information received more retweets [30]. Misinformation relies heavily on the implication of uncertainty and ambiguity of the situation, and it

creates fear, anxiety, or negative emotions in readers through its content. This may also explain the observation of higher engagements for online information during the outbreak, the time that the general public might have perceived the most uncertainty towards the outbreak's evolution, as well as the many measures implemented to contain it. Yet, this phenomenon of higher engagement for unverified information indicated that false or partially false information was more likely to spread and engage users. This implies not only the wide spread of misinformation, but also the danger of infodemics during times of uncertainty during the pandemic. Public health agencies, governments, and leaders should recognize this threat and strategize to effectively counter misinformation and unverified information, as the role of health education is crucial during pandemic times.

Our study was subject to several limitations. We did not explore in more detail the account characteristics or the semantics used, which would have given a more comprehensive depiction of not only the content of the posts, but also the entities that spread such information. We also limited our analysis to three sentiment statuses (positive,

negative, or neutral). A more in-depth study on post sentiments might have captured more accurately the emotions underlying the generation of online information, and more importantly, misinformation. Thirdly, we selected only the 500 most influential posts among all collected posts due to limited resources for data processing. Fourth, due to the nature of a retrospective study, the accurate rate of information in real time during the outbreak might not have been captured since some information might have been retracted or deleted. Moreover, our study was limited to only five distinct public health interventions over a short period of time, which might affect the generalizability of the findings. Nevertheless, we believe that, overall, our study offers a robust and valid analysis of an online infodemic related to a serious public health threat during the ongoing COVID-19 pandemic.

Our study provides important evidence about the volume and sentiment dynamics of misinformation and unverified information as part of the online infodemic during the ongoing COVID-19 pandemic. While the volume of incorrect or unverifiable online information was low overall, we showed that social media were not the only affected type of online platform. The choice of words, sentiment, and influence of the source had strong impacts on their distribution. This study offers important insights for public health decision-makers in Vietnam and other countries in the region with high rates of internet use to understand the public perceptions of health interventions in response to COVID-19.

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

We acknowledge great contributions from members of INFORE Company, Ministry of Science and Technology, and Rapid Response Team of National Steering Committee of COVID-19 Prevention and Control. This research was conducted as part of the Master of Applied Epidemiology program of the Australian National University in collaboration with National Institute of Hygiene and Epidemiology, Vietnam for HLQ. HLQ and NAH are trainees of the program and received scholarships from the ASEAN-Australia Health Security Fellowship by the Commonwealth Department of Foreign Affairs and Trade.

ORCID

Ha-Linh Quach (<https://orcid.org/0000-0001-7160-8329>)
 Thai Quang Pham (<https://orcid.org/0000-0002-3796-6162>)
 Ngoc-Anh Hoang (<https://orcid.org/0000-0002-5675-9396>)
 Viet-Cuong Nguyen (<https://orcid.org/0000-0002-8008-582X>)
 Son Hong Le (<https://orcid.org/0000-0001-5260-7486>)
 Thanh Cong Le (<https://orcid.org/0000-0003-2191-8868>)
 Dang Hai Le (<https://orcid.org/0000-0002-4276-9796>)
 Anh Duc Dang (<https://orcid.org/0000-0002-2380-2431>)
 Duong Nhu Tran (<https://orcid.org/0000-0002-9159-1103>)
 Nghia Duy Ngu (<https://orcid.org/0000-0002-6548-0530>)
 Florian Vogt (<https://orcid.org/0000-0003-2686-3149>)
 Cong-Khanh Nguyen (<https://orcid.org/0000-0002-3775-9139>)

Supplementary Materials

Supplementary materials can be found via <https://doi.org/10.4258/hir.2022.28.4.307>.

