JMIR Infodemiology

Journal Impact Factorâ ¢ (JIF) (2023): 3.5
Volume 5 (2025) ISSN 2564-1891 Editor-in-Chief: Tim Ken Mackey, MAS, PhD

Contents

Original Papers

Original Paper

How Patients With Cancer Use the Internet to Search for Health Information: Scenario-Based Think-Aloud Study

Fiorella Huijgens[1,2,3], MSc; Pascale Kwakman[1,2,3], MSc; Marij Hillen[1,2,3], PhD; Julia van Weert[4,5], Prof Dr; Monique Jaspers[5,6], Prof Dr; Ellen Smets[1,2,3], Prof Dr; Annemiek Linn[4,7], PhD

[1]Department of Medical Psychology, Amsterdam University Medical Centers, location Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands

[2]Quality of Care, Amsterdam Public Health, Amsterdam, Netherlands

[3]Cancer Treatment and Quality of Life, Cancer Center Amsterdam, Amsterdam, Netherlands

[4]Amsterdam School of Communication Research/ASCoR, Department of Communication Science, University of Amsterdam, Amsterdam, Netherlands

[5]Digital Health, Amsterdam Public Health, Amsterdam, Netherlands

[6]Department of Medical Informatics, Amsterdam University Medical Centers, location Academic Medical Center, University of Amsterdam, Amsterdam, Netherlands

[7]Health Behaviors and Chronic Diseases, Amsterdam Public Health, Amsterdam, Netherlands

Corresponding Author:
Fiorella Huijgens, MSc
Department of Medical Psychology
Amsterdam University Medical Centers, location Academic Medical Center
University of Amsterdam
Meibergdreef 9
Amsterdam, 1105 AZ
Netherlands
Phone: 31 623715595
Email: f.l.huijgens@amsterdamumc.nl

Abstract

Background: Patients with cancer increasingly use the internet to seek health information. However, thus far, research treats web-based health information seeking (WHIS) behavior in a rather dichotomous manner (ie, approaching or avoiding) and fails to capture the dynamic nature and evolving motivations that patients experience when engaging in WHIS throughout their disease trajectory. Insights can be used to support effective patient-provider communication about WHIS and can lead to better designed web-based health platforms.

Objective: This study explored patterns of motivations and emotions behind the web-based information seeking of patients with cancer at various stages of their disease trajectory, as well as the cognitive and emotional responses evoked by WHIS via a scenario-based, think-aloud approach.

Methods: In total, 15 analog patients were recruited, representing patients with cancer, survivors, and informal caregivers. Imagining themselves in 3 scenarios—prediagnosis phase (5/15, 33%), treatment phase (5/15, 33%), and survivor phase (5/15, 33%)—patients were asked to search for web-based health information while being prompted to verbalize their thoughts. In total, 2 researchers independently coded the sessions, categorizing the codes into broader themes to comprehend analog patients' experiences during WHIS.

Results: Overarching motives for WHIS included reducing uncertainty, seeking reassurance, and gaining empowerment. At the beginning of the disease trajectory, patients mainly showed cognitive needs, whereas this shifted more toward affective needs in the subsequent disease stages. Analog patients' WHIS approaches varied from exploratory to focused or a combination of both. They adapted their search strategy when faced with challenging cognitive or emotional content. WHIS triggered diverse emotions, fluctuating throughout the search. Complex, confrontational, and unexpected information mainly induced negative emotions.

Conclusions: This study provides valuable insights into the motivations of patients with cancer underlying WHIS and the emotions experienced at various stages of the disease trajectory. Understanding patients' search patterns is pivotal in optimizing web-based health platforms to cater to specific needs. In addition, these findings can guide clinicians in accommodating patients' specific needs and directing patients toward reliable sources of web-based health information.

(JMIR Infodemiology 2025;5:e59625) doi:10.2196/59625

KEYWORDS

web-based health information seeking; think aloud; scenario based; cancer; patient evaluation; information seeking; web-based information; health information; internet; pattern; motivation; cognitive; emotional; response; patient; survivor; caregiver; interview; scenario; women; men

Introduction

Background

Patients with cancer increasingly use web-based platforms to seek information about their diagnosis, treatment, and implications thereof in the short and long term. In the Netherlands, 85% of patients with cancer use the internet [1,2], a rate comparable to that in most Asian countries [3] and other European countries [4,5]. The internet offers a wealth of information that can be readily accessed. It provides practically limitless opportunities for finding health information and support from both lay and expert perspectives, making it a highly popular source of information for many patients.

Within the context of cancer, patients' web-based health information seeking (WHIS) behaviors have been explained through theories of coping behavior. Most often, cancer literature on information-seeking patterns revolves around coping behaviors such as monitoring and blunting. Studies suggest that most patients manage health threats by proactively seeking information, a behavior referred to as monitoring coping style, whereas others choose to avoid information and opt for distraction, known as blunting coping style [6,7]. However, some studies indicate that the WHIS behaviors of patients with cancer could be explained via a broader range of approaches than merely through theories of coping behavior [8-10]. For instance, patients with cancer could also differ in their choices regarding the kind, quantity, and origins of the sought information, as well as the strategies used for information management. These approaches are based on patients' perceptions of self-care, which means that patients vary in their WHIS based on what they need to adequately take care of themselves [10]. In addition, the reasons behind seeking information and emotional support on the web are contingent on how patients use the internet [9].

Another factor that could explain variations in how people use the internet is patients' disease and treatment stage—which may predict different needs concerning the type and amount of information [11,12]. However, studies investigating WHIS and particularly the motives to engage in WHIS often treat the behavior as a one-time event. By treating WHIS as a one-time event, researchers tend to overlook the dynamic nature of health information needs and fail to capture the evolving motivations that patients experience throughout their disease trajectory. Considering that searching for health information is a rather longitudinal behavior, especially for patients moving through different stages of the disease trajectory, a longitudinal lens is required when studying WHIS [11].

In addition to the different phases in the disease trajectory influencing how patients use the internet, WHIS may also vary depending on patients' *motives* for going on the web. For example, patients may do so to address their cognitive (ie, the need for understanding) and affective (ie, the need to be understood) needs [13]. Cognitive needs (eg, engaging with the internet to enhance preparedness and comprehension of the information provided during a consultation or to validate or challenge the information offered by the provider) will lead to diverse forms of WHIS compared to affective needs (eg, using the internet for peer interaction). In other words, patients' specific goals regarding information seeking could also impact their search queries [13]. However, these motives are often not sufficiently taken into account when studying WHIS behavior.

Finally, in the period between diagnosis and cure or remission, patients often experience a range of emotions, including (but not limited to) uncertainty, hope, fear, and anxiety. These feelings and emotions are important motivators for many patients to seek out information to cope with their illness [14]. For example, when just diagnosed with cancer, individuals might be concerned about the unpredictable aspects of the disease, leading them to search for information to better manage and cope with their newly discovered illness. Apart from instigating patients' WHIS behavior, these emotions may also influence decisions to continue, expand, or terminate WHIS [10,14-16]. Earlier qualitative studies have identified various WHIS patterns and the emotions associated with them, ranging from intense to guarded information seeking [10,16,17]. While all participants in these studies expressed a desire for basic information about their diagnosis, they also exhibited diversity in their motivations for seeking cancer information; the emotions experienced; and the nature, quantity, and sources of the sought information, along with the strategies used to manage this information. However, interviews rely on patients' subjective, retrospective reporting and, therefore, do not provide a comprehensive overview of WHIS behavior.

Hence, it is thus far largely unknown how various motives and emotions guide WHIS behavior in various phases of the cancer disease trajectory, whereas such insights can lead to better designed web-based health platforms catering to patients' changing requirements and supporting them effectively throughout their health journey. In addition, having a comprehensive understanding of how patients navigate information acquisition on the internet is crucial for establishing effective patient-provider communication that accommodates patients' specific needs. These insights may also make health care providers aware of the potential impact that WHIS has on patients and, consequently, on the consultation.

Objectives

Studying the impact of motives and emotions on information-seeking behavior during the disease trajectory poses several challenges that have not been taken into account in previous studies. First, as most WHIS occurs in private settings, such as at home, most of these studies use data collection methods that rely on patients' subjective, retrospective reporting, such as surveys, focus groups, and interviews. Using these retrospective methods presents significant drawbacks, including recall bias, which may lead to inaccurate results [18]. In particular, information collected before or during diagnosis is considered challenging as this often entails a short and stressful period for many patients [19]. New research methods such as the think-aloud method enable participants to verbalize what they are thinking and doing while performing a certain task [20]; this allows researchers to observe patients' WHIS more precisely. This includes assessing attention to web-based information, choices made while selecting information, and people's thoughts and feelings evoked during exposure to information [21]. When combining the think-aloud method with vignettes representing different scenarios at various stages of the disease trajectory, research has the potential to provide a more comprehensive and naturalistic view on the WHIS of patients with cancer. Therefore, this study aimed to explore patterns of motivations and emotions behind the web-based information seeking of patients with cancer at different stages of their disease trajectory, as well as the cognitive and emotional responses evoked via a scenario-based, think-aloud approach. This study adopted a unique explorative approach by observing analog patients (ie, patients or healthy participants putting themselves in the position of a patient [22]) as they engaged in WHIS during different phases of their disease trajectory.

Methods

Study Design, Setting, and Population

We used a scenario-based, think-aloud approach followed by a semistructured interview to obtain more in-depth information regarding analog patients' search strategy, their reasoning and emotions behind this strategy (ie, motives), and the emotions experienced throughout. To increase feasibility and for ethical reasons, we decided to rely on analog patients (patients or healthy participants who are asked to imagine themselves in the role of the patients), who are considered valid proxies for clinical patients [23,24]. The COREQ (Consolidated Criteria for Reporting Qualitative Research) guidelines were used to report the methods (Multimedia Appendix 1).

Analog patients were recruited from a local panel of patients with cancer, survivors, and their informal caregivers who were willing to participate in scientific research on patient-provider communication and health information provision [25]. In this way, we ensured that the analog patients had some personal experience with cancer. Via email, panel members were informed about the study purpose and invited to complete a screening questionnaire to establish their eligibility, that is, whether they were aged ≥18 years, had previously used the internet to search for health information, and owned a computer or laptop with internet connection. The screening questionnaire also included panel members' age, gender, and educational attainment to allow for purposive sampling based on these characteristics as research shows that individuals differing in these characteristics navigate the web differently and differ in information needs [26]. In addition, we strived for diversity in relation to cancer experience (eg, "I have (had) cancer" or "My partner has (had) cancer"), cancer type, and frequency of using the internet for health information in the previous year (eg, "1-5 times," "6-10 times," "11-30 times," or "more than 30 times").

In total, 75 panel members indicated an interest in participating. Of these 75 members, we invited 34 (45%) individuals based on purposive sampling to take part in the scenario-based, think-aloud study. Eventually, of the 34 individuals, 5 (15%) participated in the pilot study, and 15 (44%) participated in the think-aloud sessions, 5 (33%) for each scenario. Among the 34 individuals, there were 9 (26%) nonresponses, 1 (3%) failed recording, and 4 (12%) who opted out.

Procedure

The scenario-based, think-aloud sessions were conducted between May 2021 and December 2021 by 3 researchers (PK, FH, and an undergraduate student). PK and the student have a health communication background, and FH has a health science and health care management background. PK is trained in qualitative research. Due to the COVID-19 pandemic, the sessions were held on the web using videoconferencing software (ie, Zoom [Zoom Video Communications] or Microsoft Teams [Microsoft Corp]) and were recorded with video. Analog patients could participate in the sessions from the comfort of their home while using their own devices, thereby enhancing ecological validity.

We used a protocol for the scenario-based, think-aloud sessions, including a semistructured interview guide. This protocol was pilot-tested with 15% (5/34) of the analog patients. On the basis of the pilot, we decided to develop a video tutorial explaining the think-aloud procedure and a written manual explaining the use of the videoconferencing software (eg, "How do I share my screen?"). We also adapted the interview guide by adding questions focusing on analog patients' explanations of and reflections on their WHIS behavior (Multimedia Appendix 2). Participating analog patients received an email including an information letter and the video tutorial.

At the start of each session, the researcher explained the nature of the scenario-based, think-aloud method to the analog patients and asked for their personal experience with WHIS. Then, to become familiar with the process of thinking aloud, the analog patients were presented with a practical task (ie, to find a recipe for a pie or a cake containing apples) [27].

After familiarizing the analog patients with the think-aloud procedure, the researcher asked them to imagine themselves in one of the three following scenarios: (1) being an individual who experienced symptoms that could point toward non-Hodgkin lymphoma (NHL), hereinafter referred to as analog prediagnostic patient; (2) being a patient who is about to receive treatment for NHL, hereinafter referred to as analog patient with cancer; or (3) being a survivor of NHL 2 months after having finished treatment, hereinafter referred to as analog survivor of cancer (Multimedia Appendix 3). We use the general term *analog patients* when referring to 2 or 3 scenarios.

Each scenario was based on real patient experiences that were reported in blogs and discussion groups of the largest cancer-related website in the Netherlands [28] and was reviewed by a survivor of cancer to optimize external validity [29]. Analog patients were assigned to the scenario that was most appropriate given their health status and relationship to cancer.

To enhance identification, analog patients were asked to report in their own words what they had just heard in the scenario. In addition, the researcher asked analog patients to discuss any thoughts or feelings that were evoked by the scenario and score their stress, anxiety, worries about cancer, hope, and uncertainty on an 11-point thermometer-style scale (0=*not at all*; 10=*an extreme amount*). Next, analog patients were asked to go on the web imagining themselves as the described patient in the scenario. While performing the various tasks, analog patients were asked to share their screen. The researcher instructed analog patients to indicate when they wanted to stop their web-based search. If analog patients fell silent during the session, the researcher reminded them to voice their thoughts.

After the think-aloud process, a short semistructured interview was conducted in which the researchers probed for analog patients' motives (eg, what made them choose particular search terms or why they decided to end their search) and their satisfaction with the content (Multimedia Appendix 2). Each interview session ended with a questionnaire assessing the analog patients' coping style (Dutch Threatening Medical Situations Inventory [30,31]), uncertainty intolerance (Dutch version of the short Intolerance of Uncertainty Scale [32]), information needs [33], and eHealth literacy (Dutch eHealth Literacy Scale [34]). These measures were used to be able to describe the sample.

Data Analysis

In total, 2 coders (FH and PK) first familiarized themselves with the data by watching the recordings and reading the interviewer field notes. Second, they independently selected and transcribed parts of each recording that seemed relevant to the research questions (eg, motives and emotions related to WHIS and search strategies). During the analysis, they focused on the analog patients' actions (observations), their verbalized thoughts during the scenario-based, think-aloud process (what they did vs what they said), and their reflections (interview). What was considered relevant was first discussed with a third team member (AL). Third, the coders independently double coded all relevant fragments. Fragments were coded inductively based on the sensitizing concepts as discussed in the introduction (ie, emotions and motives to seek web-based health information, search strategy used, and type of emotions evoked). During the observations, the coders closely examined the search terms used by the analog patients and the content viewed to deduce the analog patients' underlying motives. Fourth, the coders met and discussed their codes after each session to reach an agreement on the coding scheme together with a third team member (AL). Fifth, after completion of the coding process, the codes were aggregated into potential overarching themes and subthemes through comparisons and discussion between the coders. To improve reliability, validity, and generalizability, the results were substantiated using vivid quotes, and a continuous process of reflection and discussion among the coauthors (FH, PK, AL, and ES) was used. To improve the readability of the overall analysis (N=15), we decided to use the term *most* when the analysis applied to >10 analog patients, *several* when it applied to between 5 and 10 analog patients, and *some* when the analysis applied to <5 analog patients. For scenario-specific analysis (5/15, 33%), we decided to use the term *most* when the analysis applied to 3 or 4 analog patients and the term *some* when the analysis applied to 2 analog patients.

Ethical Considerations

The Amsterdam School of Communication Research Ethical Review Board approved this study at the University of Amsterdam (ethics approval code: 2021-PC-13493). Informed consent was verbally obtained from analog patients at the start of the scenario-based, think-aloud session. Analog patients could withdraw their consent at any time. The data could not be anonymized as the think-aloud interviews were video recorded. The data are saved on a secured drive of the Amsterdam University Medical Center. No compensation was provided to the participants.

Results

Sample Characteristics

Among the 15 participating analog patients (n=9, 60% women and n=6, 40% men), the ages ranged from 28 to 72 years (mean 56.9, SD 12.5 years). Most were former patients with cancer and reported having used the internet for seeking health information >6 times in the foregoing year. In total, the sessions lasted between 25 and 70 minutes, and the web-based search lasted between approximately 6 and 26 minutes. The number of web pages visited ranged from 3 to 15 per session, and changes in search terms ranged from 1 to 16 per session. Table 1 shows the sample characteristics, and Tables 2-4 provide descriptions of the individual search sessions.

Table 1. Analog patient characteristics (N=15).

	Prediagnosis stage (n=5)	Treatment stage (n=5)	Survivor stage (n=5)	Total
Age (y), mean (SD; range)	59.6 (8.1; 51-72)	54.6 (14.9; 28-63)	56.4 (15.7; 29-66)	56.9 (12.5; 28-72)
Gender, n (%)				
Woman	3 (60)	3 (60)	3 (60)	9 (60)
Man	2 (40)	2 (40)	2 (40)	6 (40)
Educational level, n (%)[a]				
Low	1 (20)	1 (20)	1 (20)	3 (20)
Middle	0 (0)	1 (20)	2 (40)	3 (20)
High	4 (80)	3 (60)	2 (40)	9 (60)
Relationship to cancer, n (%)				
Having cancer	0 (0)	2 (40)	1 (20)	3 (20)
Having had cancer	2 (40)	3 (60)	4 (80)	9 (60)
Having a relative with cancer	3 (60)	0 (0)	0 (0)	3 (20)
Frequency of web-based health information seeking in the previous year, n (%)				
1-5 times	3 (60)	2 (40)	1 (20)	6 (40)
6-10 times	1 (20)	0 (0)	2 (40)	3 (20)
11-30 times	0 (0)	3 (60)	1 (20)	4 (27)
>30 times	1 (20)	0 (0)	1 (20)	2 (13)
Uncertainty intolerance score, mean (SD; range)	36.2 (7.9; 25-46)	31.8 (9.4; 24-47)	25.6 (7.4; 15-36)	31.2 (8.9; 15-47)
eHEALS[b] score, mean (SD; range)	34.6 (3.8; 31-40)	34.0 (5.3; 27-40)	36.6 (2.1; 34-39)	35.1 (3.8; 27-40)
Monitoring coping style score, mean (SD; range)	11.8 (2.6; 8-15)	13.0 (2.3; 10-15)	8.2 (1.3; 6-9)	11.0 (2.9; 6-15)
Information preference, n (%)				
"I want to know as much as possible, both positive and negative information."	4 (80)	4 (80)	3 (60)	11 (73)
"I want to know as much as possible, both positive and negative information, but in a dosed way (little by little)."	1 (20)	1 (20)	1 (20)	3 (20)
"I want mainly positive information."	0 (0)	0 (0)	1 (20)	1 (7)
"I don't need to know that much."	0 (0)	0 (0)	0 (0)	0 (0)

[a]Low: secondary education; middle: senior general secondary education, secondary vocational education, and preuniversity education; high: college or university.

[b]eHEALS: eHealth Literacy Scale.

Table 2. Characteristics of the participants and search sessions in the prediagnosis phase.

	Participant S01	Participant S05	Participant S06	Participant S08	Participant S10	Values, mean (SD)
Age (y)	61	51	54	60	72	59.6 (8.1)
Gender	Man	Woman	Man	Woman	Woman	—[a]
Educational level[b]	High	High	High	Low	High	—
Search time	15 min 52 s	8 min 57 s	16 min 51 s	6 min 11 s	7 min 55 s	11 min 9 s (4 min 51 s)
Times changing search terms, N	9	4	8	4	1	5.2 (3.3)
Search engine used	Google	Google	Google	Google	Google	—
Total web pages visited, N	5	9	9	3	5	6.2 (2.7)
Uncertainty intolerance score[c]	35	34	41	46	25	36.2 (7.9)
eHealth literacy score[d]	33	32	37	31	40	34.6 (3.8)
Monitoring coping style score[e]	11	8	15	12	13	11.8 (2.6)
Thermometer score[f]						
Feelings of stress and anxiety	7	6.5	8	7	5	6.7 (1.1)
Worries about cancer	7	7.5	6	5.5	5	6.2 (1.0)
Hope	—	—	—	—	—	—
Uncertainty	7	7.5	8	6	10	7.7 (1.5)

[a]Not applicable.

[b]Low: secondary education; middle: senior general secondary education, secondary vocational education, and preuniversity education; high: college or university.

[c]Uncertainty intolerance was measured using the Intolerance of Uncertainty Scale. Total sum scores can range from 12 to 60. A higher score indicates a higher level of intolerance of uncertainty.

[d]Self-perceived eHealth literacy was measured using the eHealth Literacy Scale (eHEALS). Scores on the eHEALS are summed and can range from 8 to 40, with higher scores representing higher self-perceived eHealth literacy.

[e]Monitoring coping style was measured using the Threatening Medical Situations Inventory. Scores are based on the sum of items and can range from 3 to 16. A higher score means a higher monitoring coping style.

[f]Analog patients were asked before the search session to score feelings of stress and anxiety (thermometer 1), worries about cancer (only for the first scenario), hope (only for the second and third scenarios; thermometer 2), and uncertainty (thermometer 3) on an 11-point thermometer-style scale (*0=not at all*; *10=an extreme amount*).

Table 3. Characteristics of the participants and search sessions in the treatment phase.

	Participant S23	Participant S24	Participant S25	Participant S27	Participant S28	Values, mean (SD)
Age (years)	61	62	63	59	28	54.6 (14.9)
Gender	Woman	Man	Woman	Woman	Man	—[a]
Educational level[b]	Low	High	High	Middle	High	—
Search time	9 min 55 s	13 min 40 s	16 min 34 s	24 min 55 s	16 min 35 s	16 min 19 s (5 min 31 s)
Times changing search terms, N	5	5	9	11	9	7.8 (2.7)
Search engines used	Google	Google and Microsoft Bing	Google, Firefox, and Norton Safe Search	Google	Google	—
Total web pages visited, N	5	3	10	11	9	7.6 (3.4)
Uncertainty intolerance score[c]	32	32	24	47	24	31.8 (9.4)
eHealth literacy score[d]	36	40	30	36	27	34 (5.3)
Monitoring coping style score[e]	15	15	10	14	11	13 (2.3)
Thermometer score[f]						
Feelings of stress and anxiety	7	8	7	9	8	7.8 (0.8)
Worries about cancer	—	—	—	—	—	—
Hope	9	3	9.5	4	4.5	6 (3.0)
Uncertainty	8	8.5	2	9	5.5	6.6 (2.9)

[a]Not applicable.

[b]Low: secondary education; middle: senior general secondary education, secondary vocational education, and preuniversity education; high: college or university.

[c]Uncertainty intolerance was measured using the Intolerance of Uncertainty Scale. Total sum scores can range from 12 to 60. A higher score indicates a higher level of intolerance of uncertainty.

[d]Self-perceived eHealth literacy was measured using the eHealth Literacy Scale (eHEALS). Scores on the eHEALS are summed and can range from 8 to 40, with higher scores representing higher self-perceived eHealth literacy.

[e]Monitoring coping style was measured using the Threatening Medical Situations Inventory. Scores are based on the sum of items and can range from 3 to 16. A higher score means a higher monitoring coping style.

[f]Analog patients were asked before the search session to score feelings of stress and anxiety (thermometer 1), worries about cancer (only for the first scenario), hope (only for the second and third scenarios; thermometer 2), and uncertainty (thermometer 3) on an 11-point thermometer-style scale (0=*not at all*; 10=*an extreme amount*).

Table 4. Characteristics of the participants and search sessions in the survivor phase.

	Participant S32	Participant S34	Participant S35	Participant S36	Participant S37	Values, mean (SD)
Age (y)	66	63	66	29	58	56.4 (15.7)
Gender	Woman	Man	Man	Woman	Woman	—[a]
Educational level[b]	Low	High	High	Middle	Middle	—
Search time	15 min 11 s	21 min 40 s	8 min 40 s	25 min 55 s	23 min 35 s	19 min 00 s (7 min 01 s)
Times changing search terms, N	6	16	4	13	8	9.4 (5.0)
Search engines used	Microsoft Bing	Google and Microsoft Bing	Google	Microsoft Bing	Google	—
Total web pages visited, N	8	13	4	15	12	10.4 (4.4)
Uncertainty intolerance score[c]	25	26	15	26	36	25.6 (7.4)
eHealth literacy score[d]	38	39	34	37	35	36.6 (2.1)
Monitoring coping style score[e]	9	6	9	9	8	8.2 (1.3)
Thermometer score[f]						
Feelings of stress and anxiety	5	8	6	3.5	8	6.1 (1.9)
Worries about cancer	—	—	—	—	—	—
Hope	3	6.5	8	10	8	7.1 (2.6)
Uncertainty	7.5	6	0	5	9	5.5 (3.4)

[a]Not applicable.

[b]Low: secondary education; middle: senior general secondary education, secondary vocational education, and preuniversity education; high: college or university.

[c]Uncertainty intolerance was measured using the Intolerance of Uncertainty Scale. Total sum scores can range from 12 to 60. A higher score indicates a higher level of intolerance of uncertainty.

[d]Self-perceived eHealth literacy was measured using the eHealth Literacy Scale (eHEALS). Scores on the eHEALS are summed and can range from 8 to 40, with higher scores representing higher self-perceived eHealth literacy.

[e]Monitoring coping style was measured using the Threatening Medical Situations Inventory. Scores are based on the sum of items and can range from 3 to 16. A higher score means a higher monitoring coping style.

[f]Analog patients were asked before the search session to score feelings of stress and anxiety (thermometer 1), worries about cancer (only for the first scenario), hope (only for the second and third scenarios; thermometer 2), and uncertainty (thermometer 3) on an 11-point thermometer-style scale (0=*not at all*; 10=*an extreme amount*).

Start of the Search Session

Analog patients reported starting their search session with various associations and reactions evoked by the scenario. For example, in the scenario in which they were experiencing symptoms, analog prediagnostic patients were immediately worried about cancer or felt alarmed by specific symptoms. This was reflected in their search terms, showing a predominant focus on searching for information about these symptoms. This was also reflected in their thoughts as patients expressed concern about the symptoms. Whenever the general practitioner in the scenario showed concern, analog patients more often showed signs of feeling distressed:

> *The word tumor immediately pops into my mind. This is serious. These are symptoms I would not trust.* [S01; analog prediagnostic patient]

> *You do not immediately think the best, especially sweating attacks and weight loss are warning signs.* [S05; analog prediagnostic patient]

Most analog patients with cancer assigned to the scenario of undergoing cancer treatment started their search by expressing fear about the upcoming challenges, particularly the apprehension of chemotherapy, and harboring doubts about the effectiveness of the treatment. The aggressive nature of NHL added to their anxiety, with a lack of optimistic information causing visible distress and confusion about the treatment process:

> *I am scared of what's coming and scared of the chemo. And I am not so hopeful because of my doubt whether the treatment will work.* [S27; analog patient with cancer]

Despite these negative emotions, some analog patients with cancer still remained combative or hopeful:

> *Damn, I have cancer again, now I have to have another treatment, but well I am going for it, because I am far from finished living.* [S25; analog patient with cancer]

This fear was also reflected in their search, with all analog patients with cancer being prone to mainly focus on using search words that were used in the scenario (*(aggressive) non-Hodgkin* and *R-CHOP* [rituximab, cyclophosphamide, hydroxydaunorubicin, oncovin, and prednisone regimen]).

Finally, those who were allocated to the survivor case ("analog survivors of cancer") generally voiced uncertainty at the beginning of the search about whether the cancer was definitely gone. They showed concerns about cancer recurrence and recovery and were somewhat skeptical about recovery:

> *Should I really be happy with being cancer-free? What if it comes back? Before this, I had not felt anything. Now, I do not know what I should and should not feel anymore.* [S32; analog survivor of cancer]

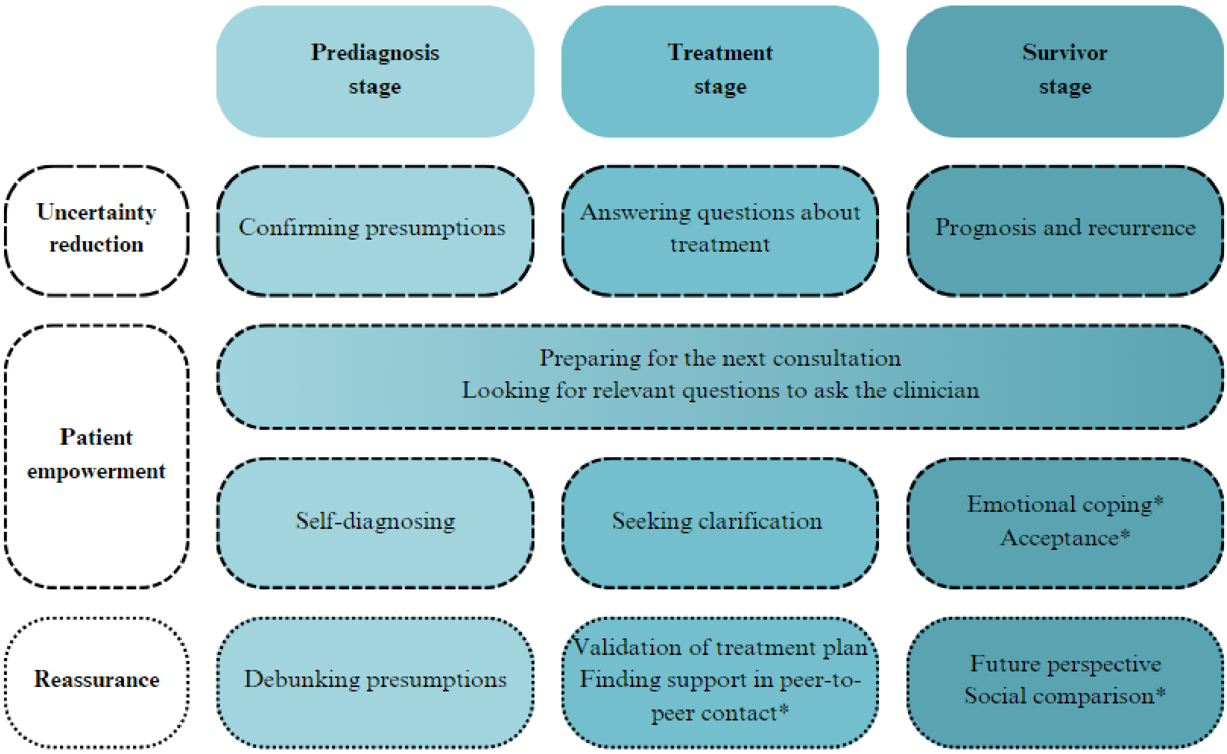
Analog survivors of cancer voiced that, most of all, they wanted to return to their normal lives before the diagnosis and, accordingly, started with search terms related to this desire to get back to the normality of their lives (eg, *out of cancer treatment, what now?*).

Search Motives

Overview

On the basis of the search strategies observed and the thoughts voiced, we were able to distinguish 3 overarching motives guiding patients' WHIS. These overarching motives were prevalent regardless of the allocated stage in the disease trajectory. Each overarching motive was expressed differently throughout the various disease stages (Figure 1). The first motive was *uncertainty reduction* to cope with the anxiety and health threats as most analog patients started their search by expressing uncertainty about what was going to happen to them. The second motive was *empowerment* (ie, "the process of increasing the capacity of individuals (or groups) to make choices and to transform those choices into desired actions and outcomes" [35]) as most analog patients searched content to pursue an active role in their own care process, for example, by actively preparing for the next consultation and looking for relevant questions to ask the clinician. The third motive was *finding reassurance* as analog patients wished to find content that would give them some hope. The 3 overarching motives were not mutually exclusive; they could go hand in hand.

Figure 1. Expressions of the 3 overarching motives for web-based health information seeking—uncertainty reduction, patient empowerment, and reassurance—within the 3 disease stages (prediagnosis stage, treatment stage, and survivor stage). *Affective needs (ie, need to be understood).



Motives in the Prediagnosis Stage

Analog prediagnostic patients wanted to diminish their anxiety and reduce their uncertainty by starting their search with confirming their presumptions. One analog prediagnostic patient immediately used the search term *characteristics cancer*, looked on different websites to compare symptoms, and said the following at search onset:

> *You do not know anything for sure...Apart from the fact that I initially think it is cancer, I still want to confirm that by searching the internet.* [S10; analog prediagnostic patient]

Furthermore, they mostly used the internet to empower themselves by attempting to self-diagnose and prepare for the next consultation. When trying to self-diagnose, they used symptom-related search terms, such as *fatigue*, *swollen glands*, *(unexplained) weight loss*, and *night sweats*. After encountering content about possible diagnoses, some changed their search terms to *symptoms of non-Hodgkin* and *symptoms of cancer* while simultaneously explaining this change:

> *I am actually finding several causes now and cancer is also mentioned. However, I am not quite happy with the information I'm getting yet. But since cancer has come up a few times, I am going to search for symptoms of cancer, so I'm turning it [the search terms] around now [searches for: symptoms of cancer].* [S06; analog prediagnostic patient]

The motive *empowerment* was apparent in one analog prediagnostic patient who used the search terms *preparing consult internist* and read the text *What can you do to prepare for the first visit with an internist?*:

> *What I would do now, because I am going to the hospital, is that I am going to prepare. So, I am now going to search on prepare consult internist. I think I am going to an internist, but obviously I'm not sure yet. [reads text on how to prepare for a visit] I would also like to know, what are useful questions? [clicks on other website] Okay, I have pretty much got everything now I need to consider, only I have to go through the 3 good questions again which I can ask the internist [opens the online brochure about 3 good questions].* [S06; analog prediagnostic patient]

The extent to which analog prediagnostic patients in this phase narrowed down their search to know their exact (possible) diagnosis differed. Some searched various options related to the symptoms, one settled for the likely diagnosis "cancer," whereas others continued their search until they had a specific idea about the type of cancer. Those who searched for various possible diagnoses wanted to be reassured that the symptoms could be anything other than a serious illness such as cancer. They tried to debunk their presumptions, as reflected in the following observation and quote:

> *[reads content about causes of swollen lymph nodes] Infection, which could also be, that makes sense. Then I see here swollen nodes due to a systemic disease. Then I am thinking about Lyme disease, okay. That is different from a tumor. Autoimmune disease is potentially on the table. I already see that swollen nodes can be caused by many factors, which is somewhat reassuring.* [S01; analog prediagnostic patient]

Motives in the Treatment Stage

Analog patients with cancer mostly appeared to use the internet to answer their remaining questions to reduce uncertainty. Reducing uncertainty seemed to be combined with increasing their feeling of empowerment as they appeared to seek for more clarification about diagnosis and treatment. Both uncertainty reduction and empowerment were reflected in search terms such as *What is non-Hodgkin lymphoma?*, *R-CHOP*, *side effects*, and *immunotherapy* (ie, cognitive needs). While searching these terms, they said the following:

> *More than 50% of patients with an aggressive non-Hodgkin lymphoma in an advanced stage are cured after treatment with R-CHOP. Okay, that is quite a lot. But, hmm, yes, the other half does not. It is not clear to me whether the half that does not recover remains chronically ill or simply succumbs to death. I would like to know that in that sense.* [S28; analog patient with cancer]

The motive to obtain reassurance via web-based information was reflected in analog patients with cancer using the internet to validate whether the treatment (as being proposed in the scenario) was indeed the right treatment for them. They specifically searched for websites and information that would convince them of this treatment being the best option:

> *And I would definitely, before starting that treatment, request a second opinion from another institution to ensure that I...um...yes, receive the correct diagnosis or the right treatment [searches for other hospitals].* [S24; analog patient with cancer]

One analog patient with cancer also seemed to use the internet to obtain reassurance via socioemotional content. This was reflected in the search term *experiences with R CHOP*. Of note, none of the analog patients with cancer used search terms indicating a need to know more about the prognosis of NHL.

Motives in the Survivor Stage

Analog survivors of cancer seemed to use the internet to reduce uncertainty only to a limited extent. When they used the internet for that purpose, they wanted to know more about prognosis and recurrence, as reflected in search terms such as *prognosis*, *late effects*, and *what to expect*. While using these search terms, they said the following:

> *Yes, you are quite uncertain about how everything will unfold. There are still quite a few questions, and that diminishes over time, but especially in the beginning after that hospital period, you still have quite a lot of questions.* [S37; analog survivor of cancer]

Analog survivors of cancer mainly used the internet to search for socioemotional content related to pursuing an active role in their own recovery (ie, patient empowerment). This was reflected in search terms regarding feelings, experiences, and emotions (eg, *uncertainty after cancer* and *feelings after non-Hodgkin treatment*). Pursuing an active role in their own recovery mainly encompassed (emotional) coping and finding acceptance (eg, returning to their normal life before diagnosis). Apparently, to satisfy these motives, they often visited blogs of survivors of cancer writing about feelings and experiences and providing advice on coping with survivorship (eg, *how to deal with emotions/fatigue/work/daily life*). Some searched for psychologists or for recovery programs offered by patient organizations or hospitals, which could also be seen as an expression of empowerment:

> *Not because I do not trust my own hospital, but I just want to look further. What do other hospitals offer their patients? Is there anything I can take advantage of?* [S32; analog survivor of cancer]

To a lesser extent, analog survivors of cancer went on the web to seek reassurance about their future. They seemed to be reassured when encountering people with similar experiences. For example, one survivor stated the following:

> *Okay, I found something here, there are more people like me. Shared sorrow is half sorrow.* [S34; analog survivor of cancer]

Overall WHIS Patterns

The web-based source that analog patients eventually selected seemed to depend on their cancer-specific knowledge, cancer-related experience, and search experience. The use of cancer-specific knowledge and experience was reflected in selecting familiar and well-known websites about cancer. The use of search experience was reflected in analog patients using strategies that they reported to prefer (eg, preferring to use the search bar on specific websites instead of the regular search engine or the other way around). Analog patients mentioned different reasons for selecting content. The most prevalent reasons were familiarity with a website or organization (eg, the Dutch Cancer Society) or previous experience with a website. Some also mentioned that they selected certain websites as part of habitual behavior rather than for specific reasons. Notably, analog patients also visited websites while voicing doubt about their trustworthiness. It seemed that those analog patients thought that it was more important to find information relevant to fulfill their motives than looking for trustworthy information.

WHIS Approaches

In total, 2 overarching WHIS approaches could be identified: explorative and focused. Explorative approaches consisted of spontaneously selecting information seemingly without having an explicit information need. Analog patients who used this explorative approach mainly guided their searches by clicking on referral links and using suggestions made by search features on Google, such as the *autocomplete* (a feature within Google Search that makes it faster to complete searches that users start to type. Google's automated systems generate predictions that help users save time by allowing them to quickly complete the search they already intended to do) and *people also ask* (a feature within Google Search that provides users with additional questions related to their original search query and quick answers to them) functions. Analog patients were considered to use a focused approach when they seemed to search more purposefully (ie, mainly selecting information aligned with their verbally expressed specific information needs). For instance, an analog prediagnostic patient searched *symptoms of cancer* and exclusively selected content related to these search terms.

Unlike analog patients using an explorative approach, patients using a focused approach only made use of Google features when these explicitly helped them meet their self-reported information needs. For example, an analog patient with cancer searched for and read information about R-CHOP and subsequently encountered the following suggestions from the Google feature *people also ask*: *What does R-CHOP mean?* and *What is a CHOP cure?*

Several analog patients used both explorative and focused approaches. Some started with a clearly focused search strategy based on an information need but appeared to become emotionally distracted by the encountered content and started to use a more explorative approach. Others started with an explorative approach and were triggered by specific content that led them to adopt a new, more focused approach (eg, understanding difficult, complex words or confirming assumptions). In other words, information needs evolved while searching. WHIS approaches seemed independent of the disease stage that analog patients were allocated to.

Dissatisfying Content

All analog patients came across dissatisfying content while searching (in other words, content that did not satisfy the wishes of the patients). Examples of dissatisfying content were difficulty navigating systems on websites, cookies, or information not being in line with search motives. When this dissatisfying content was encountered, analog patients most often changed their search terms or quickly moved on to other web pages (the number of web pages visited ranged from 3 to 15 per session). Search terms were frequently changed during a search session (range 1-16 times per session), mostly because of dissatisfying content:

> *So, I'm not getting anywhere with this either, because I don't need to know what the cancer looks like...So I guess I'm not getting anywhere with this search term, with the search things. Uhm how am I going to do that?* [S35; analog survivor of cancer]

Impact of WHIS on Emotions and Dealing With Content

Emotions

Regardless of the stage of the disease, emotions were present throughout the entire search process, ranging from anxiety and worry to hope. These emotions fluctuated, and negative emotions were often induced when confrontational, complex, or unwanted information was found. Confrontational content included information on symptoms suggesting cancer or thyroid problems, information on treatment side effects such as hair loss and nausea, or a confronting picture:

> *I am not happy with the image I see here. That photo confirms the nightmare I have about chemotherapy. This is someone surrounded by nurses, being injected, and she has no hair, so that picture embodies for me everything that is wrong with this disease in one image. They have succeeded very skillfully in capturing all of that in one photo, but I do not think that was the intention of the person who took the photo. However, that is how it comes across at me: the embodiment of a mountain of misery.* [S27; analog patient with cancer]

Complex information included content containing medical jargon, such as *malignancies*; *cachexia*; or drug names such as *rituximab*, *cyclophosphamide*, and *hydroxydaunorubicin*. Most analog patients seemed to be affected by complex words:

> *This is getting annoying because I already see a word here that I do not know at all. I'm getting a lot of medical terms here that do not mean much to me...* [S06; analog prediagnostic patient]

Sometimes, positive emotions emerged from information that gave hope (eg, indolent NHL more often has a chance of recurrence than aggressive NHL). Moreover, analog patients who doubted their own navigation skills while searching on the web reported high levels of distress. Some of the analog patients also experienced cognitive dissonance (ie, a mental state of having conflicting beliefs, thoughts, values, or attitudes), as reflected in the following quote:

> *Everything in you says that it is better not to click on it, because you don't want to know it. But if you see the option then you just need to click on it.* [S27; analog patient with cancer]

Dealing With Emotionally Difficult Content

When encountering cognitively or emotionally difficult (or unwanted) information, analog patients with cancer dealt with the content in various ways. They adapted their search strategy, ignored the information by quickly clicking away from it and shifting toward other information, or stopped searching:

> *I immediately find myself with types of cancer, um...all the hits are related to Hodgkin; [scrolling back and forth through search results on the first Google page, but not clicking on anything]. Yes, I find this difficult; I think I will check the next Google pages to see what else comes up, what comes after Hodgkin.* [S01; analog prediagnostic patient]

Several analog patients also mentioned that they would normally seek information multiple times briefly or seek a distraction from the confronting information, such as watching Netflix or having some tea.

End of the Search

As mentioned previously, one of the reasons to stop searching was encountering cognitively or emotionally difficult information (confronting, upsetting, or confusing). This was mostly the case for analog prediagnostic patients and analog patients with cancer. The following quote illustrates this "overload":

> *Nothing [information found] makes me happy. Yeah, you can find information, but I believe I would make a cup of coffee now. I cannot say I'm a lot wiser now.* [S25; analog patient with cancer]

Another reason to stop searching was that analog patients saw their health care provider as a gatekeeper and their primary source of information about their disease and treatment. During the interview, they indicated that they preferred to talk with their clinician to clarify the encountered information instead of looking for more web-based information:

> *I believe that this information is quite overwhelming me right now, so I would put it away for a while. And I would talk it through first at a subsequent consultation with my doctor before I start worrying and assuming things that are not an issue at all...So, I think I will stop looking for now until I have spoken to the doctor again. It is a lot of information, and it is also complicated. So, I want to consult the doctor first.* [S23; analog patient with cancer]

All analog patients with cancer indicated ending their search sessions with many unanswered questions and an increase in uncertainty (compared to the start of the search). Unlike analog patients with cancer, analog prediagnostic patients and analog survivors of cancer ended their search more often with their information needs being fulfilled, as reflected in the following quote during the interview:

> *I do think it is very true. I'm at a point now where I do think: yeah, I'm reading this now, I'm not really getting very comfortable with this. I do not think there is any point in continuing to search now. I think I am now on a trustworthy site, and I find this a very upsetting story now that I see this. I cannot do much but wait and see. I don't know if I'm happy I've figured this out now...* [S01; analog prediagnostic patient]

Compared to analog patients in other disease stages, analog survivors of cancer ended their search most often satisfied and with more positive emotions; they felt less uncertain and found useful (practical) information on ways to cope with the future:

> *I definitely did become a bit wiser, because I can move on: I can go to physio, psychologist and I have a phone line which I can call.* [S36; analog survivor of cancer]

Discussion

Principal Findings

Using a comprehensive scenario-based, think-aloud approach, we were able to show that (1) patients' overarching motives for WHIS were mainly to reduce uncertainty, obtain reassurance, and increase empowerment; (2) these motives differed depending on the disease stage (at the beginning of the disease trajectory, patients mainly showed cognitive needs, whereas this shifted more toward affective needs in the subsequent disease stages); (3) analog patients' WHIS approaches varied from exploratory to focused to a combination of both; and (4) positive (hope and reassurance) and negative (anxiety and worry) emotional responses played an important role in patients' search strategies.