References

1. World Health Organization. Fighting misinformation in the time of COVID-19, one click at a time [Internet]. Geneva, Switzerland: World Health Organization; 2021 [cited 2022 Sep 30]. Available from: <https://www.who.int/news-room/feature-stories/detail/fighting-misinformation-in-the-time-of-covid-19-one-click-at-a-time>.
2. World Health Organization. Understanding the infodemic and misinformation in the fight against COVID-19 [Internet]. Geneva, Switzerland: World Health Organization; 2020 [cited 2022 Sep 30]. Available from: https://iris.paho.org/bitstream/handle/10665.2/52052/Factsheet-infodemic_eng.pdf?sequence=14.
3. Tang L, Bie B, Park SE, Zhi D. Social media and outbreaks of emerging infectious diseases: a systematic review of literature. *Am J Infect Control* 2018;46(9):962-72. <https://doi.org/10.1016/j.ajic.2018.02.010>
4. Towers S, Afzal S, Bernal G, Bliss N, Brown S, Espinoza B, et al. Mass media and the contagion of fear: the case of Ebola in America. *PLoS One* 2015;10(6):e0129179. <https://doi.org/10.1371/journal.pone.0129179>
5. Takahashi B, Tandoc EC, Carmichael C. Communicating on Twitter during a disaster: an analysis of tweets during Typhoon Haiyan in the Philippines. *Comput Human Behav* 2015;50:392-8. <https://doi.org/10.1371/journal.pone.0150190>
6. World Health Organization. Managing the COVID-19

- infodemic: promoting healthy behaviours and mitigating the harm from misinformation and disinformation [Internet]. Geneva, Switzerland: World Health Organization; 2020 [cited 2022 Sep 30]. Available from: <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>.
7. Quach HL, Nguyen KC, Hoang NA, Pham TQ, Tran DN, Le MT, et al. Association of public health interventions and COVID-19 incidence in Vietnam, January to December 2020. *Int J Infect Dis* 2021;110 Suppl 1:S28-S43. <https://doi.org/10.1016/j.ijid.2021.07.044>
 8. Le TH, Tran TP. Alert for COVID-19 second wave: a lesson from Vietnam. *J Glob Health* 2021;11:03012. <https://doi.org/10.7189/jogh.11.03012>
 9. Internet World Stats. Top 20 countries in Internet users [Internet]. [place unknown]: Internet World Stats; 2019 [cited 2022 Sep 30]. Available from: <https://www.internetworldstats.com/top20.htm>.
 10. Chou WS, Oh A, Klein WM. Addressing health-related misinformation on social media. *JAMA* 2018;320(23):2417-8. <https://doi.org/10.1001/jama.2018.16865>
 11. Cuan-Baltazar JY, Munoz-Perez MJ, Robledo-Vega C, Perez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the Internet: infodemiology study. *JMIR Public Health Surveill* 2020;6(2):e18444. <https://doi.org/10.2196/18444>
 12. Ferrara E, Cresci S, Luceri L. Misinformation, manipulation, and abuse on social media in the era of COVID-19. *J Comput Soc Sci* 2020;3(2):271-7. <https://doi.org/10.1007/s42001-020-00094-5>
 13. Alamoodi AH, Zaidan BB, Al-Masawa M, Taresh SM, Noman S, Ahmaro IY, et al. Multi-perspectives systematic review on the applications of sentiment analysis for vaccine hesitancy. *Comput Biol Med* 2021;139:104957. <https://doi.org/10.1016/j.combiomed.2021.104957>
 14. Alonso MA, Vilares D, Gomez-Rodríguez C, Vilares J. Sentiment analysis for fake news detection. *Electronics* 2021;10(11):1348. <https://doi.org/10.3390/electronics10111348>
 15. Zaeem RN, Li C, Barber KS. On sentiment of online fake news. *Proceedings of 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*; 2020 Dec 7-10; The Hague, Netherlands. p. 760-7. <https://doi.org/10.1109/ASONAM49781.2020.9381323>
 16. Ahinkorah BO, Ameyaw EK, Hagan JE, Seidu AA, Schack T. Rising above misinformation or fake news in Africa: another strategy to control COVID-19 spread. *Front Commun* 2020;5:45. <https://doi.org/10.3389/fcomm.2020.00045>
 17. Hou Z, Du F, Zhou X, Jiang H, Martin S, Larson H, et al. Cross-country comparison of public awareness, rumors, and behavioral responses to the COVID-19 epidemic: infodemiology Study. *J Med Internet Res* 2020;22(8):e21143. <https://doi.org/10.2196/21143>
 18. Enders AM, Uscinski JE, Klofstad C, Stoler J. The different forms of COVID-19 misinformation and their consequences. *Harv Kennedy Sch Misinformation Rev* 2020;1(8):1-21.
 19. World Health Organization. WHO public health research agenda for managing infodemics [Internet]. Geneva, Switzerland: World Health Organization; 2021 [cited 2022 Sep 30]. Available from: <https://www.who.int/publications/i/item/9789240019508>.
 20. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, et al. Coronavirus goes viral: quantifying the COVID-19 misinformation epidemic on twitter. *Cureus* 2020;12(3):e7255. <https://doi.org/10.7759/cureus.7255>
 21. Nguyen-Nhat DK, Duong HT. One-document training for Vietnamese sentiment analysis. In: Tagarelli A, Tong H, editors. *Computational data and social networks*. Cham, Switzerland: Springer; 2019, p. 189-200. https://doi.org/10.1007/978-3-030-34980-6_21
 22. Vu T, Nguyen DQ, Nguyen DQ, Dras M, Johnson M. VnCoreNLP: a Vietnamese natural language processing toolkit. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT)*; 2018 Jun 2-4; New Orleans, LA. p. 56-60.
 23. Singh L, Bansal S, Bode L, Budak C, Chi G, Kawintiranon K, et al. A first look at COVID-19 information and misinformation sharing on Twitter. *ArXiv [Preprint]* 2020;arXiv:2003.13907v1.
 24. Yang KC, Torres-Lugo C, Menczer F. Prevalence of low-credibility information on twitter during the covid-19 outbreak [Internet]. Ithaca (NY); arXiv.org; 2020 [cited 2022 Sep 30]. Available from: <https://arxiv.org/abs/2004.14484>.
 25. Gallotti R, Valle F, Castaldo N, Sacco P, De Domenico M. Assessing the risks of 'infodemics' in response to COVID-19 epidemics. *Nat Hum Behav* 2020;4(12):1285-93. <https://doi.org/10.1038/s41562-020-00994-6>
 26. Fung IC, Fu KW, Chan CH, Chan BS, Cheung CN, Abraham T, et al. Social media's initial reaction to in-