We found 3 overarching motives (ie, reducing uncertainty, obtaining reassurance, and increasing empowerment) for patients to search on the web. With these findings, we not only confirm the problem-solving model in the context of patient motivations to go on the web throughout their illness journey but also extend this model. According to Wilson [36], the process of problem-solving is the result of patients' wishes to reduce uncertainty. Patients' uncertainty at either the prediagnosis or treatment phase concerned various topics, clearly showing that these motives change over time. However, we also discovered 2 other important motives for patients to engage in a problem-solving process, namely, reassurance and empowerment [36]. In addition, the study's findings revealed a potential conflict between patient empowerment and uncertainty reduction in the context of WHIS. When patients seek web-based information to empower themselves, they gain a better understanding of their situation, which could enable them to ask informed questions to their clinicians. However, this increased knowledge may also give rise to new questions and uncertainties, leading to a potential challenge in fulfilling the motive of uncertainty reduction.

Moreover, our findings provide insights into the search behavior of patients with cancer at various stages of their disease trajectory and how these behaviors vary. In the initial phase of prediagnosis, patients often engaged in self-diagnosis. The results of this study extend those of previous research [9] by showing that patients prepare for a consultation by using the internet not only to help them formulate questions but also to self-diagnose. Despite the popularity of this search approach, research on self-diagnosing remains limited. In the context of web-based self-diagnosis for minor ailments, research shows that using the internet for self-diagnosis can be helpful as 44% of participants achieved accurate final diagnoses after searching the internet compared to 11% before searching the internet [37]. Another study shows that web-based self-diagnosing has the potential to empower patients in appraising and challenging clinicians' advice and requesting further diagnostic procedures [38]. However, web-based self-diagnosis can also be counterproductive if the patient misdiagnoses themselves, leading to unnecessary concerns. In addition, problems may occur if patients visit their clinician with a preconceived diagnosis, potentially causing disagreements about their condition [39]. During the treatment phase, the search strategy of patients with cancer focused on cognitive needs by seeking clarification, gathering more information, and preparing. However, we only observed a shift in search strategies toward affective needs by seeking emotional coping resources for dealing with the disease after patients completed treatments and were declared cancer free. In other words, at the beginning of the disease trajectory, analog patients had mainly cognitive needs, whereas analog survivors also showed affective needs and used the internet for emotional support. The change from more cognitive needs to more affective needs could be explained using the social-cognitive processing model. According to this model, seeking emotional support may facilitate emotional adjustment to traumatic experiences, such as cancer diagnosis and treatment [40]. Potentially, survivors have more mental space to cope with the situation and reflect on what has happened in the past months.

Our results further show that patients tend to use different search strategies: explorative, focused, or a combination of both. Previous research has demonstrated that individuals who are more exploratory seekers tend to tackle unfamiliar problems by using a broader search strategy (symptom exploration), resulting in a broader range of new information [37]. By encountering a broad range of information, patients are possibly confronted with new and unknown content, which could increase their level of uncertainty [41]. Our results also suggest that an exploratory search strategy increased the risk of being confronted with unwanted information. On the other hand, those who are more focused seekers tend to have a clear idea and a specific plan, leading them to research within a limited set of results (hypothesis testing) [37]. Such hypothesis testing can be problematic because it skews the way in which patients process information and distorts their perception of reality—a phenomenon known as confirmation bias [42]. It occurs when patients seek, interpret, or favor information that confirms their existing beliefs while ignoring or downplaying evidence that contradicts those beliefs [43]. Pang et al [41] argue that seekers within one internet visit alternate between exploratory and focused search strategies as new, unknown topics often lead to more exploratory searches. If the topic to be searched becomes clearer, the seeker may use a more focused approach. Our results confirm those of this previous study by showing that patients used both explorative and focused approaches. Some started with a focused search but became emotionally distracted and switched to an explorative approach. Others began exploratively and shifted to a focused search after encountering specific content.

Furthermore, our findings show that positive (hope and reassurance) and negative (anxiety and worry) emotional responses were present before, during, and after the search sessions. On the basis of patients' voiced thoughts and observed behavior, we conclude that these emotions impacted their search behavior. This is in line with the functionalist perspective of emotions, which argues that emotional responses may motivate people to behave in particular ways [44-46]. For instance, hope is seen as a motivating force that helps individuals move toward desired outcomes even in the face of uncertainty [47]. It is a future-oriented emotion as it involves visualization of positive future situations [48], and thus, hope could explain why patients are motivated to seek reassurance. Worry, on the other hand, is seen as an uncertainty-associated emotion and can increase a patient's desire for obtaining additional information [15]. Studies show a positive relationship between worry and the perceived need for additional information [49-51], and thus, worry could explain why patients are motivated to reduce uncertainty by searching for additional information. However, we also observed that patients who were worried ignored or avoided specific information. A possible explanation is that hope and worry are intertwined during WHIS [16]. Confronting or complex information poses a threat to hope, and thus, ignoring certain information may serve as a self-protective behavior to stay hopeful [16].

In our study, patients in the treatment phase were most worried after their search session. This is in contrast to existing literature indicating that perceived knowledge through web-based information seeking decreased patients' worry [15]. WHIS has also been found to help searchers fill information voids and enhance their coping abilities [52]. Although we did find some comparable results for the prediagnosis and survivor phases regarding decrease in worry and enhancing coping abilities, we did not find this for patients in the treatment phase. A possible explanation is that complex or confrontational information (eg, jargon for medicines and treatments and intense side effects) may have induced worries in analog patients in this phase. This inconsistency with the existing literature could further be explained by our design, which involved one search session only at one specific moment rather than multiple search sessions by one individual patient. Possibly, patients who search for more information at multiple times will eventually be less worried as they become more familiar with the difficult and complex information. Therefore, future research should investigate the longitudinal search behaviors of individual patients during their disease trajectory and the effects of multiple shorter search sessions within a particular disease phase.

Limitations and Strengths

First, a strength of our approach is that we not only observed patients' WHIS behaviors but simultaneously gained insights into their thoughts. During the interview, the interviewer made use of techniques such as paraphrasing and checking to clarify the meaning of the interviewee, thereby enhancing the validity of our findings. This innovative, comprehensive scenario-based, think-aloud approach exhibits strength in its consideration of the intuitive nature of web-based searching while overcoming challenges such as recall bias in retrospective methods. However, certain limitations should be considered. Some remarks suggested that participants may have felt limited in their choice of search engine and might have perceived an obligation to use a specific search platform, such as Google. Furthermore, during the think-aloud sessions, participants did not explore the use of social media channels (eg, Facebook, Instagram, or Twitter [subsequently rebranded X]). Use of social media may have been limited as participants could perceive it as an intrusion into their personal lives. Another reason could be that these communication channels may represent more spontaneous ways through which patients acquire unplanned or unexpected web-based health information while scrolling through their social media timeline [53]. The scenario-based, think-aloud approach as used in this study does not provide any insights in how social media has an effect on patients' WHIS strategies, motives, and emotions. Furthermore, the relatively small sample size used in this study calls for caution when generalizing the findings. It is important to account for variations in patients' (eHealth) literacy, education, and cultural backgrounds [54]. Although previous research demonstrates overlap in WHIS among patients from different countries, it also identifies distinct country-specific differences even when the countries have comparable welfare and health status [5]. As this study was an explorative qualitative study, and despite our relatively small sample size, we believe we achieved thematic saturation during the iterative process as no new codes emerged toward the end of our analysis. Moreover, it is important to bear in mind when interpreting the findings that our sample consisted of analog patients who were presented with a scenario. This may have biased our results as using analog patients is different from using patients with NHL. However, participants in this study possessed preexisting familiarity with cancer; our sample consisted of patients with cancer (other than NHL), survivors of cancer, and informal caregivers of patients with cancer. Thus, this sample's strength lies in their ability to strongly identify with the scenarios presented, which is also reflected in their quotes, the emotions showed during the think-aloud process, and their scores on the thermometers [24]. Furthermore, participants possessed experience in web-based cancer information seeking. Many of them were acquainted with patient advocacy organizations, and a subset even served as administrators for certain web-based platforms dedicated to cancer information and peer support groups. In addition, they had previously encountered medical terminology in the context of their own medical conditions, thus acquiring a degree of familiarity with medical jargon. Consequently, our sample likely possessed a higher level of proficiency in navigating the internet for cancer-related information compared to the average patient with cancer. Despite their advanced familiarity with the subject, the results still indicated that patients encountered difficulties in navigating the internet and understanding medical jargon.

Practical Implications

Knowing how patients with cancer search for web-based health information is a first step toward optimizing web-based health platforms such that patients with cancer can (more) easily find and navigate through information that fits their needs. On the basis of the study results, there are various implications for the development of cancer websites. First, web-based health platforms could use less complex words and show content warnings about confrontational prognostic or side effect–related information on web pages. The latter could warn searchers about unwanted information, which is especially relevant for exploratory searchers. Second, websites should enable users to self-pace and allow for user-initiated tailoring (ie, allowing users to tailor the information themselves based on their information needs). For example, information should be minimalized, with the possibility to read more if wanted (eg, with the use of hyperlinks). Third, it should also be clear to the user whether platforms are expert generated or peer generated as these platforms differ in content focusing on cognitive needs (addressing the needs of analog prediagnostic patients and analog patients with cancer) and affective needs (addressing the needs of analog survivors of cancer) [13]. In the Netherlands, multiple cancer platforms already make use of such features, which patients in our sample experienced as convenient. In addition to these implications for websites, another important finding is that patients see their health care providers as their primary source of information when it comes to their disease and treatment. Patients indicated that they had various remaining questions and considerable uncertainty after their search, which they wanted to resolve during their interaction with their health care provider. Therefore, it is important that, within consultations, there is room for questions arising from WHIS. Furthermore, health care providers can guide patients in the search process by giving tips and tricks on how (not) to use the internet to search for health information and how to cope with any uncertainty that may result from such a search.

Conclusions

This study provides valuable insights into the real-time WHIS strategies of patients with cancer, the motivations behind seeking web-based health information, and the emotions experienced at various stages of the disease trajectory. Understanding patients' search patterns is pivotal in optimizing web-based health platforms to cater to their specific needs. In addition, these findings can guide clinicians in directing patients toward reliable sources of web-based health information.

Acknowledgments

The authors would like to thank all participants for allowing them to watch how they searched the internet for health information while sharing their thoughts with them. Furthermore, the authors would like to thank Diana Aksan for her work in the selection process and pilot-testing of the protocol. This work was supported by the Dutch Cancer Society. No generative artificial intelligence was used in any portion of manuscript writing.

Data Availability

The datasets generated during and analyzed during this study are not publicly available due to the anonymity of the participants but are available from the corresponding author on reasonable request.

Authors' Contributions

The conceptualization of the study was carried out by ES, MH, JvW, and AL. ES acquired funding for this project. FH and PK gathered and analyzed the data, and AL was responsible for the validation process. The original draft of the manuscript was prepared by FH and AL. All authors reviewed the protocols, contributed to conceptualization and methodology, and reviewed the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
COREQ (Consolidated Criteria for Reporting Qualitative Research) checklist.
[DOCX File , 18 KB - infodemiology_v5i1e59625_app1.docx]

Multimedia Appendix 2
Final think-aloud protocol, including semistructured interview guide.
[DOCX File , 15 KB - infodemiology_v5i1e59625_app2.docx]

Multimedia Appendix 3
Think-aloud scenarios.
[DOCX File , 15 KB - infodemiology_v5i1e59625_app3.docx]

References

1. van Eenbergen MC, Vromans RD, Boll D, Kil PJ, Vos CM, Krahmer EJ, et al. Changes in internet use and wishes of cancer survivors: a comparison between 2005 and 2017. Cancer 2020 Jan 15;126(2):408-415 [FREE Full text] [doi: 10.1002/cncr.32524] [Medline: 31580497]
2. van Eenbergen MC, Vromans RD, Tick LW, Vreugdenhil G, Krahmer EJ, Mols F, et al. Comparing survivors of cancer in population-based samples with those in online cancer communities: cross-sectional questionnaire study. JMIR Cancer 2022 Mar 08;8(1):e19379 [FREE Full text] [doi: 10.2196/19379] [Medline: 35258460]
3. He Y, Wang Y, Luan F, Yu Z, Feng H, Chen B, et al. Chinese and global burdens of gastric cancer from 1990 to 2019. Cancer Med 2021 May;10(10):3461-3473 [FREE Full text] [doi: 10.1002/cam4.3892] [Medline: 33931958]
4. Capocaccia R, Colonna M, Corazziari I, de Angelis R, Francisci S, Micheli A, et al. Measuring cancer prevalence in Europe: the EUROPREVAL project. Ann Oncol 2002 Jun;13(6):831-839 [FREE Full text] [doi: 10.1093/annonc/mdf152] [Medline: 12123329]
5. Link E, Baumann E, Linn A, Fahr A, Schulz PJ, Abuzahra ME. Influencing factors of online health information seeking in selected European countries. Eur J Health Commun 2021 May 19;2(1):29-55. [doi: 10.47368/ejhc.2021.002]
6. Lazarus RS, Folkman S. Stress, Appraisal, and Coping. Cham, Switzerland: Springer; 1984.
7. Miller SM. When is a little information a dangerous thing? Coping with stressful events by monitoring versus blunting. In: Levine S, Ursin H, editors. Coping and Health. Cham, Switzerland: Springer; 1980:145-169.
8. Bol N, Linn AJ, Smets EM, Verdam MG, van Weert JC. Tailored communication for older patients with cancer: using cluster analysis to identify patient profiles based on information needs. J Geriatr Oncol 2020 Jul;11(6):944-950 [FREE Full text] [doi: 10.1016/j.jgo.2020.01.004] [Medline: 31974067]

9. Duimel SL, Linn AJ, Smets EM, Smit ES, van Weert JC. Profiling cancer patients based on their motives for seeking informational and emotional support online. Health Commun 2023 Dec;38(14):3223-3237. [doi: 10.1080/10410236.2022.2144287] [Medline: 36415021]
10. Lambert SD, Loiselle CG, Macdonald ME. An in-depth exploration of information-seeking behavior among individuals with cancer: part 1: understanding differential patterns of active information seeking. Cancer Nurs 2009;32(1):11-23. [doi: 10.1097/01.NCC.0000343372.24517.bd] [Medline: 19104197]
11. Bautista JR, Zhang Y, Gwizdka J, Chang YS. Consumers' longitudinal health information needs and seeking: a scoping review. Health Promot Int 2023 Aug 01;38(4):daad066. [doi: 10.1093/heapro/daad066] [Medline: 37432774]
12. Rutten LJ, Arora NK, Bakos AD, Aziz N, Rowland J. Information needs and sources of information among cancer patients: a systematic review of research (1980-2003). Patient Educ Couns 2005 Jun;57(3):250-261. [doi: 10.1016/j.pec.2004.06.006] [Medline: 15893206]
13. Sanders R, Linn AJ, Araujo TB, Vliegenthart R, van Eenbergen MC, van Weert JC. Different platforms for different patients' needs: automatic content analysis of different online health information platforms. Int J Hum Comput Stud 2020 May;137:102386. [doi: 10.1016/j.ijhcs.2019.102386]
14. Savolainen R. Information need as trigger and driver of information seeking: a conceptual analysis. ASLIB J Inf Manag 2017 Jan 16;69(1):2-21 [FREE Full text] [doi: 10.1108/AJIM-08-2016-0139]
15. Lee SY, Hawkins RP. Worry as an uncertainty-associated emotion: exploring the role of worry in health information seeking. Health Commun 2016 Aug;31(8):926-933. [doi: 10.1080/10410236.2015.1018701] [Medline: 26752071]
16. Leydon GM, Boulton M, Moynihan C, Jones A, Mossman J, Boudioni M, et al. Cancer patients' information needs and information seeking behaviour: in depth interview study. BMJ 2000 Apr 01;320(7239):909-913 [FREE Full text] [doi: 10.1136/bmj.320.7239.909] [Medline: 10742000]
17. Ofran Y, Paltiel O, Pelleg D, Rowe JM, Yom-Tov E. Patterns of information-seeking for cancer on the internet: an analysis of real world data. PLoS One 2012;7(9):e45921 [FREE Full text] [doi: 10.1371/journal.pone.0045921] [Medline: 23029317]
18. Schmier JK, Halpern MT. Patient recall and recall bias of health state and health status. Expert Rev Pharmacoecon Outcomes Res 2004 Apr;4(2):159-163. [doi: 10.1586/14737167.4.2.159] [Medline: 19807519]
19. Stehl ML, Kazak AE, Alderfer MA, Rodriguez A, Hwang WT, Pai AL, et al. Conducting a randomized clinical trial of an psychological intervention for parents/caregivers of children with cancer shortly after diagnosis. J Pediatr Psychol 2009 Sep;34(8):803-816 [FREE Full text] [doi: 10.1093/jpepsy/jsn130] [Medline: 19091806]
20. Bolle S, Romijn G, Smets EM, Loos EF, Kunneman M, van Weert JC. Older cancer patients' user experiences with web-based health information tools: a think-aloud study. J Med Internet Res 2016 Jul 25;18(7):e208 [FREE Full text] [doi: 10.2196/jmir.5618] [Medline: 27457709]
21. Marcu A, Muller C, Ream E, Whitaker KL. Online information-seeking about potential breast cancer symptoms: capturing online behavior with an internet browsing tracking tool. J Med Internet Res 2019 Feb 06;21(2):e12400 [FREE Full text] [doi: 10.2196/12400] [Medline: 30724741]
22. Blanch-Hartigan D. Patient satisfaction with physician errors in detecting and identifying patient emotion cues. Patient Educ Couns 2013 Oct;93(1):56-62. [doi: 10.1016/j.pec.2013.04.010] [Medline: 23669151]
23. van Vliet LM, van der Wall E, Albada A, Spreeuwenberg PM, Verheul W, Bensing JM. The validity of using analogue patients in practitioner-patient communication research: systematic review and meta-analysis. J Gen Intern Med 2012 Nov;27(11):1528-1543 [FREE Full text] [doi: 10.1007/s11606-012-2111-8] [Medline: 22700392]
24. Smets EM, van Vliet LM, Pieterse AH, Visser LN, Labrie NH, Hillen MA. Analogue patients. In: Ho EY, Bylund CL, van Weertj JC, Basnyat I, Bol N, Dean M, editors. The International Encyclopedia of Health Communication. Hoboken, NJ: John Wiley & Sons; 2023:1-10.
25. PanelCom. URL: https://www.panelcom.nl/ [accessed 2021-07-30]
26. Agree EM, King AC, Castro CM, Wiley A, Borzekowski DL. "It's got to be on this page": age and cognitive style in a study of online health information seeking. J Med Internet Res 2015 Mar 24;17(3):e79 [FREE Full text] [doi: 10.2196/jmir.3352] [Medline: 25831483]
27. Young K. Direct from the source: the value of 'think-aloud' data in understanding learning. J Educ Enquiry 2005;6(1):25 [FREE Full text]
28. In de wereld van kanker wijzen wij je de weg. kanker.nl. URL: https://www.kanker.nl/ [accessed 2021-07-30]
29. Kim J. Scenarios in information seeking and information retrieval research: a methodological application and discussion. Libr Inf Sci Res 2012 Oct;34(4):300-307. [doi: 10.1016/j.lisr.2012.04.002]
30. van Zuuren FJ, de Groot KI, Mulder NL, Peter M. Coping with medical threat: an evaluation of the Threatening Medical Situations Inventory (TMSI). Pers Individ Dif 1996 Jul;21(1):21-31. [doi: 10.1016/0191-8869(96)00029-3]
31. Bronner MB, Nguyen MH, Smets EM, van de Ven AW, van Weert JC. Anxiety during cancer diagnosis: examining the influence of monitoring coping style and treatment plan. Psychooncology 2018 Feb;27(2):661-667. [doi: 10.1002/pon.4560] [Medline: 28976610]
32. de Bruin GO, Rassin E, van der Heiden C, Muris P. Psychometric properties of a Dutch version of the intolerance of uncertainty scale. Neth J Psychol 2006 Dec;62(2):87-92. [doi: 10.1007/bf03061055]

33. Thomas R, Kaminski E, Stanton E, Williams M. Measuring information strategies in oncology - developing an information satisfaction questionnaire. Eur J Cancer Care (Engl) 2004 Mar;13(1):65-70. [doi: 10.1111/j.1365-2354.2004.00445.x] [Medline: 14961777]
34. van der Vaart R, van Deursen AJ, Drossaert CH, Taal E, van Dijk JA, van de Laar MA. Does the eHealth Literacy Scale (eHEALS) measure what it intends to measure? Validation of a Dutch version of the eHEALS in two adult populations. J Med Internet Res 2011 Nov 09;13(4):e86 [FREE Full text] [doi: 10.2196/jmir.1840] [Medline: 22071338]
35. Aymé S, Kole A, Groft S. Empowerment of patients: lessons from the rare diseases community. Lancet 2008 Jun 14;371(9629):2048-2051. [doi: 10.1016/S0140-6736(08)60875-2] [Medline: 18555918]
36. Wilson T. Exploring models of information behaviour: the ‘uncertainty’ project. Inf Process Manag 1999 Nov;35(6):839-849. [doi: 10.1016/S0306-4573(99)00029-1]
37. Kwakernaak J, Eekhof JA, de Waal MW, Barenbrug EA, Chavannes NH. Patients' use of the internet to find reliable medical information about minor ailments: vignette-based experimental study. J Med Internet Res 2019 Nov 11;21(11):e12278 [FREE Full text] [doi: 10.2196/12278] [Medline: 31710304]
38. Mueller J, Jay C, Harper S, Todd C. The role of web-based health information in help-seeking behavior prior to a diagnosis of lung cancer: a mixed-methods study. J Med Internet Res 2017 Jun 08;19(6):e189 [FREE Full text] [doi: 10.2196/jmir.6336] [Medline: 28596146]
39. Aboueid S, Liu RH, Desta BN, Chaurasia A, Ebrahim S. The use of artificially intelligent self-diagnosing digital platforms by the general public: scoping review. JMIR Med Inform 2019 May 01;7(2):e13445 [FREE Full text] [doi: 10.2196/13445] [Medline: 31042151]
40. Lepore SJ. A social–cognitive processing model of emotional adjustment to cancer. In: Baum A, Andersen BL, editors. Psychosocial Interventions for Cancer. Washington, DC: American Psychological Association; 2001:99-116.
41. Pang PC, Verspoor K, Chang S, Pearce J. Conceptualising health information seeking behaviours and exploratory search: result of a qualitative study. Health Technol 2015 Feb 10;5(1):45-55. [doi: 10.1007/S12553-015-0096-0]
42. Keselman A, Browne AC, Kaufman DR. Consumer health information seeking as hypothesis testing. J Am Med Inform Assoc 2008;15(4):484-495 [FREE Full text] [doi: 10.1197/jamia.M2449] [Medline: 18436912]
43. Meppelink CS, Smit EG, Fransen ML, Diviani N. "I was right about vaccination": confirmation bias and health literacy in online health information seeking. J Health Commun 2019;24(2):129-140. [doi: 10.1080/10810730.2019.1583701] [Medline: 30895889]
44. Lazarus RS. Emotion and Adaptation. Oxford, UK: Oxford Academic Press; 1991.
45. Frijda NH. The laws of emotion. Am Psychol 1988 May;43(5):349-358. [doi: 10.1037//0003-066x.43.5.349] [Medline: 3389582]
46. Izard CE. Differential emotions theory. In: Izard CE, editor. Human Emotions. Cham, Switzerland: Springer; 1977:43-66. [doi: 10.1007/978-1-4899-2209-0_3]
47. Lazarus RS. Hope: an emotion and a vital coping resource against despair. Soc Res 1999;66(2):653-678 [FREE Full text]
48. de los Santos TM, Nabi RL. Emotionally charged: exploring the role of emotion in online news information seeking and processing. J Broadcast Electron Media 2019 Mar 18;63(1):39-58. [doi: 10.1080/08838151.2019.1566861]
49. Hovick SR, Kahlor L, Liang MC. Personal cancer knowledge and information seeking through PRISM: the planned risk information seeking model. J Health Commun 2014 Apr;19(4):511-527 [FREE Full text] [doi: 10.1080/10810730.2013.821556] [Medline: 24433251]
50. ter Huurne EF, Griffin RJ, Gutteling JM. Risk information seeking among U.S. and Dutch residents. Sci Commun 2009 Feb 27;31(2):215-237. [doi: 10.1177/1075547009332653]
51. Yang ZJ, Kahlor L. What, me worry? The role of affect in information seeking and avoidance. Sci Commun 2013;35(2):189-212. [doi: 10.1177/1075547012441873]
52. Morahan-Martin JM. How internet users find, evaluate, and use online health information: a cross-cultural review. Cyberpsychol Behav 2004 Oct;7(5):497-510. [doi: 10.1089/cpb.2004.7.497] [Medline: 15667044]
53. Fergie G, Hilton S, Hunt K. Young adults' experiences of seeking online information about diabetes and mental health in the age of social media. Health Expect 2016 Dec;19(6):1324-1335 [FREE Full text] [doi: 10.1111/hex.12430] [Medline: 26647109]
54. Hofstede G. Dimensionalizing cultures: the Hofstede model in context. Online Read Psychol Cult 2011 Dec 01;2(1):22. [doi: 10.9707/2307-0919.1014]

Abbreviations

COREQ: Consolidated Criteria for Reporting Qualitative Research
NHL: non-Hodgkin lymphoma
R-CHOP: rituximab, cyclophosphamide, hydroxydaunorubicin, oncovin, and prednisone
WHIS: web-based health information seeking

Edited by R Cuomo; submitted 17.04.24; peer-reviewed by PCI Pang, W van Harten; comments to author 05.06.24; revised version received 30.07.24; accepted 21.11.24; published 16.01.25.

Please cite as:
Huijgens F, Kwakman P, Hillen M, van Weert J, Jaspers M, Smets E, Linn A
How Patients With Cancer Use the Internet to Search for Health Information: Scenario-Based Think-Aloud Study
JMIR Infodemiology 2025;5:e59625
URL: https://infodemiology.jmir.org/2025/1/e59625
doi:10.2196/59625
PMID:

Original Paper

Assessment of Reliability and Validity of Celiac Disease–Related YouTube Videos: Content Analysis

Yunus Halil Polat[1*], MD; Rasim Eren Cankurtaran[2*], MD

[1]Department of Gastroenterology, Ankara Training and Research Hospital, Ankara, Turkey

[2]Department of Gastroenterology, Ankara Etlik City Hospital, Ankara, Turkey

[*]all authors contributed equally

Corresponding Author:
Yunus Halil Polat, MD
Department of Gastroenterology
Ankara Training and Research Hospital
bağlıca neighbourhood 1336 street no:7/7 Etimesgut
PO:06870
Ankara, 06870
Turkey
Phone: 90 5556801815
Email: yunushpolat@gmail.com

Abstract

Background: YouTube is an increasingly used platform for medical information. However, the reliability and validity of health-related information on celiac disease (CD) on YouTube have not been determined.

Objective: This study aimed to analyze the reliability and validity of CD-related YouTube videos.

Methods: On November 15, 2023, a search was performed on YouTube using the keyword "celiac disease." This search resulted in a selection of videos, which were then reviewed by 2 separate evaluators for content, origin, and specific features. The evaluators assessed the reliability and quality of these videos using a modified DISCERN (mDISCERN) score, the *Journal of the American Medical Association* (*JAMA*) benchmark criteria score, the usefulness score, video power index (VPI), and the Global Quality Scale (GQS) score.

Results: In the analysis of 120 initially screened CD videos, 85 met the criteria for inclusion in the study after certain videos were excluded based on predefined criteria. While the duration of the videos uploaded by health care professionals was significantly longer than the other group (P=.009), it was concluded that the median scores for mDISCERN (4, IQR 4-5 vs 2, IQR 2-3; P<.001), GQS (4, IQR 4-5 vs 3, IQR 2-3; P<.001), *JAMA* (4, IQR 3-4 vs 2, IQR 2-3; P<.001), and usefulness (8, IQR 7-9 vs 6, IQR 3-6; P<.001) of the videos from this group were significantly higher than those from non–health care professionals. Video interaction parameters, including the median number of views, views per day, likes, dislikes, comments, and VPI, demonstrated no significant difference between the 2 groups.

Conclusions: This study showed that YouTube videos about CD vary significantly in reliability and quality depending on their source. Increasing the production of reliable videos by health care professionals may help to improve patient education and make YouTube a more reliable resource.

(JMIR Infodemiology 2025;5:e58615) doi:10.2196/58615

KEYWORDS

gastroenterology; celiac; YouTube; internet-based information; medical information; health-related; reliability; validity; quality; videos; celiac sprue; sprue; gluten enteropathy; cross-sectional

Introduction

Celiac disease (CD) is an autoimmune disorder that occurs in genetically predisposed individuals as a result of the immune reaction to gluten, primarily affecting the small intestine [1]. Symptoms range from asymptomatic to digestive problems and nutritional deficiencies due to malabsorption of nutrients. Treatment includes a gluten-free diet [1]. Over the past few decades, CD has been estimated to affect around 1% of the world's population [2]. Despite the increasing prevalence of

CD, the majority of the patients with CD remain undiagnosed [1].

In recent years, the internet has become an important source of health information for the public. It has been reported that 80% of internet users use social media (SM) platforms to get information about their disease. Patients with chronic diseases in particular are increasingly relying on SM platforms to manage their conditions [3]. In a recent study investigating the use of SM by patients with CD and parents of patients with CD, it was reported that 96% of participants used SM for disease management [4]. YouTube (Google), is one of the world's most popular video-sharing platforms. Currently, YouTube has more than 1 billion registered users, and billions of videos are watched every day, about 30 million of which are health-related. Health-related videos can be uploaded by anyone, but the content of these videos may contain inaccurate or misleading information without being reviewed by health care professionals.

There are studies in the literature evaluating the reliability and quality of YouTube videos for many diseases [5,6]. There are few studies evaluating CD-related YouTube videos [7,8]. However, one of these studies evaluated non-English language videos [8]. The other study did not measure CD-related YouTube videos with the tests developed for these studies and did not include videometric parameters (such as the number of likes and dislikes) in the evaluation [7]. Unlike previous studies, which either focused on non-English videos or lacked comprehensive quality metrics, this research provides a more robust and comparative analysis of CD-related video content on YouTube.

We could not find any studies in the literature that evaluated the reliability and validity of YouTube videos about CD. This study aims to evaluate the quality and reliability of YouTube videos about CD using validated scoring tools and detailed content analysis.

Methods

Study Design

In this cross-sectional study, videos were collected using the keyword "Celiac Disease" in YouTube's search engine on November 15, 2023. The search was conducted in a Google Chrome browser in incognito mode, logged out of any user account, and using a standard IP address in Turkey. This was chosen because it is the most common keyword that holistically assesses all aspects of the disease, such as clinical, pathogenesis, diet, and nutrition. YouTube's default relevance mode was used to simulate the average consumer's search habits. It is recognized that most viewers rarely venture beyond the first few pages of results. Therefore, the first 120 videos about CD were selected, similar to previous studies. Based on the search results, a total of 120 videos were saved for further analysis, ranging from the most viewed video to the least viewed video. Video sampling criteria were determined with reference to similar studies [5,9].

The following factors were considered as exclusion criteria in the research: (1) videos in languages other than English, (2) videos with muted or poor picture quality, (3) videos containing advertisements, (4) videos with content unrelated to CD, and (5) videos with repetitive content.

Data Review

Data such as video type (real and animation), video length (min), time since upload (d), number of views, number of daily views (number of views/d since upload), number of likes, number of daily likes (number of likes/d since upload), number of dislikes, and number of comments were recorded. In our study, we categorized video sources into two groups: educational content of health care professionals (doctors, academic institutions or professional organizations, and health-related websites) and personal narratives of non–health care professionals (patients, independent users). The videos were independently analyzed by 2 raters (YHP and REC) and coded according to the themes of "Educational content" and "Personal narratives." Discrepancies in coding were resolved through repetitive discussions and consensus, ensuring a reliable and consistent categorization process. This method of assessment has been used in similar studies of other diseases [10].

Video Usefulness

The usefulness score is a usefulness scale defined by Lee et al [11]. Each video is rated with a score between 0 and 10 depending on the content of the video, such as causes, symptoms, diagnosis, diagnosis, and recovery status. According to the total score obtained, it is categorized as follows: 0=not useful, 1-3=less useful, 4-7=useful, and 8-10=very useful.

Video Popularity

The video power index (VPI) developed by Erdem et al [12] shows the popularity of videos and has been used in many studies [9]. The VPI calculation is as follows: VPI=(×100/[number of likes+number of dislikes]) × (number of views/number of d since upload)/100.

Quality and Reliability Evaluation

The Global Quality Scale (GQS) assesses the quality by providing the interpretation and usefulness of the videos for patients based on the flow of information. GQS has a 5-point Likert structure according to the quality, flow, and ease of use of the analyzed videos [13]. As used in similar studies, scores 1-2 were considered as low quality (inadequate in terms of patient information, contains incomplete information), 3 as medium quality (video flow is poor, some information is available but important issues are not addressed), and 4-5 (contains sufficient and useful information for patients) as high quality [14].

The quality assessment included the *Journal of the American Medical Association* (*JAMA*) benchmark criteria for determining authorship, attribution, disclosure, and currency. Each of these criteria was given a score of 1, with a maximum score of 4 [15].

The mDISCERN scale developed by Charnock et al [16] and later adapted to YouTube videos by Singh et al [17] was used to assess the reliability of the videos. The mDISCERN scale consists of 5 questions and is a questionnaire about information sources, purpose, reliability, bias, additional sources, and areas of uncertainty. Each question can be answered yes or no. Each yes answer is worth 1 point and 5 points represent the highest quality.

The video content was evaluated and graded according to the most recent American College of Gastroenterology guidelines for the management of CD [18]. These guidelines emphasize accurate symptom identification, diagnostic criteria, and effective dietary management strategies. Videos were scored for reliability, usefulness, and consistency with evidence-based practice.

Statistical Analyses

The SPSS (version 25.0 for Windows; IBM Corp) package program was used. Continuous variables were evaluated using the Shapiro-Wilk test to determine whether they were normally distributed. Continuous variables are reported as median and IQR, while categorical variables are presented as counts and percentages. Chi-square tests were used to analyze categorical variables and Mann-Whitney *U* test for numerical variables. The significance level was set at P=.05 for all analyses.

Ethical Considerations

The study adhered to the ethical standards outlined in the Helsinki Declaration and complied with national regulations in the respective field. Since the study did not involve the use of human or animal data, ethics committee approval was not necessary. This study analyzed publicly available YouTube videos. No identifiable personal data was used, and all results are presented in aggregate. Therefore, formal ethics approval was not required.

Results

Main Characteristics of Videos and Video Analysis

In total, 120 videos were analyzed and 85 videos met the study criteria and were included. A total of 35 videos were excluded from the study, including 2 non-English language videos, 13 videos with repetitive content, 12 videos with advertising content, and 8 videos with poor picture and sound quality. Most (22/85, 25.9%) were published by universities and other organizations, and most (50/85, 59%) were uploaded by health care professionals. A total of 68.2% (58/85) of the videos consisted of real images. Descriptive statistics of the above characteristics and other variables are shown in Table 1.

Table 1. Main characteristics of the analyzed videos. Categorical variables are expressed as n (%), and numerical variables are expressed as median (Q1-Q3).

Characteristics	Values
Source, n (%)	
Physicians	12 (14
Universities and professional organizations	22 (26)
Health information websites	16 (19)
Independent users	16 (19)
Patient	19 (22)
Source, n (%)	
Health care professionals	50 (59)
Non–health care professionals	35 (41)
Image type	
Real image, n (%)	58 (68)
Animation, n (%)	27 (32)
Number of views, median (IQR)	17,026 (2860-46,358)
Number of likes, median (IQR)	306 (45-820)
Number of dislikes, median (IQR)	6 (1-20)
Duration (min), median (IQR)	6.3 (3.4-12.1)
Days on YouTube, median (IQR)	1381 (572-2290)
Number of comments, median (IQR)	27 (5-130)
Views per day, median (IQR)	13.1 (4-33.2)
Likes per day, median (IQR)	0.2 (0.1-0.7)

Content Analysis and Source Evaluation of Videos

In the health care professional group, most (37/85, 43.1%) of the videos were uploaded by universities and other organizations, whereas in the non–health care professional group, most (19/34, 55.9%) of the videos were uploaded by "patients" (P<.001). While the duration of the videos uploaded by health care professionals was significantly longer than the other group (P=.009), it was concluded that the median scores for mDISCERN (4, IQR 4-5 vs 2, IQR 2-3; P<.01), GQS (4, IQR 4-5 vs 3, IQR 2-3; P<.001), *JAMA* (4, IQR 3-4 vs 2, IQR 2-3; P<.001), and usefulness (8, IQR 7-9 vs 6, IQR 3-6; P<.001) of the videos from this group were significantly higher than those from non–health care professionals. (Tables 2 and 3)

Table 2. The average scales of the analyzed videos.

Video scales	Values, median (IQR)
mDISCERN[a]	3 (3-4)
GQS[b]	4 (3-4)
JAMA [c]	3 (2-4)
VPI[d]	12.8 (4-33)
Usefulness	7 (5-9)

[a]mDISCERN: modified DISCERN score.
[b]GQS: Global Quality Scale score.
[c]*JAMA*: Journal of the American Medical Association.
[d]VPI: video power index.

Table 3. Comparison of videos according to source status. Categorical variables are expressed as n (%), and numerical variables as median (Q1-Q3).

Variables	Source		*P* value
	Health care professionals	Non–health care professionals	
Image			
Real image, n (%)	31 (62)	27 (77.1)	.21
Animation, n (%)	19 (38)	8 (22.9)	
Number of views, median (IQR)	16,657 (4858-57,896)	17,851.5 (1907-43,310)	.87
Number of likes, median (IQR)	297 (52-774)	373 (22-846)	.67
Number of dislikes, median (IQR)	6 (1-24)	8.5 (0-18)	.92
Duration (min), median (IQR)	7.4 (4.2-16.4)	3.9 (2.5-8.2)	.009
Days on YouTube, median (IQR)	1291 (516-2290)	1467.5 (832-2470)	.64
Number of comments, median (IQR)	21 (6-79)	67 (3-170)	.52
View per day, median (IQR)	12.8 (4.6-40.9)	15.6 (2.1-33.2)	.50
Like per day, median (IQR)	0.23 (0.07-1)	0.18 (0.03-0.73)	.39
mDISCERN[a], median (IQR)	4 (4-5)	2 (2-3)	<.001
GQS[b], median (IQR)	4 (4-5)	3 (2-3)	<.001
JAMA[c], median (IQR)	4 (3-4)	2 (2-2)	<.001
VPI[d], median (IQR)	12.3 (4.6-41)	15.3 (2.1-33)	.72
Usefulness, median (IQR)	8 (7-9)	5 (3-6)	<.001

[a]mDISCERN: modified DISCERN score.
[b]GQS: Global Quality Scale score.
[c]*JAMA: Journal of the American Medical Association.*
[d]VPI: video power index.

Themes Identified in Videos

From the 85 included videos, two major themes were identified.

Educational Content

These videos, primarily created by health care professionals, provided detailed information about CD symptoms, diagnosis, treatment, and long-term management. This category accounted for 59% (50/85) of all videos and demonstrated significantly higher scores in quality and reliability metrics (mDISCERN, GQS, *JAMA*, and Usefulness; P<.001).

Personal Narratives

Uploaded by patients or non–health care professionals, these videos focused on personal journeys, sharing challenges, and tips for living with CD. They received moderate interaction metrics (likes, comments) but were lower in quality and reliability scores (*P*<.001).

Discussion

Principal Findings

In this study, we analyzed YouTube videos about CD, an important disease that can occur at any age. We found that CD videos uploaded by health care professionals were significantly more reliable, adequate, useful, and quality information sources than those uploaded by non–health care professionals. Another striking result of the study was that there was no difference in video interaction parameters between those with and without health care professionals as video sources.

Recently, SM has become a popular way to access medical information and knowledge. Patients with many chronic diseases, including CD, have been reported to use SM as a source of information since adolescence [19]. Especially YouTube, a video sharing website, has become an important source of information in the field of health. In a recent nationally based survey study, it was reported that younger patient groups and patients with chronic diseases such as hypertension, diabetes mellitus, and chronic lung disease were more likely to watch YouTube videos as a source of health-related information [20].

As in other chronic diseases, SM use among patients with CD and their families has become widespread in recent years [4]. When we consider the importance of increasing adherence to a gluten-free diet as well as the diagnosis, risk factors, and clinical presentation of the disease, access to real and adequate information through SM becomes even more important. In a recent survey of patients with CD, two-thirds of the patients used SM every day for an average of 60 minutes per day. The 3 most common reasons for using SM were researching gluten-free diet products, obtaining information about diet, and CD. In the study, it was stated that the most frequently used platform was WhatsApp (Meta), and it was suggested that YouTube usage was 4% [4]. Although this rate may vary according to regional and cultural differences, it is still a relatively low rate and suggests that the use of YouTube may be higher than this data. In another similar survey study conducted in Japan, 27% of more than 2000 participants with chronic diseases stated that they used the YouTube platform related to their disease [20].

One of the studies evaluating YouTube videos on CD was a study in which 100 videos were evaluated in 2019. In this study, it was examined whether there was a difference between sources in 31 different topics such as etiology, symptoms, diagnosis, and treatment of the disease, and it was stated that there was no significant difference in terms of content in all remaining topics except 3 [7]. However, none of the video reliability-efficacy tests used in our study were used in this study. Nevertheless, it differs from our study because it claims that there is mostly no significant difference between videos whose source is health care professionals and other videos in terms of topics. Another study in the literature evaluated Polish-language videos, so it does not seem possible to make a comparison with our study [8].

Among the videos analyzed in our study, the fact that the reliability, usefulness, and quality scores of the videos of health care professionals were significantly higher than those of non–health care professionals was also observed in similar studies evaluating other diseases [21]. One of the most remarkable findings of our study is that there was no significant difference between the groups in terms of views, likes, dislikes, and VPI. There are many factors that can contribute to this, such as the visual presentation of the video, the demographic and cultural make-up of the viewers, the video's viral status, and the influencer's effect [22,23]. In a recent study investigating the influencer effect on SM related to dermatology, it was shown that dermatologists without competence and certification had as high a level of interaction as those with competence and certification [23]. This finding shows us that videos that may be insufficient as a source of information may also have high interaction and accordingly may cause misinformation and negative effects on patients and their families.

Based on these findings, we believe that in order for YouTube to be an accurate source of information about CD, many organizations and institutions, such as professional associations and universities, should provide training for health care professionals to produce high-quality videos that can provide more interaction and raise awareness among health care professionals about this issue. On the other hand, it is also important to raise patient awareness of the possibility that patients may be exposed to misinformation when using YouTube. We think that more use of YouTube and other SM platforms by health care professionals and peer review of health-related video content may reduce misinformation.

Limitations

There were some limitations in our study. The first 120 videos searched with the keyword "Celiac disease'" in the search results were analyzed and the other videos were not analyzed. In addition, since YouTube is a dynamic SM platform, video interaction parameters such as daily views, likes, and dislikes can change every day. Finally, the fact that only English videos were analyzed in our study can be considered among the limitations.

Conclusions

This study showed that YouTube videos about CD vary significantly in reliability and quality depending on their source. Increasing the production of reliable videos by health care professionals may help to improve patient education and make YouTube a more reliable resource.

Authors' Contributions

REC and YHP contributed to conceptualization, resources, data curation, writing (original draft preparation and review & editing), formal analysis, project administration, software, validation, and visualization. REC was responsible for developing the methodology, conducting the statistical analyses, and investigating and supervising the project.

Conflicts of Interest

None declared.