- formation and misinformation on Ebola, August 2014: facts and rumors. *Public Health Rep* 2016;131(3):461-73. <https://doi.org/10.1177/003335491613100312>
27. Shahi GK, Dirkson A, Majchrzak TA. An exploratory study of COVID-19 misinformation on Twitter. *Online Soc Netw Media* 2021;22:100104. <https://doi.org/10.1016/j.osnem.2020.100104>
28. Kim J, Yoo J. Role of sentiment in message propagation: reply vs. retweet behavior in political communication. *Proceedings of 2012 International Conference on Social Informatics*; 2012 Dec 14-16; Alexandria, VA. pp. 131-6. <https://doi.org/10.1109/SocialInformatics.2012.33>
29. Stieglitz S, Dang-Xuan L. Political communication and influence through microblogging: an empirical analysis of sentiment in Twitter messages and retweet behavior. *Proceedings of 2012 45th Hawaii International Conference on System Sciences*; 2012 Jan 4-7; Maui, HI. pp. 3500-9. <https://doi.org/10.1109/HICSS.2012.476>
30. Medford RJ, Saleh SN, Sumarsono A, Perl TM, Lehmann CU. An "Infodemic": leveraging high-volume twitter data to understand early public sentiment for the coronavirus disease 2019 outbreak. *Open Forum Infect Dis* 2020;7(7):ofaa258. <https://doi.org/10.1093/ofid/ofaa258>