References

1. Lebwohl B, Sanders DS, Green PHR. Coeliac disease. Lancet 2018;391(10115):70-81. [doi: 10.1016/S0140-6736(17)31796-8] [Medline: 28760445]
2. Singh P, Arora A, Strand TA, Leffler DA, Catassi C, Green PH, et al. Global prevalence of celiac disease: systematic review and meta-analysis. Clin Gastroenterol Hepatol 2018;16(6):823-836.e2. [doi: 10.1016/j.cgh.2017.06.037] [Medline: 29551598]
3. Fox S. Online health search 2006. Pew Research Center. 2006. URL: https://www.pewresearch.org/internet/2006/10/29/online-health-search-2006/ [accessed 2025-01-22]
4. Al Sarkhy A. Social media usage pattern and its influencing factors among celiac patients and their families. Saudi J Gastroenterol 2020;26(2):99-104 [FREE Full text] [doi: 10.4103/sjg.SJG_495_19] [Medline: 32031161]
5. Karakoyun A, Yildirim A. YouTube videos as a source of information concerning Behçet's disease: a reliability and quality analysis. Rheumatol Int 2021;41(12):2117-2123. [doi: 10.1007/s00296-021-05009-9] [Medline: 34590188]
6. Tutan D, Kaya M. Evaluation of YouTube videos as a source of information on hepatosteatosis. Cureus 2023;15(10):e46843 [FREE Full text] [doi: 10.7759/cureus.46843] [Medline: 37829652]
7. Basch CH, Hillyer GC, Garcia P, Basch CE. Content of widely viewed YouTube videos about celiac disease. Public Health 2019;167:147-151. [doi: 10.1016/j.puhe.2018.11.004] [Medline: 30682698]
8. Kiedrowski M, Mróz A, Gajewska D, Nurzyński P, Deptała A. Celiac disease on YouTube - a study of the Polish content available on the popular video-sharing website. Pol Merkur Lekarski 2017;43(256):168-171. [Medline: 29084190]
9. Pons-Fuster E, Ruiz Roca J, Tvarijonaviciute A, López-Jornet P. YouTube information about diabetes and oral healthcare. Odontology 2020;108(1):84-90. [doi: 10.1007/s10266-019-00445-3] [Medline: 31396751]
10. Uz C, Umay E, Ballı Uz F, Cankurtaran D, Gundogdu I. Assessment of the quality and reliability of the information on spasticity on YouTube. J Clin Neurosci 2023;113:142-146. [doi: 10.1016/j.jocn.2023.05.018] [Medline: 37269749]
11. Lee JS, Seo HS, Hong TH. YouTube as a source of patient information on gallstone disease. World J Gastroenterol 2014;20(14):4066-4070 [FREE Full text] [doi: 10.3748/wjg.v20.i14.4066] [Medline: 24744597]
12. Erdem MN, Karaca S. Evaluating the accuracy and quality of the information in kyphosis videos shared on YouTube. Spine (Phila Pa 1976) 2018;43(22):E1334-E1339. [doi: 10.1097/BRS.0000000000002691] [Medline: 29664816]
13. Bernard A, Langille M, Hughes S, Rose C, Leddin D, Veldhuyzen van Zanten S. A systematic review of patient inflammatory bowel disease information resources on the World Wide Web. Am J Gastroenterol 2007;102(9):2070-2077. [doi: 10.1111/j.1572-0241.2007.01325.x] [Medline: 17511753]
14. Kocyigit BF, Akaltun MS, Sahin AR. YouTube as a source of information on COVID-19 and rheumatic disease link. Clin Rheumatol 2020;39(7):2049-2054 [FREE Full text] [doi: 10.1007/s10067-020-05176-3] [Medline: 32447603]
15. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the internet. JAMA 1997;277(15):1244-1245. [doi: 10.1001/jama.1997.03540390074039]
16. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999;53(2):105-111 [FREE Full text] [doi: 10.1136/jech.53.2.105] [Medline: 10396471]
17. Singh AG, Singh S, Singh PP. YouTube for information on rheumatoid arthritis--a wakeup call? J Rheumatol 2012;39(5):899-903. [doi: 10.3899/jrheum.111114] [Medline: 22467934]
18. Rubio-Tapia A, Hill ID, Semrad C, Kelly CP, Greer KB, Limketkai BN, et al. American College of Gastroenterology guidelines update: diagnosis and management of celiac disease. Am J Gastroenterol 2023;118(1):59-76. [doi: 10.14309/ajg.0000000000002075] [Medline: 36602836]
19. De Nardi L, Trombetta A, Ghirardo S, Genovese MRL, Barbi E, Taucar V. Adolescents with chronic disease and social media: a cross-sectional study. Arch Dis Child 2020;105(8):744-748. [doi: 10.1136/archdischild-2019-317996] [Medline: 31941715]
20. Mitsutake S, Takahashi Y, Otsuki A, Umezawa J, Yaguchi-Saito A, Saito J, INFORM Study Group. Chronic diseases and sociodemographic characteristics associated with online health information seeking and using social networking sites: nationally representative cross-sectional survey in Japan. J Med Internet Res 2023;25:e44741 [FREE Full text] [doi: 10.2196/44741] [Medline: 36862482]
21. Onder ME, Zengin O. Quality of healthcare information on YouTube: psoriatic arthritis. Z Rheumatol 2023;82(Suppl 1):30-37 [FREE Full text] [doi: 10.1007/s00393-021-01069-1] [Medline: 34468808]

22. Rao A, Morstatter F, Lerman K. Partisan asymmetries in exposure to misinformation. Sci Rep 2022;12(1):15671 [FREE Full text] [doi: 10.1038/s41598-022-19837-7] [Medline: 36123387]
23. O'Brien KF, Newsom EC, Park JH, Lawrence N. Increasing a dermatologist's footprint on Instagram: an analysis of top influencers performing nonsurgical cosmetic procedures. Dermatol Surg 2021;47(8):1093-1097. [doi: 10.1097/DSS.0000000000003088] [Medline: 33988555]

Abbreviations

CD: celiac disease
GQS: Global Quality Scale
***JAMA*:** *Journal of the American Medical Association*
mDISCERN: modified DISCERN
SM: social media
VPI: video power index

Edited by T Mackey; submitted 20.03.24; peer-reviewed by S Guandalini, H Meteran; comments to author 24.05.24; revised version received 03.06.24; accepted 05.01.25; published 04.02.25.

Please cite as:
Polat YH, Cankurtaran RE
Assessment of Reliability and Validity of Celiac Disease–Related YouTube Videos: Content Analysis
JMIR Infodemiology 2025;5:e58615
URL: https://infodemiology.jmir.org/2025/1/e58615
doi:10.2196/58615
PMID:39756057

Original Paper

Unveiling Topics and Emotions in Arabic Tweets Surrounding the COVID-19 Pandemic: Topic Modeling and Sentiment Analysis Approach

Farah Alshanik[1], PhD; Rawand Khasawneh[2], PharmD; Alaa Dalky[3], PhD; Ethar Qawasmeh[4], MS

[1]Department of Computer Science, Faculty of Computer and Information Technology, Jordan University of Science and Technology, Irbid, Jordan

[2]Department of Clinical Pharmacy, Faculty of Pharmacy, Jordan University of Science and Technology, Irbid, Jordan

[3]Department of Health Management and Policy, Faculty of Medicine, Jordan University of Science and Technology, Irbid, Jordan

[4]Department of Computer Science, Faculty of Computer Science and Engineering, The Ohio State University, Columbus, OH, United States

Corresponding Author:
Farah Alshanik, PhD
Department of Computer Science
Faculty of Computer and Information Technology
Jordan University of Science and Technology
Alhusun St
Irbid, 22110
Jordan
Phone: 962 2 7201000 ext 23130
Fax: 962 2 7095123
Email: fmalshanik@just.edu.jo

Abstract

Background: The worldwide effects of the COVID-19 pandemic have been profound, and the Arab world has not been exempt from its wide-ranging consequences. Within this context, social media platforms such as Twitter have become essential for sharing information and expressing public opinions during this global crisis. Careful investigation of Arabic tweets related to COVID-19 can provide invaluable insights into the common topics and underlying sentiments that shape discussions about the COVID-19 pandemic.

Objective: This study aimed to understand the concerns and feelings of Twitter users in Arabic-speaking countries about the COVID-19 pandemic. This was accomplished through analyzing the themes and sentiments that were expressed in Arabic tweets about the COVID-19 pandemic.

Methods: In this study, 1 million Arabic tweets about COVID-19 posted between March 1 and March 31, 2020, were analyzed. Machine learning techniques, such as topic modeling and sentiment analysis, were applied to understand the main topics and emotions that were expressed in these tweets.

Results: The analysis of Arabic tweets revealed several prominent topics related to COVID-19. The analysis identified and grouped 16 different conversation topics that were organized into eight themes: (1) preventive measures and safety, (2) medical and health care aspects, (3) government and social measures, (4) impact and numbers, (5) vaccine development and research, (6) COVID-19 and religious practices, (7) global impact of COVID-19 on sports and countries, and (8) COVID-19 and national efforts. Across all the topics identified, the prevailing sentiments regarding the spread of COVID-19 were primarily centered around anger, followed by disgust, joy, and anticipation. Notably, when conversations revolved around new COVID-19 cases and fatalities, public tweets revealed a notably heightened sense of anger in comparison to other subjects.

Conclusions: The study offers valuable insights into the topics and emotions expressed in Arabic tweets related to COVID-19. It demonstrates the significance of social media platforms, particularly Twitter, in capturing the Arabic-speaking community's concerns and sentiments during the COVID-19 pandemic. The findings contribute to a deeper understanding of the prevailing discourse, enabling stakeholders to tailor effective communication strategies and address specific public concerns. This study underscores the importance of monitoring social media conversations in Arabic to support public health efforts and crisis management during the COVID-19 pandemic.

(JMIR Infodemiology 2025;5:e53434) doi:10.2196/53434

KEYWORDS
topic modeling; sentiment analysis; COVID-19; social media; Twitter; public discussion

Introduction

Background

Throughout history, humanity has faced numerous outbreaks of infectious diseases that have resulted in significant loss of life and economic impact. Toward the end of 2019, the World Health Organization reported a series of pneumonia cases in Wuhan, which were later identified as COVID-19. As a novel infectious disease transmitted through respiratory droplets and contact, COVID-19 quickly spread across the globe, leading to an unprecedented impact on global public health, businesses, and economies. As of February 7, 2023, there have been >676 million confirmed cases and 500,000 reported deaths in >200 countries [1]. Social media platforms, particularly Twitter, have emerged as valuable sources of information for understanding and predicting disease outbreaks. Text mining techniques allow for the extraction of relevant health information from user-generated content on social media platforms. Twitter, in particular, provides researchers with vast amounts of real-time data, enabling early response strategies and enhancing situational awareness. Analyzing Twitter data has become a crucial area of focus in medical informatics research [2,3].

COVID-19 emerged as a prominent and sustained topic on Twitter starting from January 2020, and its discussion has persisted uninterrupted up to the present day [4]. With quarantine measures implemented worldwide, individuals increasingly relied on social media to access news and express their opinions. Twitter data offer valuable insights into public discussions, sentiments, and real-time updates during global pandemics [2,5]. Using Twitter as a data source enables infodemiology studies, providing health authorities with opinions and concerns to inform their responses [6].

Since the outset of the COVID-19 outbreak, an escalating number of studies have been harnessing Twitter data to delve into the public's reactions and discussions surrounding the COVID-19 pandemic. In their respective studies, researchers used distinct methodologies to explore COVID-19–related discussions and sentiments. For instance, Xue et al [4,7] used latent Dirichlet allocation (LDA) for topic identification. Similarly, a study by Alharbi and Alkhateeb [8] investigated the sentiment of the Arabic public on Twitter, using natural language processing (NLP) and machine learning techniques, finding that the long short-term memory model outperformed the naive Bayes model with an accuracy rate of 99% [8]. Another study focused on Arabic sentiment analysis for vaccine-related COVID-19 tweets, introducing the first and largest human-annotated dataset in Arabic for this purpose; it used advanced models such as the stacked gated recurrent unit and AraBERT, achieving a 7% accuracy enhancement [9]. During the COVID-19 pandemic, a separate study analyzed online learning–related tweets in Arabic, using various classification algorithms and achieving a maximum accuracy of approximately 89.6% using the Support Vector Machine classifier to analyze public perceptions of the coronavirus [10].

In addition, research conducted in Saudi Arabia showed a significant increase in negative sentiments during the COVID-19 pandemic, with deep learning algorithms achieving high accuracy rates [11]. Other studies explored sentiment differences between countries and in response to events, using topic modeling and sentiment analysis to reveal previously unreported patterns [12]. Furthermore, a study from Morocco compared different machine learning algorithms for tweet classification, finding logistic regression to yield the best sentiment predictions [13].

Recent advancements in NLP have shown significant potential in transforming various aspects of health care, including clinical decision support, patient management, and automated analysis of health records. Recent studies, such as the one by Tamang et al [14], highlight the use of NLP for optimizing patient outcome predictions and identifying disease patterns through electronic health record data. Similarly, a study by Elbattah et al [15] explores the role of NLP in extracting actionable insights from unstructured medical texts, further underscoring the growing relevance of NLP in enhancing the health care decision-making processes.

COVID-19 remains a scientifically and medically novel disease that requires in-depth and consistent research. Leveraging social media data, particularly from platforms such as Twitter, is essential for syndromic surveillance and understanding public health–related concerns. Twitter, as a prominent communication modality during disease outbreaks, offers valuable insights into public awareness and provides real-time reflections of public sentiment. Despite extensive research on COVID-19, limited studies have used social media data, specifically Twitter, to address conclusive themes and sentiment analysis in Arab regions during the early stages of the COVID-19 pandemic.

While numerous studies have investigated similar themes in different languages and contexts, there remains a notable gap in the analysis of Arabic tweets [16-22]. The Arabic-speaking population plays a significant role in the global discourse on COVID-19, and their perspectives and sentiments warrant dedicated exploration. Building on previous research, and to bridge this gap, our study used a combination of topic modeling techniques, specifically LDA, and sentiment analysis methods to uncover the predominant topics of discussion and the prevailing emotional tones within this corpus.

This Study

This study aims to analyze Twitter posts during the early stages of the COVID-19 pandemic in Arab regions to provide valuable insights into public sentiment, concerns, and awareness regarding COVID-19 in Arab communities. To achieve this, >1 million tweets posted between March 1 and March 31, 2020, were collected and analyzed. Through this analysis, we hope to assist policy makers in making informed decisions, enhancing public health communication, and implementing effective interventions to mitigate the impact of future outbreaks.

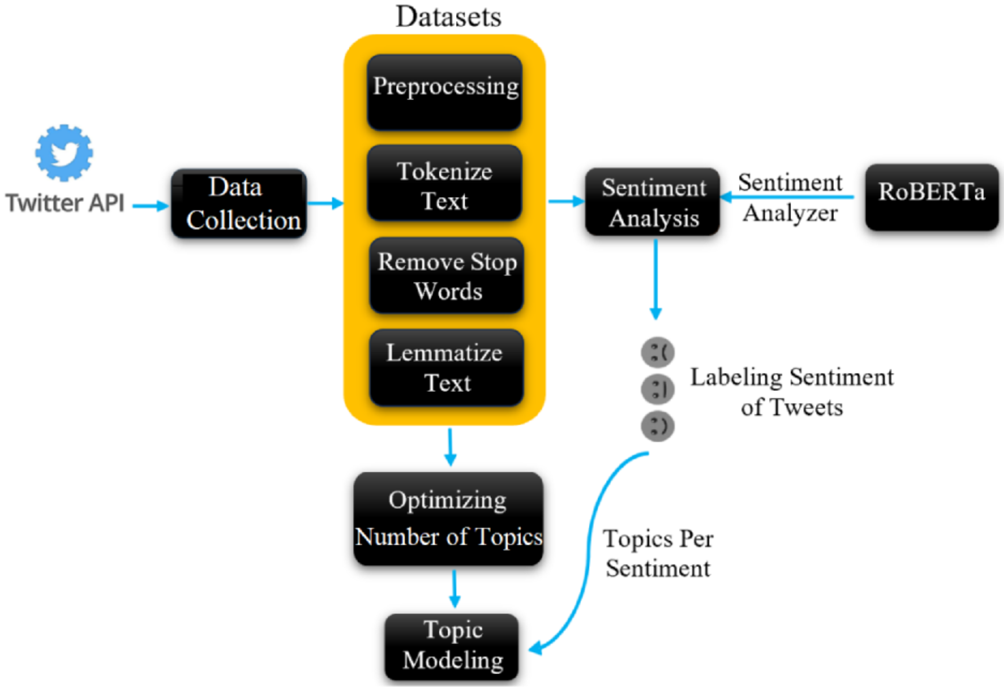
Although this study was conducted during the COVID-19 pandemic, its scope extends beyond the immediate implications of the COVID-19 pandemic. The primary goal of this research is to enhance health care planning and resource allocation in Jordan, which remains a critical issue regardless of pandemic conditions. The findings are designed to inform strategies that could be beneficial in various health care scenarios, whether in routine health care management or in response to other emergent public health challenges. Therefore, the study's relevance persists even in a postpandemic context, making it valuable for long-term health care system improvements.

Methods

Research Design

This study uses LDA for topic modeling and a sentiment analysis emotion detection tool to uncover topics and emotions in Twitter data related to COVID-19 in the Arab region. The methodological flowchart is depicted in Figure 1. Our approach to mining Twitter data adheres to the following 4 primary steps: data collection, data preprocessing, sentiment analysis, and topic modeling. The flowchart in Figure 1 illustrates how these steps are interconnected and carried out in our data analysis pipeline. Through these methods, we aim to gain valuable insights into the topics of discussion and the emotional responses of individuals in the Arab region concerning the COVID-19 pandemic.

Figure 1. Twitter data mining pipeline. API: application programming interface.



Data Collection

In our research, we harnessed the GeoCoV19 dataset, a multilingual COVID-19 Twitter dataset that spans a significant period of 90 days, from February 1, 2020, to May 1, 2020. This extensive dataset comprises hundreds of millions of tweets and is enriched with a diverse set of multilingual hashtags and keywords to ensure its comprehensiveness [23]. The dataset primarily provides tweet IDs, which presented us with the task of retrieving the actual tweet text associated with these IDs. To accomplish this, we made effective use of the Twarc application programming interface (API), a robust and efficient tool explicitly designed for this purpose [24]. The Twarc library was chosen due to its robustness in handling large-scale data collection, effective management of Twitter's API rate limits, seamless integration with existing data pipelines, and support for extended tweet metadata, making it an ideal tool for ensuring the integrity and completeness of the dataset required for this study. The Twarc API streamlined the process of collecting tweet texts corresponding to the tweet IDs provided. As we gathered all the tweets, we applied a language filter to focus exclusively on Arabic tweets. This selective filtering step was crucial for tailoring the dataset to our specific analysis, concentrating on tweets in the Arabic language.

Data Preprocessing

Data preprocessing plays a pivotal role in text mining, and it serves as a fundamental step in this domain. The purpose of this preprocessing is 2-fold: it optimizes the efficiency of prediction algorithms by eliminating potentially detrimental words, and it conserves storage space, contributing to improved computational performance [25]. In our analysis, we worked with Arabic text data, which requires thorough preprocessing to filter out any noise or irrelevant elements. The initial raw Arabic text underwent a series of transformations as part of this preprocessing effort. These transformations involved tokenization and the removal of various elements such as white spaces, punctuation marks, special characters, emojis, and URLs.

To accomplish this, we used a set of established methods for Arabic text preprocessing, including the use of Farasa [26]. Farasa proved invaluable in normalizing Arabic characters, stripping away diacritics, erasing punctuation marks, and eliminating repetitive characters, collectively enhancing the quality and relevance of the text data for our analysis.

Sentiment Analysis

Overview

To classify the primary sentiments expressed in Twitter messages, such as fear and joy, we used sentiment analysis, an NLP technique [27]. Our approach involved deploying the RoBERTa-base model, meticulously trained on a vast corpus of approximately 58 million tweets and further fine-tuned for precise emotion recognition leveraging the TweetEval benchmark [28]. This specific model, known as Twitter-RoBERTa-Base-Emotion [29], has been purposefully tailored for the nuanced task of emotion recognition within Twitter text data. It adeptly classifies text into various emotion categories, including joy, sadness, anger, fear, surprise, disgust, anticipation, and trust. Our sentiment analysis process unfolded in a sequence of four distinct steps, described in the following sections.

Step 1: Translation to English

As a reliable Arabic emotion detection API was not readily available, we initiated the process by translating Arabic tweets to English. To accomplish this, we leveraged the Google Translation API. We established an account and procured the necessary translation service. It is worth noting that the cost associated with using the Google Translation API amounts to US $20 per 1 million characters. Given that we were dealing with a substantial volume of data, encompassing 5.1 million Arabic tweets with a staggering 970,801,329 characters, the estimated cost tallied up to US $19,420. Consequently, we opted to translate 1 entire month of tweets. March was selected as the ideal candidate for translation, primarily due to its status as the month with the highest tweet volume. In addition, March witnessed several pivotal events, including Trump's declaration of COVID-19 as a national emergency, the implementation of travel bans on non-US citizens traveling from Europe, and the World Health Organization's formal declaration of the coronavirus as a global pandemic. To verify the quality of the translations, a sample of 5000 tweets was randomly selected and evaluated both before and after translation. Bilingual experts reviewed these tweets, comparing the original Arabic content with the translated English text. This review process focused on ensuring that the translations accurately conveyed the original meaning, context, and sentiment. On the basis of their feedback, we confirmed that the translations were of high quality, making them suitable for further analysis.

Step 2: English Text Preprocessing

Once the translation was complete, we embarked on preprocessing the English text. This entailed removing common stop words such as "and," "the," and "to."

Step 3: Stemming

To further refine the text data, we applied a stemming process, which involves eliminating predefined prefixes and suffixes. This step aids in reducing words to their root form. For instance, it transforms "running" into "run" through stemming.

Step 4: Emotion Determination

The final step involved determining the emotion expressed in the tweets using Twitter-RoBERTa-Base-Emotion.

Table 1 illustrates the distribution of emotions across the analyzed tweets, providing valuable insights into the prevailing sentiments during the specified time frame.

Table 1. Number of tweets per emotion.

Emotion	Tweets, n
Anger	182,105
Disgust	150,022
Joy	141,446
Anticipation	60,449
Sadness	44,591
Surprise	30,666
Fear	28,439

Topic Modeling Using LDA

In our analysis, we harnessed the power of LDA as a formidable tool for uncovering latent topics within our extensive dataset. LDA, a generative probabilistic model, proves exceptionally useful for extracting these hidden themes from a vast collection of documents. Its underlying mechanism involves representing documents as random combinations of latent topics and characterizing each topic as a distribution of words [30]. This framework of the LDA model adheres to a 3-level Bayesian approach to effectively capture the generative process. However, before delving into the application of LDA or any other probabilistic topic modeling techniques, a critical step is to determine and define the number of topics often denoted as "k" [31]. This crucial decision significantly impacts the outcomes of the topic modeling process.

Qualitative Analysis

To strengthen the reliability of our findings obtained through the LDA model, we integrated a qualitative method focused on gaining a more profound insight into the identified themes. In particular, we followed the established 6-step thematic analysis framework outlined by Braun and Clarke [32] and successfully used by Xue et al [33]. This framework includes the following steps: (1) familiarizing ourselves with the keyword data and reviewing the most representative tweets for each topic, (2) generating initial codes to summarize key themes, (3) searching for thematic patterns by grouping similar topics, (4) reviewing and refining these potential themes to ensure coherence and consistency, (5) defining and naming themes based on their overall significance and contribution to the research question, and (6) reporting and documenting the final themes. This process was iterative and reflexive, involving multiple rounds of discussion and reassessment. Two researchers with extensive experience in social media analysis and public health independently reviewed and documented the initial codes. These codes were then examined by 2 additional researchers to refine the themes, ensuring that they accurately captured the essence of the topics.

Ethical Considerations

This study analyzed publicly available data collected from Twitter. The dataset consisted of tweet IDs, and no personally identifiable information was included in the analysis. All tweet texts were retrieved in compliance with Twitter's terms of service. Ethics approval was not sought, as the study used publicly accessible data, ensuring that no identifiable personal information was involved. To maintain the highest ethical standards, all results are presented in aggregate, guaranteeing the anonymity and privacy of individuals represented in the dataset.

Results

Descriptive Results

A total of 637,718 tweets were included in the final dataset after processing raw data. The analysis focused on identifying the most frequently tweeted bigrams (pairs of words) related to COVID-19. Bigrams are 2 consecutive words, regardless of their grammar structure or semantic meaning. They may not be self-explanatory, as in the case of the bigram "social distancing," which does not convey the meaning of either word on its own. Such an approach was adopted by Xue et al [4], and it was proved that bigrams can be a useful way to identify the most prominent topics and themes in Twitter conversations. The identified bigrams included pairs of words such as "virus corona," "stay home," "home order," "travel curfew," "new coronavirus," "spread virus," "home quarantine," "health quarantine," "coronavirus pandemic," "new infected," and "new case." Among the popular unigrams were words such as "coronavirus," "virus," "home," "new," "health," "world," "visit," "pandemic," "stay," "case," "quarantine," and "curfew." Most common unigrams and bigrams related to COVID-19, and pertinent details are listed in Table 2 (original Arabic tweets are provided in Multimedia Appendix 1).

Table 2. Top 50 unigrams and bigrams and their distributions.

	Values (%)
Top 50 unigrams	
Coronavirus	6.558451
Virus	2.350919
Home	0.921041
New	0.857981
Health	0.614924
Kuwait	0.576566
Condition	0.551307
Saudi Arabia	0.503562
World	0.491143
Country	0.487031
Visit	0.392251
Pandemic	0.391468
Curfew	0.359459
Stay	0.359077
Country	0.352204
Spread	0.34872
Infected	0.340486
Quarantine	0.339662
Case	0.335292
Disease	0.331376
Infected	0.328934
Urgent	0.314949
Egypt	0.313753
Virus	0.288958
People	0.272675
Minister	0.263771
People	0.257506
Health	0.244108
China	0.243201
Good	0.241965
Travel	0.241181
Citizen	0.239945
COVID	0.238966
King	0.238255
New	0.220993
Procedure	0.213274
Lebanon	0.211883
Wanted	0.209183
Confrontation	0.205782
Education	0.205174
In	0.198331

	Values (%)
Infection	0.193302
Thanks	0.187623
Announced	0.186263
Prevention	0.185222
Nation	0.184861
Iran	0.180255
House	0.178111
Italy	0.174504
In house	0.172979
Top 50 bigrams	
Virus, coronavirus	2.029932
Coronavirus, new	0.526419
Stay home	0.325347
Coronavirus, coronavirus	0.302665
Visit, health	0.263658
Virus, coronavirus	0.19593
Coronavirus, Kuwait	0.194992
Coronavirus, new	0.192446
Curfew, travel	0.18009
Spread, virus	0.155542
Coronavirus, virus	0.146133
Quarantine, home	0.138868
Quarantine, health	0.123512
New, virus	0.122492
Coronavirus, Lebanon	0.108992
Pandemic, coronavirus	0.108868
Home, coronavirus	0.107683
Coronavirus, Saudi Arabia	0.105704
Coronavirus, Egypt	0.103818
Infected, virus	0.102376
New, case	0.09342
Coronavirus, COVID	0.091503
Kuwait, coronavirus	0.089236
New, coronavirus	0.088587
Health, global	0.08464
Stay, home	0.083898
Minister, health	0.083743
Crisis, coronavirus	0.083589
Coronavirus, stay	0.076416
Organizer, health	0.073128
Confrontation, coronavirus	0.068563
Condition, in	0.06845
Saudi Arabia, coronavirus	0.064812

	Values (%)
Coronavirus, wanted	0.061967
Coronavirus, urgent	0.060535
Recording, case	0.055537
Confrontation, virus	0.054918
Spread, virus	0.053424
Spread, coronavirus	0.053187
Coronavirus, curfew	0.050755
Curfew, curfew	0.04958
Procedure, precautionary	0.049426
United, State	0.048818
Staying, home	0.048519
Disease, coronavirus	0.047993
Infected, coronavirus	0.047849
Citizen, resident	0.047684
Servant, holy mosque	0.04552
Prevention, travel	0.045458
Coronavirus, visit	0.044582

COVID-19–Related Topics

In our study, we used the LDA technique to identify and categorize frequently co-occurring words associated with COVID-19. The LDA algorithm allowed us to manually determine the number of topics we wanted to generate. In this study, we used 2 widely recognized metrics, CaoJuan2009 and Deveaud2014, available through the R package (R Foundation for Statistical Computing), to determine the optimal number of topics for our dataset. These metrics provided a robust framework for evaluating the coherence and distinctiveness of the topics, ensuring that the final model best captured the underlying structure of the data. The CaoJuan2009 measure is minimized when the number of topics aligns with the data's intrinsic structure, while the Deveaud2014 measure is maximized to indicate topic coherence and separation. These metrics were used to assess and validate the number of topics to ensure they reflect the data's diversity and relevance. By leveraging these 2 complementary metrics, we ensured that the selected number of topics provided meaningful insights and reduced the risk of overfitting. The number of topics was determined when these metrics stabilized, indicating a consistent result.

Upon evaluating the metrics, it was found that the CaoJuan2009 score converged at its minimum value with 16 topics, while the Deveaud2014 score peaked at its maximum value with the same number of topics. On the basis of this, we concluded that the optimal number of topics, denoted as "k," is 16, as shown in Figure 2.

In addition, we calculated the topic distance and visualized the intertopic relationships using a 2D plane [34]. Each circle in the plot represents a distinct topic, ranging from topic 1 to topic k. The positioning of these circles reflects the calculated distances between topics, offering a visual representation of their relationships.

It is also worth noting that cross-validation is less commonly applied in topic modeling for several reasons. These include computational challenges associated with applying cross-validation to unsupervised models, the interpretive nature of topic models, and the emphasis on qualitative coherence over predictive performance. Most studies on LDA and related techniques do not apply cross-validation, as the focus of topic modeling is on the interpretability and coherence of the topics rather than on predictive performance. Instead, topic models are typically evaluated using internal coherence and stability measures, such as the CaoJuan2009 and Deveaud2014 metrics, which prioritize the coherence of the topics and the consistency of the results across multiple runs. This approach is consistent with what is found in most related work on LDA. For example, Blei et al [30] introduced LDA and highlighted that the evaluation of topic models is traditionally done using measures such as coherence scores.

In Table 3 (original Arabic tweets are provided in Multimedia Appendix 2), we present the findings of the 16 LDA topics, revealing the most frequently occurring words within each topic along with the percentage of tweets falling under each respective topic. Among all 16 topics, topic 5 stands out with the highest percentage (9.98%) of tweets associated with it. In topic 5, we observed a significant co-occurrence of specific words, including "coronavirus," "increase," "health," "new," "infected," "death," "recovery," and "case." This combination of words indicates an escalation in the number of COVID-19 infections, leading to unfortunate fatalities and the emergence of new cases. Moreover, the presence of the term "recovery" implies that some individuals who were previously infected are now undergoing healing and improvement. Furthermore, we calculated the topic distance and illustrated the intertopic distance [35] in a 2D plane, as depicted in Figure 3. Each circle on the plot corresponds to a topic, ranging from topic 1 to topic 16 in this study. The positions of these circles were determined based on the calculated distances between the topics. Notably, in the visualization, the circles were not overlapping, which served as a validation of the 16 topics.

Figure 2. Metrics for estimating the optimal number of topics, ranging from 2 to 25 topics.

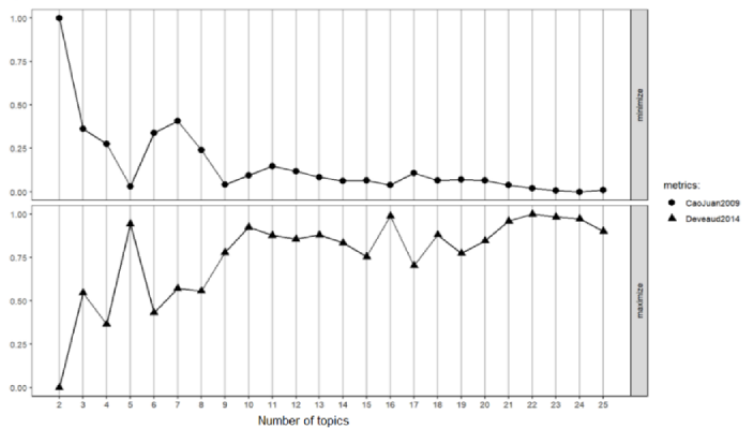
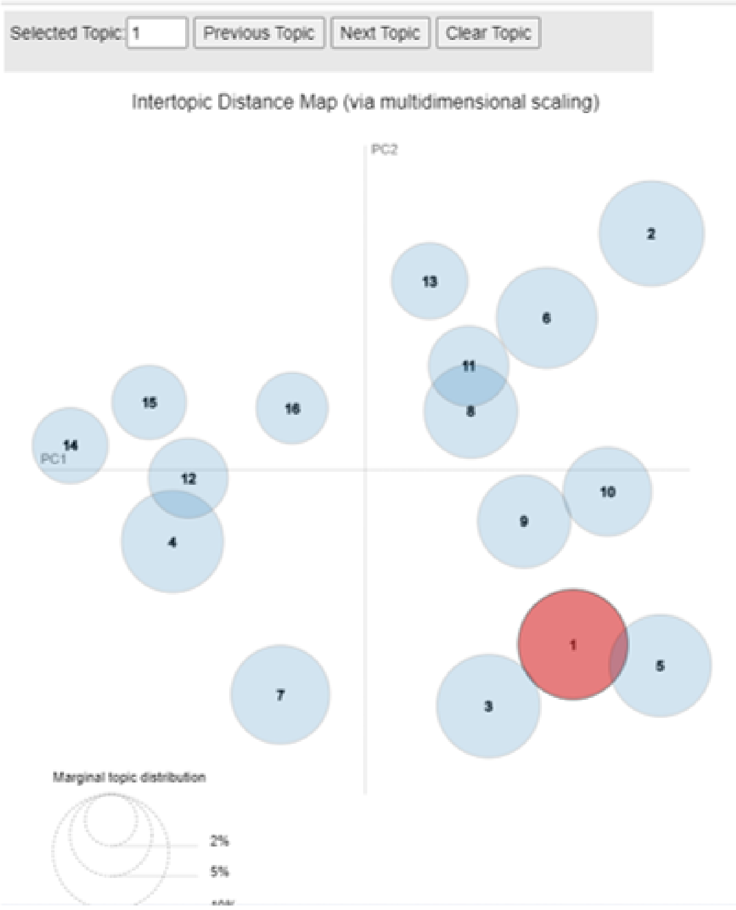


Table 3. Topic, words, and percentage of tweets.

Topic	Words	Values (%)
0	country, corona, Kuwait, praise, protection, gratitude to god, blessing, people, protect, people or nation, state, goodness, world, Saudi Arabia, Muslim, illness, thanks, pandemic, virus, Egypt	6.31
1	corona, affliction, pandemic, goodness, virus, Muslim, mercy, supplication/prayer, new, mind, world, lift or remove, great, illness, heart, raise, evil, people, mercy, Earth	8.5
2	corona, hand, virus, mask, washing, people, new, water, sanitizer, way, discount, knowledge, world, wear, person, soap, usage, glove, mask, beautiful	4.69
3	corona, virus, illness, Iran, medical, infected, hospital, doctor, treatment, Iraq, examination or test, health, person, device, hospital, Bahrain, infected, transmission, Italy, system	7.28
4	corona, virus, Kuwait, Egypt, new, emerging, COVID, health, visited, suspension, Saudi Arabia, corona, statement, Kuwaiti, confrontation, Emirate, study, crew, state, prevention	6.09
5	corona, virus, condition, new, case, infected, health, infected, died, infection, urgent, recording, death, announced, visited, increase, recovery, recorded, total, rose	9.98
6	corona, virus, education, visited, minister, confrontation, support, private, health, student, bank, spread, sector, state, responsible, crisis, communication, community, request, home	8.16
7	corona, China, state, virus, world, pandemic, union, Italy, hate, Europe, league, America, new, spread, presented, condition, European, action, player, east	4.13
8	corona, virus, house, scene, protect, country, Algeria, Egypt, died, rest, detail, video, lead, people, young man, Morocco, new, image, wanted, film	3.83
9	house, corona, stay, curfew, quarantine, wandering, home based, virus, new, Saudi Arabia, home, Kuwait, responsible, effectiveness, roaming, health, wanted, complete, goodness, Zoom	7.18
10	corona, virus, world, Trump, Oman, new, vaccine, president, faced, America, China, treatment, wanted, news, Chinese, partnership, vaccine, COVID, American, Palestine	4.68
11	corona, virus, spread, health, state, pandemic, prevention, illness, enemy, awareness, danger, way, gathering, must, home, country, avoidance, citizen, world, prevention	5.86
12	corona, virus, spread, health, state, pandemic, prevention, illness, enemy, awareness, danger, way, gathering, must, home, country, avoidance, citizen, world, prevention	5.86
13	corona, China, state, virus, world, pandemic, union, Italy, hate, Europe, league, America, new, spread, presented, condition, European, action, player, east	4.13
14	corona, Saudi Arabia, thanks, Kuwait, king, health, protection, country, homeland, virus, citizen, people or nation, visited, effort, state, sanctuary, praise, Salman, pandemic, protect	7.35
15	corona, Lebanon, people, one, age, went out, quarantine, meant, topic, condition, house, what, virus, safety, health, Egypt, people or nation, world	4.98

Figure 3. Latent Dirichlet allocation—intertopic distance.



COVID-19–Related Themes

Through the process of thematic analysis, we were able to categorize the identified topics, bigrams, and representative tweet samples into distinct themes, as shown in Table 4 (original Arabic tweets are provided in Multimedia Appendix 3).

The sample tweets provided in Table 4 are excerpts taken from the original tweets. These 16 topics have been categorized into eight overarching themes, summarized below.

1. Preventive measures and safety (“public health measures”): this theme focuses on various measures to prevent the spread of COVID-19, such as wearing masks, washing hands, using sanitizers, and practicing social distancing.
2. Medical and health care aspects: this theme encompasses topics related to the medical and health care aspects of COVID-19, including hospitals, doctors, treatments, testing, and recovery.
3. Government and social measures: this theme covers government actions, social measures, and policies implemented to address the COVID-19 pandemic, including lockdowns, travel restrictions, home orders, suspending schools, avoiding gatherings, closing shops, staying at home, and support measures.
4. Impact and numbers: this theme involves discussions about the impact of COVID-19, including the number of cases, deaths, recoveries, and updates on the situation.
5. Vaccine development and research: this theme revolves around vaccine development, clinical trials, and scientific research related to finding a solution to COVID-19.
6. COVID-19 and religious practices: this topic discusses how COVID-19 has impacted religious practices and gatherings. It mentions places of worship ( ) and the importance of adhering to prayers ( ) and religious guidelines ( ) during the COVID-19 pandemic, especially during occasions such as Ramadan ( ). The theme also includes expressions of gratitude and good wishes for nations and people (, ,).
7. Global impact of COVID-19 on sports and countries: this topic discusses the spread of COVID-19 in different countries, including China, Italy, and the United States, and its impact on various aspects, such as sports events and leagues in Europe and the Middle East. It also mentions the virus as a global pandemic and its effects on athletes and players ( ) as well as its presence in different regions around the world.
8. COVID-19 and national efforts: this theme focuses on the efforts of different nations, including Saudi Arabia and Kuwait, in combating COVID-19. It mentions leaders (, ,) and their efforts to protect the health and well-being of their citizens (,). The theme includes expressions of gratitude for the nation’s efforts in managing the COVID-19 pandemic ( ) and highlights the importance of public health ( ). Textbox 1 provides a comprehensive list of topics, thoughtfully translated into English for better clarity and accessibility.

Table 4. Themes based on topic classification, bigrams, and sample tweets.

Theme and topic	Bigrams	Sample tweets
Preventive measures and safety		
Face mask	Wear mask	A note for your safety from the new coronavirus infection: Avoid social gatherings with more than 1 person. Avoid crowded areas or places where you might interact with individuals who are sick. Avoid handshakes as they are among the primary causes of virus transmission. Wear a mask whenever possible.
Hands	Wash hands, use sanitizers	Avoid gatherings, closed spaces, and crowded areas, along with regularly washing your hands with water and soap or sanitizing them with alcohol-based disinfectants. By God's will, you will be protected from contracting the new coronavirus.
Social distancing	Social distancing	Social distancing means staying away from gatherings and crowded places. If you must leave your home, maintain a distance of at least 2 meters from the people around you. Source: Cleveland Clinic, COVID-19.
Medical and health care aspects		
Health authorities	Precautionary measures, followed the instructions	Home quarantine protects against the risk of a person spreading the coronavirus without showing symptoms, making them a potential source of transmission to various groups. Preventive measures against COVID-19 ease the burden on health care providers, enabling them to fulfill their roles in treating other illnesses and performing preventive tasks, including COVID-19 detection. Voice of the physician.
Recovery	Case recovery	Breaking: The Ministry of Health announces the recovery of the first coronavirus case in the kingdom. This concerns the young man who returned from Italy and was previously announced as the first imported case of the virus in Morocco. COVID-19, Morocco, Recovery, Ministry of Health.
Treatment	Treating the infected	The Minister of Health announces the initiation of treating patients with COVID-19 with the chloroquine vaccine.
Treatment	New drug	The *Washington Post* reports that Chinese experts and physicians have successfully fought COVID-19 using chloroquine, a drug primarily used to treat malaria, and Kaletra, an HIV medication that combines lopinavir and ritonavir. Emirati physician Omar Al Hammadi shares the success of this trial.
Hospital	Field hospital	Starting Sunday, a physician will accompany every ambulance, and a field hospital will be established inside the trade unions complex. Dr Ali Al-Abous, President of the Jordanian Medical Association, comments on the nationwide curfew in Jordan due to the COVID-19 pandemic.
Government and social measures		
Lockdowns and suspending	Closing shops, suspending schools	Precautionary measures in Kuwait against COVID-19: suspension of studies and work, cancellation of weddings, closure of mosques, closure of malls, closure of salons, partial curfew, extension of the suspension of studies, regulation of work in central markets, closure of shops, postponement of installments.
Travel restrictions	Travel ban	Saudi Arabia: Saudi Arabia suspended studies, banned cafes and shisha, prohibited sports gatherings and cinemas, halted entertainment activities, stopped Umrah and travel, and conducted intensive testing to search for patients. All for your benefit—help your government overcome these circumstances with minimal losses.
Home orders	Stay home	Stay home and protect your family from coronavirus. Prevention guidelines. Stay home.
Curfew	Curfew	Breaking: Al Jazeera correspondent reports the sounding of alarm sirens across Jordan as the nationwide curfew begins to combat the spread of COVID-19.
Remote	Remote work	It is everyone's duty to follow the precautionary measures taken by our government, may God protect them, to prevent the spread of COVID-19. At our facility, we have informed the success team to work remotely from their homes until further notice.
Impact and numbers		
New cases	Confirmed cases, increase in cases	The Kuwaiti Ministry of Health has reported new cases of the novel coronavirus, and the total number of patients that have exited quarantine is 20.
Deaths	Coronavirus deaths	A new death has been recorded in Jordan due to COVID-19, bringing the total number of deaths to 5.
Vaccine development and research		
Religious guidelines	Prayer, supplication	Breaking: The Senior Scholars Authority calls on everyone to adhere to the instructions, guidelines, and regulations, to fear God, and to resort to prayer and supplication. COVID-19, Saudi Arabia.

Theme and topic	Bigrams	Sample tweets
Umrah	Suspension of Umrah	It was discovered during the COVID-19 crisis that preserving life is one of the most important objectives of Sharia, and everything is subordinated to it. The suspension of Umrah and prayer in mosques reflects the greatness of Islam and the depth of Sharia's objectives.
Global impact of COVID-19 on sports and countries		
Postponement of matches	Postponement of matches	The Union of European Football Association has decided to postpone all matches scheduled for next week. Sports, COVID-19.
Italy	The situation in Italy	Terrifying numbers in Italy and Iran; a video shows the spread of the coronavirus outside China until March.
COVID-19 and national efforts		
King Salman	Royal support	King Salman bin Abdulaziz and Crown Prince Mohammed bin Salman. The Saudi Arabian Monetary Authority announces support for the private sector with 1 billion Saudi riyals to face the expected financial and economic impacts of the coronavirus.
Thanks	Government gratitude	We thank God for the blessing of Islam and the blessing of Salman. Every Saudi has the right to be proud and boast about Saudi Arabia. May God protect its government and people from all harm. Saudi Arabia. COVID-19. Stay at home.

Textbox 1. Topic and words (English translations) used in the study.

- Topic 0: country, corona, Kuwait, Hamad, preserve, Alhamdulillah, blessing, people, preserve, people, state, good, world, Saudi Arabia, Muslim, disease, thanks, epidemic, virus, and Egypt
- Topic 1: corona, calamity, epidemic, good, virus, Muslim, mercy, prayer, new, by, world, lift, great, disease, heart, raise, evil, people, mercy, and land
- Topic 2: corona, hand, virus, mask, wash, people, new, water, sanitizer, road, discount, know, world, wear, person, soap, use, gloves, mask, and beautiful
- Topic 3: corona, virus, disease, Iran, medical, infected, hospital, doctor, treatment, Iraq, test, health, person, device, hospital, Bahrain, infected, transfer, Italy, and system
- Topic 4: corona, virus, Kuwait, Egypt, new, novel, Covid, health, visit, suspension, Saudi Arabia, core, statement, Kuwaiti, confront, Emirate, study, cure, country, and protection
- Topic 5: corona, virus, condition, new, condition, infected, health, infected, and, infection, urgent, registration, death, announce, visit, rise, recovery, register, total, and rise
- Topic 6: corona, virus, education, visit, minister, confront, support, special, health, student, bank, publish, sector, state, official, crisis, contact, community, request, house
- Topic 7: corona, China, country, virus, world, epidemic, union, Italy, football, Europe, league, America, new, spread, foot, player, and east
- Topic 8: corona, virus, home, scene, protect, country, Algeria, Egypt, die, wind, detail, video, top, people, young, Morocco, new, picture, wanted, and film
- Topic 9: home, corona, stay, ban, quarantine, circulation, homely, virus, new, Saudi Arabia, home, Kuwait, official, activity, circulation, health, wanted, complete, good, and old
- Topic 10: corona, virus, world, Trump, Oman, new, vaccine, president, confront, America, China, treatment, wanted, news, Chinese, company, vaccine, coveted, American, and Palestine
- Topic 11: corona, virus, spread, health, state, epidemic, protection, disease, enemy, awareness, threat, road, gathering, mandatory, country, avoid, citizen, world, and protection
- Topic 12: corona, mosque, people, gathering, prayer, congregation, Lebanon, condition, Ramadan, virus, prayer, I mean, talk, cover, world, Egypt, great, good, people, and peace
- Topic 13: corona, virus, procedure, spread, prevention, decision, sanitization, closure, local, logic, visit, urgent, Saudi Arabia, new, governor, application, shop, Riyadh, precautionary, and system
- Topic 14: corona, Saudi Arabia, thanks, Kuwait, king, health, preserve, country, homeland, virus, citizen, people, visit, effort, state, crisis, blessing, Salman, epidemic, and preserve
- Topic 15: corona, Lebanon, people, and, age, came out, quarantine, from me, subject, condition, house, and, mean, virus, peace, health, Egypt, people, world, and damn

Sentiment Analysis

We conducted sentiment analysis for each of the 16 topics and presented the results in Figure 4 and Table 5. Figure 4 visualized 7 emotions: anger, disgust, joy, anticipation, sadness, surprise, and fear. Across all 16 topics, anger (represented by the red line) was the dominant emotion in 16 topics, followed by disgust (green line), joy (blue line), and anticipation (orange line). To delve deeper into the emotional aspects of the data, we provide a breakdown of the number of tweets associated with each emotion across different topics in Table 5. For example, in topic 5, a substantial number of tweets (n=17,848) expressed anger, reflecting a strong sentiment regarding the need for essential measures and precautions. This high prevalence of anger in topic 5 stands out in comparison to the other topics. It is worth noting that excessive anger, if left unmanaged, can lead to a range of medical problems. Managing emotions such as anger is crucial not only for mental well-being but also for overall physical health.

Figure 4. Sentiment analysis for each of the 16 latent topics.

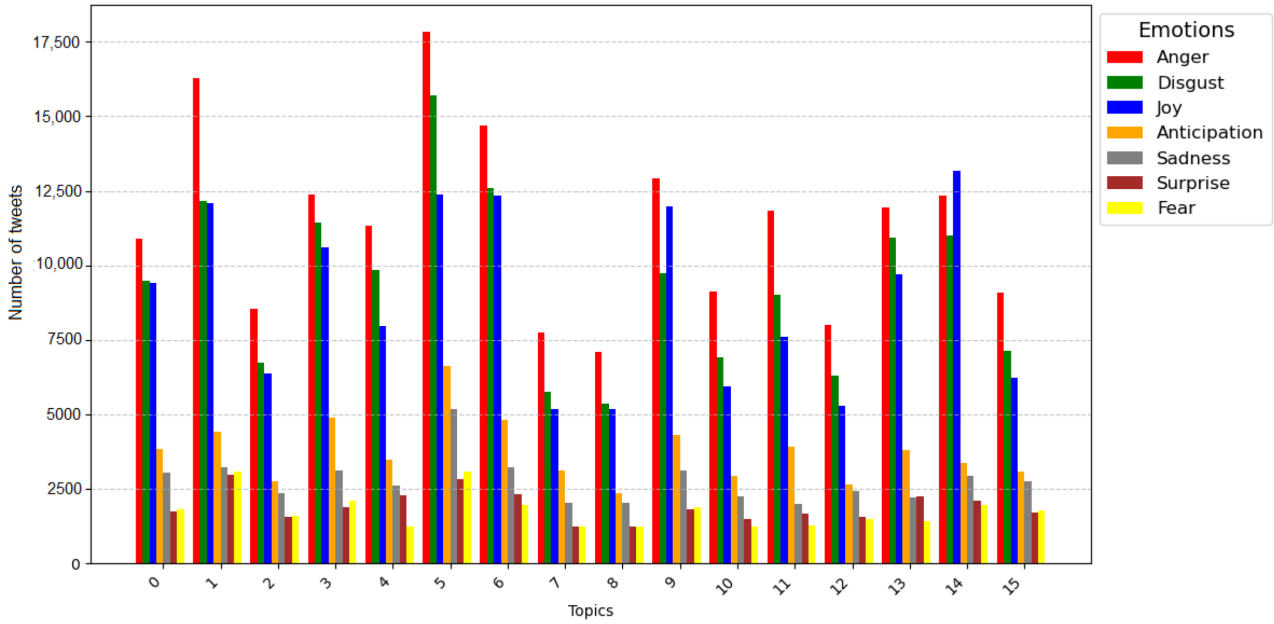


Table 5. The number of tweets for 7 emotions across 16 topics.

Topic	Anger	Anticipation	Disgust	Fear	Joy	Sadness	Surprise
0	10,896	3834	9475	1835	9428	3040	1757
1	16,295	4434	12,164	3079	12,083	3217	2965
2	8559	2772	6729	1585	6366	2346	1554
3	12,391	4894	11,424	2110	10,602	3113	1876
4	11,345	3493	9838	1246	7969	2624	2295
5	17,848	6630	15,688	3078	12,375	5192	2847
6	14,702	4820	12,601	1979	12,329	3239	2342
7	7757	3130	5752	1224	5189	2031	1230
8	7082	2367	5344	1222	5168	2035	1228
9	12,926	4321	9742	1896	11,989	3132	1802
10	9113	2948	6901	1225	5953	2245	1486
11	11,836	3915	9027	1288	7606	2015	1656
12	7988	2646	6286	1494	5282	2416	1547
13	11,942	3801	10,928	1437	9717	2231	2241
14	12,328	3357	11,007	1962	13,172	2953	2124
15	9097	3087	7116	1779	6218	2762	1716

Discussion

Principal Findings

This study delved into public discussion and emotional expressions related to COVID-19 using Arabic Twitter messages. Twitter users engaged in discussions encompassing 8 primary themes regarding COVID-19. Using topic modeling on the tweets proved valuable in uncovering insights into COVID-19–related topics and concerns. The outcomes highlighted several crucial observations.

This analysis concentrates on tweets from March 2020, a pivotal phase in the COVID-19 pandemic's unfolding narrative. During this period, the second stage of the COVID-19 pandemic emerged prominently, marked by a significant milestone as Arabic countries reported their initial cases of COVID-19. Subsequently, a cascade of vital health measures ensued, encompassing the enforcement of quarantine protocols, the temporary cessation of air travel, and the inevitable postponement or cancelation of various events. This time frame aligns logically with the peak frequency of tweets, as previously observed by Taneja et al [22] and Haouari et al [34].

Amidst the array of all 16 topics, a discernible pattern surfaced, characterized by the recurring presence of specific keywords such as "coronavirus," "increase," "health," "new," "infected," "death," "recovery," and "case." This linguistic cluster strongly implies a surge in COVID-19 infections, accompanied by lamentable loss of life and the emergence of new cases during the ongoing COVID-19 pandemic. It is imperative to emphasize that our chosen time frame aligns precisely with the onset of the COVID-19 pandemic's second phase, coinciding with heightened global concern. The substantial spike in COVID-19 cases in Italy during this period ignited a profound sense of alarm on a global scale. This surge in worldwide apprehension may have contributed to the observed increase in tweet frequency, corroborating findings from multiple studies [22,34].

Furthermore, substantial discussions revolving around the COVID-19 pandemic within diverse Arabic nations have drawn significant interest. These conversations are marked by a prevailing sense of indignation. Moreover, public sentiments concerning the spread of COVID-19 unveiled an underlying sense of anticipation toward prospective measures. These sentiments were accompanied by a mix of emotions, including anger and fear; a notable undercurrent of fear was predominant in discussions revolving around the COVID-19 crisis and the resulting fatalities. This trend aligns with global sentiments, as documented by Lwin et al [36], wherein public emotions underwent a noticeable shift from fear to anger throughout the COVID-19 pandemic, with traces of sadness and joy also emerging.

Noteworthy, the appearance of dialogues concerning COVID-19 and religious practices introduced a fresh subject not previously detected in prior research. This indicates a developing connection between COVID-19 and religious matters on the Twitter platform. This is particularly apparent due to the substantial influence of religious identity on attitudes and actions concerning the COVID-19 pandemic and vaccination efforts; the COVID-19 pandemic has significantly reshaped communal worship and gatherings as measures to curb the virus's transmission [37]. Furthermore, religious leaders have assumed a central role in championing COVID-19 vaccination campaigns, effectively addressing and mitigating vaccine hesitancy [38].

In-Depth Analysis of Findings

The application of topic modeling and sentiment analysis in this study provided several valuable insights into public sentiment and thematic discussions during the early stages of the COVID-19 pandemic in Arab regions. The findings largely align with anticipated outcomes, such as the focus on preventive measures and safety and medical and health care aspects, both of which were expected topics given the nature of the COVID-19 pandemic.

However, the emergence of discussions on COVID-19 and religious practices was a unique finding that adds depth to the understanding of public discourse in Arab communities. This theme highlights the intersection of the COVID-19 pandemic with cultural and religious practices, which had not been as thoroughly explored in previous research. It underscores the significant impact that COVID-19 had on religious identity, communal worship, and adherence to religious guidelines during pivotal periods such as Ramadan.

Another notable aspect was the attention given to the global impact of COVID-19 on sports and countries, reflecting the broad international concern and how global events, especially sports, were affected. This indicates that the COVID-19 pandemic's influence went beyond public health and extended into societal and cultural dimensions, impacting activities that are deeply integrated into daily life.

In addition, the sentiment analysis revealed a nuanced distribution of emotions, with a significant proportion of tweets expressing anger and disgust, as expected, given the uncertainty surrounding the COVID-19 pandemic. However, there was also a notable presence of positive emotions, such as hope and solidarity, particularly in tweets discussing community support and coping mechanisms. This suggests that, despite the overwhelming nature of the crisis, many users turned to social media not only to express negative emotions but also to share supportive messages and encourage others.

Overall, the identified themes and their respective discussions provide a comprehensive view of public sentiment, concerns, and priorities during the early COVID-19 pandemic period. These insights not only reflect the immediate response to the health crisis but also highlight the diverse and context-specific aspects that shaped public discourse. Such findings offer a foundation for more effective public health communication and intervention strategies, particularly in culturally sensitive contexts.

Strengths

This study provided valuable insights into the sentiments and concerns of Arabic-speaking Twitter users during the COVID-19 pandemic, underscoring the significance of social media as a means of understanding and addressing public health issues in the digital era. First, the analysis encompassed a substantial dataset of 1 million Arabic tweets, offering a comprehensive view of the sentiments and topics expressed by Twitter users in Arabic-speaking countries during a specific period of the COVID-19 pandemic. Besides, the study used a combination of machine learning techniques, including topic modeling and sentiment analysis, to uncover and categorize themes and emotions within the dataset, providing a holistic understanding of the data. By identifying and categorizing 16 conversation topics into 8 themes, the study offered a structured view of the discussions surrounding COVID-19 in the Arab region, making it easier to interpret and use the findings. Finally, the inclusion of emotion analysis adds depth to the study, revealing how Twitter users in the Arab world emotionally responded to various aspects of the COVID-19 pandemic.

Limitations

First, at the forefront of our approach, we meticulously aimed to unravel the complexities embedded within the COVID-19 pandemic's second phase. Our focus was sharp and exclusive, centered on harnessing tweets originating exclusively from March 2020. The motivation behind this specific time frame stemmed from our intention to subject translated tweets to a comprehensive sentiment analysis. This intricate process relied upon the Google API translation service, which, although effective, is accompanied by a substantial cost factor. The financial implication associated with translating the entirety of the datasets using this service was a noteworthy consideration that prompted us to make strategic choices in our analysis approach.

Second, it is crucial to recognize that Arabic is a linguistically intricate language characterized by a rich array of dialects and intricate cultural nuances. These unique linguistic qualities can present substantial challenges for automated sentiment analysis tools. While we attempted to apply automated sentiment analysis to Arabic tweets, we encountered difficulties in precisely capturing the subtleties of emotions. Automated tools often grappled with interpreting nuanced sentiments, such as sarcasm, irony, and contextual shifts in sentiment that frequently permeate social media conversations.

Third, a strategic decision was made to exclude non-Arabic tweets from our analyses. As a result, our findings were inherently confined to users who exclusively communicated in Arabic. It is essential to underscore that the fundamental objective of our research revolves around gaining insights into the opinions and reactions of Arabic countries in relation to COVID-19.

Furthermore, while our study leveraged social media data as a proxy for public sentiment, it is essential to recognize the inherent biases associated with using Twitter data. For instance, social media users may not be representative of the general population, as certain demographics might be underrepresented on platforms such as Twitter. A study by Padilla et al [39] has shown that social media content can be biased based on whether individuals are local residents or visitors and the types of activities they engage in throughout the day. Similarly, Gore et al [40] highlighted that the sentiment of tweets is often correlated with the geographical area in which they were composed, suggesting that local context and specific events may have a significant impact on sentiment analysis results. Frank et al [41] also found that emotional expressions, such as happiness, vary significantly by location, further reinforcing the influence of geographic factors on sentiment.

In addition, it is plausible that individual personality traits or political affiliations, as suggested by Auer and Elena [42], could influence whether a user expresses positive or negative sentiments. This raises an open question about the extent to which sentiment reflects variance in psychological traits versus the situational context in which those traits are expressed. These factors could contribute to biases in our dataset and should be considered as potential sources of influence on the study's outcomes.

Future Work

Regarding future studies focusing on COVID-19, first, there arises a noteworthy avenue for exploration comparing the sentiments and opinions of Arabic-speaking populations with those of individuals expressing themselves in other languages. A comprehensive approach might encompass languages such as English, Italian, French, German, and Spanish. Such comparative analyses have the potential to yield valuable insights into the cross-linguistic dynamics of perceptions and responses to the COVID-19 pandemic.

Second, another promising avenue for future research involves conducting a comparative analysis between sentiment analysis using human-labeled data and automated tools specifically tailored for Arabic languages. This comparative study should aim to ascertain the feasibility of leveraging these automated tools as an alternative to translation APIs. By meticulously comparing the results obtained from human-labeled sentiment analysis and those generated by automated tools, researchers can gauge the efficacy, accuracy, and reliability of automated sentiment analysis for Arabic tweets. The outcomes of this research hold the potential for far-reaching implications, potentially presenting a cost-effective and streamlined avenue for sentiment analysis that eliminates the reliance on costly translation APIs.

By providing an accurate and efficient mechanism for measuring sentiments in Arabic tweets, researchers and mental health professionals could identify patterns of emotional distress or psychological well-being. This could be especially pertinent during times of crises, enabling timely interventions and support for individuals experiencing heightened emotional responses. Importantly, the ability to effectively harness sentiment analysis for understanding emotional states has the potential to empower the broader field of mental health research and intervention as well as enhance our understanding of collective emotional dynamics within Arabic-speaking communities.

Third, there is an imminent need for research to unravel the stem of fabricated tweets that emerge during a pandemic. Given that Twitter users experience a heightened sense of fear, which might be exacerbated by the proliferation of misinformation, it becomes a critical endeavor to investigate the prevalence and impact of false tweets. Subsequent studies could significantly benefit from spotlighting the issue of misinformation, with a specific focus on understanding how government officials and international organizations can effectively manage the dissemination of deceptive messages targeting the public. By comprehensively addressing the challenges posed by misleading content, we can enhance our collective understanding of navigating information dissemination during such critical periods.

Conclusions

This study delves deep into the intricate web of topics and emotions found in Arabic tweets about COVID-19. It highlights how platforms such as Twitter, especially during times of global change, are crucial for capturing the diverse feelings and concerns of Arabic speakers. Through a mix of topic modeling and sentiment analysis, we revealed the basic human emotions in user responses to COVID-19 tweets from March 2020.

We used 2 methods together: topic modeling (specifically LDA) and sentiment analysis tools. These helped us uncover the main themes and feelings within the tweets. Anger was the prominent emotion tied to COVID-19 topics, accompanied by other emotions. Joy was linked to vaccine and education discussions, while authority and politics stirred up anger. Sadness emerged from topics about cases, deaths, and the impacts on families and mental health.

This study connects social media, emotions, and the global scene. It sheds light on the emotional layers of digital conversations, offering insights into COVID-19–related tweets. These findings guide better communication strategies and compassionate responses, strengthening our collective resilience in the face of challenges.

Moreover, the results and workflow of this study present actionable insights for the medical and public health communities. By integrating our findings into official government documentation or public health research, authorities can tailor their communication strategies based on public concerns and emotions. This, in turn, helps in shaping more effective educational campaigns and policy interventions. Our methodology also serves as a robust tool for continuous monitoring of public sentiment in real time, allowing policy makers to stay informed and adapt their strategies accordingly. This approach ensures that responses are not only timely but also grounded in the actual sentiments and needs of the population.

Acknowledgments

The authors would like to acknowledge the deanship of research at the Jordan University of Science and Technology for supporting this research (grant 653/2023).

Data Availability

The data sets generated and analyzed during this study are available in the GitHub repository [43].

Conflicts of Interest

None declared.

Multimedia Appendix 1
Original Arabic versions of tweets shown in Table 2.
[DOCX File , 20 KB - infodemiology_v5i1e53434_app1.docx]

Multimedia Appendix 2
Original Arabic versions of tweets shown in Table 3.
[DOCX File , 18 KB - infodemiology_v5i1e53434_app2.docx]

Multimedia Appendix 3
Original Arabic versions of tweets shown in Table 4.
[DOCX File , 25 KB - infodemiology_v5i1e53434_app3.docx]

References

1. Hasan MN, Islam MA, Sangkham S, Werkneh AA, Hossen F, Haque MA, et al. Insight into vaccination and meteorological factors on daily COVID-19 cases and mortality in Bangladesh. Groundw Sustain Dev 2023 May;21:100932 [FREE Full text] [doi: 10.1016/j.gsd.2023.100932] [Medline: 36945723]
2. Rosenberg H, Syed S, Rezaie S. The Twitter pandemic: the critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. CJEM 2020 Jul 06;22(4):418-421 [FREE Full text] [doi: 10.1017/cem.2020.361] [Medline: 32248871]
3. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]

4. Xue J, Chen J, Gelles R. Using data mining techniques to examine domestic violence topics on Twitter. Violence Gend 2019 Jun;6(2):105-114. [doi: 10.1089/vio.2017.0066]
5. Signorini A, Segre AM, Polgreen PM. The use of Twitter to track levels of disease activity and public concern in the U.S. during the influenza A H1N1 pandemic. PLoS One 2011 May 04;6(5):e19467 [FREE Full text] [doi: 10.1371/journal.pone.0019467] [Medline: 21573238]
6. Rufai S, Bunce C. World leaders' usage of Twitter in response to the COVID-19 pandemic: a content analysis. J Public Health (Oxf) 2020 Aug 18;42(3):510-516 [FREE Full text] [doi: 10.1093/pubmed/fdaa049] [Medline: 32309854]
7. Xue J, Chen J, Chen C, Zheng C, Li S, Zhu T. Public discourse and sentiment during the COVID 19 pandemic: using latent Dirichlet allocation for topic modeling on Twitter. PLoS One 2020 Sep 25;15(9):e0239441 [FREE Full text] [doi: 10.1371/journal.pone.0239441] [Medline: 32976519]
8. Alharbi NH, Alkhateeb JH. Sentiment analysis of Arabic tweets related to COVID-19 using deep neural network. In: Proceedings of the 2021 International Congress of Advanced Technology and Engineering. 2021 Presented at: ICOTEN '21; July 4-5, 2021; Taiz, Yemen p. 1-11. [doi: 10.1109/icoten52080.2021.9493467]
9. Alhumoud S, Al Wazrah A, Alhussain L, Alrushud L, Aldosari A, Altammami RM, et al. ASAVACT: Arabic sentiment analysis for vaccine-related COVID-19 tweets using deep learning. PeerJ Comput Sci 2023;9:e1507 [FREE Full text] [doi: 10.7717/peerj-cs.1507] [Medline: 38077546]
10. Ali M. Arabic sentiment analysis about online learning to mitigate COVID-19. Int J Intell Syst 2021;30(1):524-540. [doi: 10.1515/jisys-2020-0115]
11. Alqarni A, Rahman A. Arabic tweets-based sentiment analysis to investigate the impact of COVID-19 in KSA: a deep learning approach. Big Data Cogn Comput 2023 Jan 13;7(1):16 [FREE Full text] [doi: 10.3390/bdcc7010016]
12. AlAgha I. Topic modeling and sentiment analysis of twitter discussions on COVID-19 from spatial and temporal perspectives. J Inf Sci Theory Pract 2021;9(1):35-53 [FREE Full text] [doi: 10.1633/JISTAP.2021.9.1.3]
13. Habbat N, Anoun H, Hassouni L. Sentiment analysis and topic modeling on Arabic Twitter data during COVID-19 pandemic. Indones J Innov Appl Sci 2022;2(1):60-67 [FREE Full text] [doi: 10.47540/ijias.v2i1.432]
14. Tamang S, Humbert-Droz M, Gianfrancesco M, Izadi Z, Schmajuk G, Yazdany J. Practical considerations for developing clinical natural language processing systems for population health management and measurement. JMIR Med Inform 2023 Jan 03;11:e37805 [FREE Full text] [doi: 10.2196/37805] [Medline: 36595345]
15. Mahmoud E, Emilien A, Maxime G, Gilles D. The role of text analytics in healthcare: a review of recent developments and applications. In: Proceedings of the 14th International Joint Conference on Biomedical Engineering Systems and Technologies. 2021 Presented at: BIOSTEC '21; February 11-13, 2021; Virtual Event p. 825-832. [doi: 10.5220/0010414508250832]
16. Amores JJ, Blanco-Herrero D, Arcila-Calderón C. The conversation around COVID-19 on Twitter—sentiment analysis and topic modelling to analyse tweets published in English during the first wave of the pandemic. Journal Media 2023 Mar 30;4(2):467-484. [doi: 10.3390/journalmedia4020030]
17. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
18. Aslam F, Awan TM, Syed JH, Kashif A, Parveen M. Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak. Humanit Soc Sci Commun 2020 Jul 08;7(1):13. [doi: 10.1057/S41599-020-0523-3]
19. Kwok SW, 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 May 19;23(5):e26953 [FREE Full text] [doi: 10.2196/26953] [Medline: 33886492]
20. Park S, Suh YK. A comprehensive analysis of COVID-19 vaccine discourse by vaccine brand on twitter in Korea: topic and sentiment analysis. J Med Internet Res 2023 Jan 31;25:e42623 [FREE Full text] [doi: 10.2196/42623] [Medline: 36603153]
21. Wang Y, Chen Y. Characterizing discourses about COVID-19 vaccines on Twitter: a topic modeling and sentiment analysis approach. J Commun Healthc 2023 Mar 24;16(1):103-112. [doi: 10.1080/17538068.2022.2054196] [Medline: 36919802]
22. Taneja SL, Passi M, Bhattacharya S, Schueler SA, Gurram S, Koh C. Social media and research publication activity during early stages of the COVID-19 pandemic: longitudinal trend analysis. J Med Internet Res 2021 Jun 17;23(6):e26956 [FREE Full text] [doi: 10.2196/26956] [Medline: 33974550]
23. Qazi U, Imran M, Ofli F. GeoCoV19: a dataset of hundreds of millions of multilingual COVID-19 tweets with location information. SIGSPATIAL Special 2020 Jul 08;12(1):6-15. [doi: 10.1145/3404820.3404823]
24. A command line tool (and Python library) for archiving Twitter JSON. GitHub. URL: https://github.com/DocNow/twarc [accessed 2024-04-29]
25. Alshanik F, Apon A, Herzog A, Safro I, Sybrandt J. Accelerating text mining using domain-specific stop word lists. In: Proceedings of the 2020 IEEE International Conference on Big Data. 2020 Presented at: Big Data '20; December 10-13, 2020; Atlanta, GA p. 2639-2648 URL: https://ieeexplore.ieee.org/document/9378226 [doi: 10.1109/bigdata50022.2020.9378226]

26. Abdelali A, Darwish K, Durrani N, Mubarak H. Farasa: a fast and furious segmenter for arabic. In: Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations. 2016 Presented at: NAACL '16; June 12-17, 2016; San Diego, CA p. 11-16 URL: https://aclanthology.org/N16-3003.pdf [doi: 10.18653/v1/n16-3003]
27. Beigi G, Hu X, Maciejewski R, Liu H. An overview of sentiment analysis in social media and its applications in disaster relief. In: Pedrycz W, Chen SM, editors. Sentiment Analysis and Ontology Engineering: An Environment of Computational Intelligence. Cham, Switzerland: Springer; 2016:313-340.
28. Barbieri F, Camacho-Collados J, Neves L, Espinosa-Anke L. TWEETEVAL: unified benchmark and comparative evaluation for tweet classification. In: Proceedings of the 2020 Conference on Association for Computational Linguistics. 2020 Presented at: EMNLP '20; November 16-20, 2020; Virtual Event p. 1644-1650 URL: https://aclanthology.org/2020.findings-emnlp.148.pdf [doi: 10.18653/v1/2020.findings-emnlp.148]
29. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. Huggingface's transformers: state-of-the-art natural language processing. arXiv Preprint posted online October 9, 2019 [FREE Full text]
30. Blei DB, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003;3:4-5 [FREE Full text] [doi: 10.1017/9781009218245.012]
31. Blei D, Carin L, Dunson D. Probabilistic topic models: a focus on graphical model design and applications to document and image analysis. IEEE Signal Process Mag 2010 Nov 01;27(6):55-65 [FREE Full text] [doi: 10.1109/MSP.2010.938079] [Medline: 25104898]
32. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
33. Xue J, Chen J, Hu R, Chen C, Zheng C, Su Y, et al. Twitter discussions and emotions about the COVID-19 pandemic: machine learning approach. J Med Internet Res 2020 Nov 25;22(11):e20550 [FREE Full text] [doi: 10.2196/20550] [Medline: 33119535]
34. Haouari F, Hasanain M, Suwaileh R, Elsayed T. ArCOV-19: the first Arabic COVID-19 twitter dataset with propagation networks. In: Proceedings of the 6th Conference on Arabic Natural Language Processing Workshop. 2021 Presented at: WANLP '21; April 19, 2021; Kyiv, Ukraine p. 82-91 URL: https://aclanthology.org/2021.wanlp-1.9.pdf
35. Chuang J, Ramage D, Manning CD, Heer J. Interpretation and trust: designing model-driven visualizations for text analysis. In: Proceedings of the 2012 SIGCHI Conference on Human Factors in Computing Systems. 2012 Presented at: CHI '12; May 5-10, 2012; Austin, TX p. 443-452 URL: https://dl.acm.org/doi/10.1145/2207676.2207738 [doi: 10.1145/2207676.2207738]
36. Lwin MO, Lu J, Sheldenkar A, Schulz PJ, Shin W, Gupta R, et al. Global sentiments surrounding the COVID-19 pandemic on twitter: analysis of Twitter trends. JMIR Public Health Surveill 2020 May 22;6(2):e19447 [FREE Full text] [doi: 10.2196/19447] [Medline: 32412418]
37. Sisti LG, Buonsenso D, Moscato U, Costanzo G, Malorni W. The role of religions in the COVID-19 pandemic: a narrative review. Int J Environ Res Public Health 2023 Jan 17;20(3):1691 [FREE Full text] [doi: 10.3390/ijerph20031691] [Medline: 36767057]
38. Chu J, Pink SL, Willer R. Religious identity cues increase vaccination intentions and trust in medical experts among American Christians. Proc Natl Acad Sci U S A 2021 Dec 07;118(49):e2106481118 [FREE Full text] [doi: 10.1073/pnas.2106481118] [Medline: 34795017]
39. Padilla JJ, Kavak H, Lynch CJ, Gore RJ, Diallo SY. Temporal and spatiotemporal investigation of tourist attraction visit sentiment on Twitter. PLoS One 2018;13(6):e0198857 [FREE Full text] [doi: 10.1371/journal.pone.0198857] [Medline: 29902270]
40. Gore RJ, Diallo S, Padilla J. You are what you tweet: connecting the geographic variation in America's obesity rate to twitter content. PLoS One 2015;10(9):e0133505 [FREE Full text] [doi: 10.1371/journal.pone.0133505] [Medline: 26332588]
41. Frank MR, Mitchell L, Dodds PS, Danforth CM. Happiness and the patterns of life: a study of geolocated tweets. Sci Rep 2013 Sep 12;3(1):2625 [FREE Full text] [doi: 10.1038/srep02625] [Medline: 24026340]
42. Auer, Elena ML. Detecting deceptive impression management behaviors in interviews using natural language processing. Old Dominion University. URL: https://digitalcommons.odu.edu/psychology_etds/70/#:~:text=Given%20this%20limitation%2C%20natural%20language,IM%20in%20mock%20employment%20interviews [accessed 2024-04-29]
43. FarahAlshanik / COVID-Paper. GitHub. URL: https://github.com/FarahAlshanik/COVID-Paper [accessed 2024-04-29]

Abbreviations

API: application programming interface
LDA: latent Dirichlet allocation
NLP: natural language processing

Edited by T Mackey; submitted 07.10.23; peer-reviewed by M Elbattah, R Gore, MO Khursheed; comments to author 06.02.24; revised version received 06.05.24; accepted 08.12.24; published 10.02.25.

Please cite as:
Alshanik F, Khasawneh R, Dalky A, Qawasmeh E
Unveiling Topics and Emotions in Arabic Tweets Surrounding the COVID-19 Pandemic: Topic Modeling and Sentiment Analysis Approach
JMIR Infodemiology 2025;5:e53434
URL: https://infodemiology.jmir.org/2025/1/e53434
doi:10.2196/53434
PMID:39928401

Original Paper

Visualizing YouTube Commenters' Conceptions of the US Health Care System: Semantic Network Analysis Method for Evidence-Based Policy Making

Lana V Ivanitskaya[1], PhD; Elina Erzikova[2], PhD

[1]Health Administration, School of Health Sciences, The Herbert H. and Grace A. Dow College of Health Professions, Central Michigan University, Mount Pleasant, MI, United States

[2]Strategic Communication, School of Communication, Journalism and Media, College of the Arts and Media, Central Michigan University, Mount Pleasant, MI, United States

Corresponding Author:
Lana V Ivanitskaya, PhD
Health Administration, School of Health Sciences
The Herbert H. and Grace A. Dow College of Health Professions
Central Michigan University
208D Rowe Hall
Mount Pleasant, MI, 48859
United States
Phone: 1 989 774 1639
Email: ivani1sv@cmich.edu

Abstract

Background: The challenge of extracting meaningful patterns from the overwhelming noise of social media to guide decision-makers remains largely unresolved.

Objective: This study aimed to evaluate the application of a semantic network method for creating an interactive visualization of social media discourse surrounding the US health care system.

Methods: Building upon bibliometric approaches to conducting health studies, we repurposed the VOSviewer software program to analyze 179,193 YouTube comments about the US health care system. Using the overlay-enhanced semantic network method, we mapped the contents and structure of the commentary evoked by 53 YouTube videos uploaded in 2014 to 2023 by right-wing, left-wing, and centrist media outlets. The videos included newscasts, full-length documentaries, political satire, and stand-up comedy. We analyzed term co-occurrence network clusters, contextualized with custom-built information layers called overlays, and performed tests of the semantic network's robustness, representativeness, structural relevance, semantic accuracy, and usefulness for decision support. We examined how the comments mentioning 4 health system design concepts—universal health care, Medicare for All, single payer, and socialized medicine—were distributed across the network terms.

Results: Grounded in the textual data, the macrolevel network representation unveiled complex discussions about illness and wellness; health services; ideology and society; the politics of health care agendas and reforms, market regulation, and health insurance; the health care workforce; dental care; and wait times. We observed thematic alignment between the network terms, extracted from YouTube comments, and the videos that elicited these comments. Discussions about illness and wellness persisted across time, as well as international comparisons of costs of ambulances, specialist care, prescriptions, and appointment wait times. The international comparisons were linked to commentaries with a higher concentration of British-spelled words, underscoring the global nature of the US health care discussion, which attracted domestic and global YouTube commenters. Shortages of nurses, nurse burnout, and their contributing factors (eg, shift work, nurse-to-patient staffing ratios, and corporate greed) were covered in comments with many likes. Comments about universal health care had much higher use of ideological terms than comments about single-payer health systems.

Conclusions: YouTube users addressed issues of societal and policy relevance: social determinants of health, concerns for populations considered vulnerable, health equity, racism, health care quality, and access to essential health services. Versatile and applicable to health policy studies, the method presented and evaluated in our study supports evidence-based decision-making and contextualized understanding of diverse viewpoints. Interactive visualizations can help to uncover large-scale patterns and guide strategic use of analytical resources to perform qualitative research.

(JMIR Infodemiology 2025;5:e58227) doi:10.2196/58227

KEYWORDS

social media; semantic network; health system; health policy; ideology; VOSviewer; health care reform; health services; health care workforce; health insurance

Introduction

Background

The US health care system, characterized by high costs [1] and perceived to fall "far short of its potential" [2], has been a focal point for media attention and public commentary over the past decade. Discussions have revolved around topics such as the repeal of Obamacare, presidential health care agendas, the exorbitant costs of health care, comparisons to systems in other nations, and postpandemic health care personnel shortages. Throughout this period, conservative, moderate, and liberal media outlets have produced a variety of content, including newscasts, full-length documentaries, political satire, and stand-up comedy, all centered on the intricacies of the US health care system [3-6]. When disseminated through YouTube (Google Inc), the most popular platform among US social media users [7], select videos have generated millions of views and tens of thousands of comments. To the best of our knowledge, the perspectives of YouTube commenters on the US health care system and its reform, despite their considerable value for policy analysis, remain unexplored.

Objectives

Social media discussions are abundant, but they are often chaotic, noisy, indignant, and hateful [8-11]. There is a need for a method that effectively visualizes large volumes of commentary, filters out the noise, and highlights key patterns, making the information more digestible for stakeholders. The current state of social media research falls short of efficiently and clearly disseminating scientific outputs to diverse audiences. In quantitative social media studies, the constraints are statistical and graphical outputs with low idea density or high decoding requirements, which often require specialized knowledge. In qualitative studies, researchers communicate analytical outputs as summaries of themes and subthemes with representative quotes; however, they are based on limited data samples.

To address these challenges, we propose a mixed methods approach of mapping social media commentary. This approach combines automation and human judgment to create a visual representation of social media comments' contents and structure, presenting them as a semantic network [12]. This methodology is particularly relevant for researchers, policy makers, and the wider public seeking a better understanding of complex social media narratives. We repurpose VOSviewer (Centre for Science and Technology Studies at Leiden University), a user-friendly bibliometric tool, to analyze tens of thousands of social media comments on YouTube regarding the US health care system. In this study, semantic networks are graphical representations of social media comment meanings. Nodes represent terms frequently mentioned in YouTube comments, linked and grouped into clusters based on their co-occurrence.

Since their introduction in 2010, VOSviewer algorithms have been extensively applied to build term co-occurrence networks from the text of article titles and abstracts [13-20]. Visualization of nonbibliometric textual data as semantic networks in VOSviewer was proposed in 2011 [21], followed by early visualizations of Twitter and YouTube discussions ([22-25]). Subsequent explorations of VOSviewer's applications to social media comments and hashtags primarily led to cluster mapping ([26-35]). Notably, some scholars enhanced their cluster maps with informational layers called custom overlays to reveal patterns not visible in the base network [36-38].

Previous research compared VOSviewer semantic networks to networks generated from manually coded Twitter text [26]. However, there have been few systematic evaluations of VOSviewer-generated semantic networks derived from social media data. Consequently, our overarching goal is to evaluate VOSviewer's application to social media data: Can it produce credible semantic networks to be used as analytical and communication tools? We test VOSviewer's term co-occurrence map with custom-built overlays by posing 3 research questions:

1. How well does the VOSviewer network capture the content, context, and structure of social media comments?
2. What does it reveal about a decade-long online public discussion of the US health care system?
3. What is the policy analysis value of VOSviewer visualizations?

Methods

Semantic Network Construction

VOSviewer generates a custom semantic network by processing a corpus text file featuring social media comments. Our corpus comprised the text of primary comments and first-level replies to 53 videos shared by 17 US-based media outlets on their respective YouTube platforms between 2014 and 2023. The videos were sourced from news outlets such as Consumer News and Business Channel, Cable News Network, Fox News, and Public Broadcasting Service Frontline. Detailed criteria for video selection and video characteristics are outlined in the Tables S1 and S2 in Multimedia Appendix 1 [39]. After eliminating 5575 duplicate comments from the initial dataset of primary comments at first-level responses, our final corpus encompassed a total of 179,193 unique comments.

VOSviewer processes YouTube comments by detecting sentences, applying the Apache Software Foundation's OpenNLP library algorithm for part-of-speech tagging, identifying terms as nouns and the longest noun phrases, and unifying terms through various methods [17,18]. From an initial pool of 1948 terms appearing in at least 60 comments, a subset of 323 (16.58%) terms related to the US health care system, such as Obamacare, prescription, and wait time, was selected for the final semantic network. A detailed term selection process, including manual screening and thesaurus construction, is described in Multimedia Appendix 1.

By distilling 179,193 comments into a network with several hundred nodes, a macro model of YouTube video commentaries was created, providing insight into social media users' discussions on US health care. In this network, terms are interconnected and organized into distinct, nonoverlapping clusters [15,19,20]. A cluster is a group of terms tightly linked within the group and loosely connected with terms outside it. If >1 term was extracted from the text of the comment, it is possible for the same comment to be represented by multiple nodes in multiple clusters. We did a thematic analysis of clusters to gain insights about the US health system discourse.

We addressed limitations observed in previously published semantic networks by enhancing the network's informational value. First, we added custom overlays to VOSviewer's map, which displays the color of network nodes based on selected attributes. To build overlays, we coded each comment to reflect the theme of its YouTube video and added these codes, along with other comment characteristics (eg, comment date), to a scores file, which was uploaded to the VOSviewer together with our corpus file that contained YouTube comments (for more information on building corpus and scores files, refer to Multimedia Appendix 1). Second, we presented findings with hyperlinks to VOSviewer Online for broader accessibility and interactive engagement with our semantic network.

Network Interpretation and Evaluation

The evaluation of the US health care system's semantic network and its overlays was structured as follows. A comparison of 2 networks, before and after the deletion of repeated comments, served as a test of network robustness. Thematic alignment between the network terms, extracted from YouTube comments, and the videos that elicited these comments was a test of network's content representation.

To examine structural relevance, we asked if network relationships reflected the underlying meanings evident in YouTube comments. We examined clusters: Do terms in the same cluster have more similar meanings than terms in different clusters? We also examined pairs and groups of interconnected terms: Are they used together in the source data? Do their relationships align with existing knowledge? A comprehensive analysis of all pairs or term groups is outside of the scope of this study. For practical reasons, we engaged in close reading of a limited number of comments, focusing mainly on smaller nodes. When the number of comments exceeded 200, we randomly sampled 200 comments to cover discussions of different videos, taking care to sample more than once when we encountered heterogeneous ideas that required careful interpretation. When ≥2 nodes were examined, we used close reading of comments that mentioned all selected terms. Following the approach by Eve [40], network visualizations were used to locate "points of interest, which are then resynthesized into close readings."

In addition, we performed tests of semantic accuracy through raw data verification. Specifically, we cross-checked ambiguous or unexpected terms in our network against the comments that mentioned them. The analysis involved multiple readings of each comment to capture nuances of how individuals articulate their experiences or opinions of the US health care system, focusing on the words that were extracted as terms, their meaning, and context. On several occasions, for example, when performing a close reading for ideology, we offered brief summaries of the main ideas expressed by the commenters. Our validation of semantic network findings against extant comments adhered to the principles for quantitative text analysis outlined by Grimmer and Stewart [41].

Finally, we tested the usefulness of semantic network analysis for generating policy-relevant insights. We picked 4 health system design concepts—universal health care, Medicare for All, single payer, and socialized medicine—and examined how the comments mentioning these concepts were distributed across the terms we mapped. For insights into the policy ramifications of public perceptions of health system design, we focused on ideological terms and those with the highest share of comments referring to each concept.

Ethical Considerations

Ethics approval for this study was sought from Central Michigan University's Institutional Review Board (project 2023-1021-Mt. P). The study did not meet the definition of human participant research under the purview of the institutional review board according to federal regulations. The study used publicly accessible user-generated YouTube comments. The data were deidentified and aggregated before analysis. As the results are presented in an aggregate form, individual commenters cannot be identified. Informed consent has not been obtained. No compensation was provided to comment contributors.

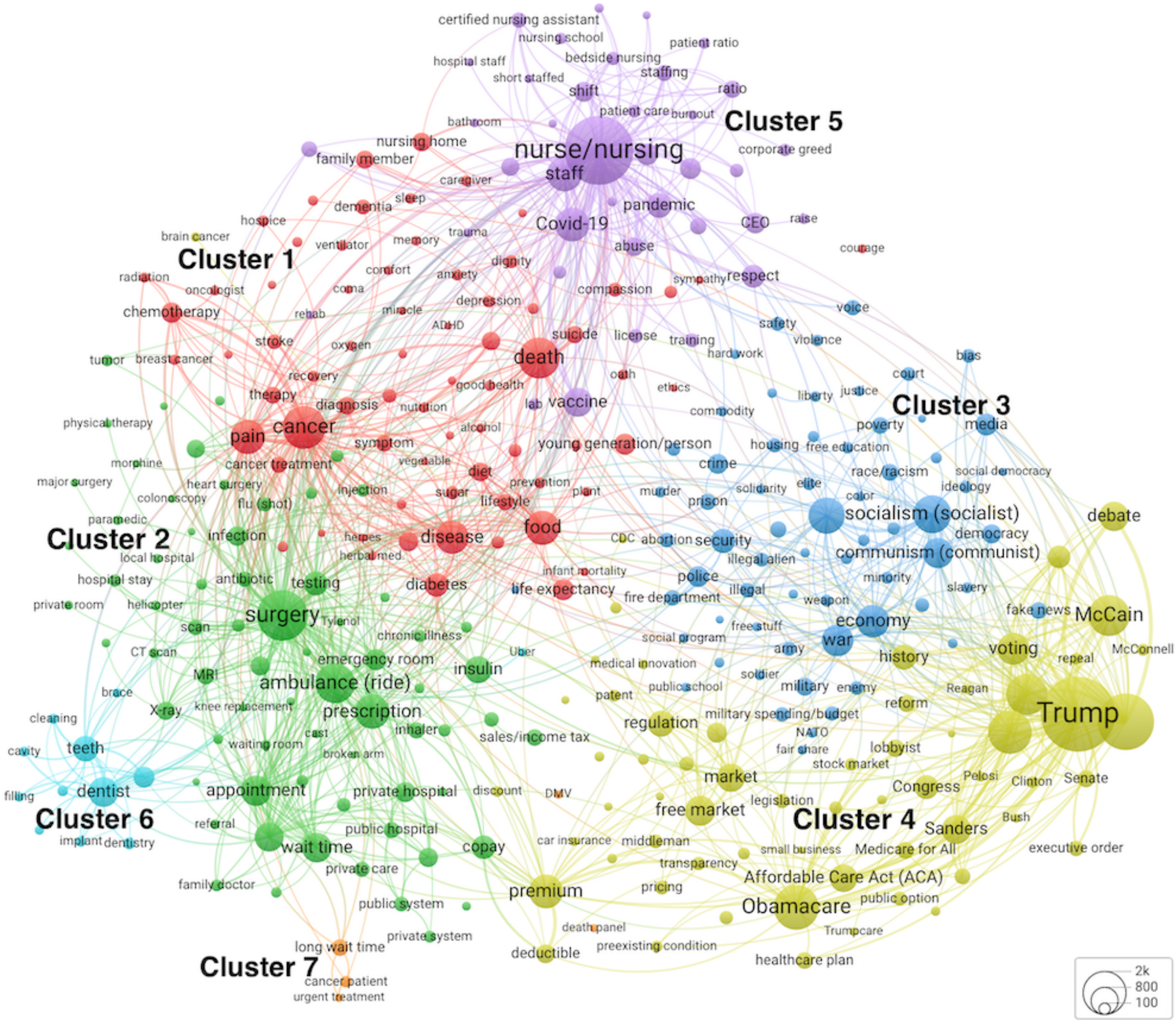
Results

A Semantic Network of Term Co-Occurrence and Clustering

From a manually screened list of 539 terms occurring in our corpus at least 60 times, VOSviewer's algorithm assisted in the selection of 323 (59.9%) most relevant terms [19]. Figure 1 [42] shows a 7-cluster solution for a term co-occurrence network.

On average, each term represented 357.74 (SD 606.88; median 163, IQR 104-321) comments. The longer the comment, the greater the likelihood that multiple terms were extracted from it. VOSviewer assigns cluster numbers based on the quantity of nodes; the same cluster numbers appear in our online interactive maps (URLs are provided in the notes of Figure 1).

Figure 1. A co-occurrence network (cluster map) of terms extracted from the comments on 53 YouTube videos about the US health care system. Binary-counted terms that occurred ≥60 times were mapped. An interactive map is available from Leiden University's VOSviewer app.



Cluster 1 (red) emerged as the largest group of nodes, covering chronic diseases, treatment, pain, and death. Its diverse terms also included topics related to disease prevention (*diet, exercise,* and *smoking*), mental health (*ADHD* [attention-deficit/hyperactivity disorder,] *anxiety*, and *depression*), and end-of-life issues (*hospice, euthanasia*, and *do-not-resuscitate*). Below it, cluster 2 (green) terms covered services, encompassing surgeries, emergency medical services, procedures, diagnostics, wait times, and discussions about public versus private health organizations and prescription medications. On the right, cluster 3 (dark blue) had terms about political ideologies, economic, societal, and cultural issues, surrounded by nodes from cluster 4 (yellow) related to political actors, institutions, the 2010 Patient Protection and *Affordable Care Act* (*ACA* or *Obamacare*), market regulation, and insurance terminology. The top of the map displayed a group of terms (cluster 5, purple) dedicated to health worker shortages, nurse-to-patient ratios, and nurses' burnout. Dental care terms formed a group on the lower left (cluster 6, light blue). Finally, a 5-node group (cluster 7, orange) at the bottom of the map had terms related to long wait times by patients with cancer who required urgent treatments, as well as further away terms *DMV* (Department of Motor Vehicles) and *death panel*.

The network displayed a rather coherent collection of terms, the meaning of which could be intuitively understood within the context of the US health care, with a few exceptions. For instance, as we manually selected terms for map inclusion, we checked the use of an ambiguous term *DMV* in YouTube users' comments. *DMV* was mentioned as a metaphor in a debate of government-managed health care efficiency. It was retained due to its relevance to the health care discourse.

The interpretive value of our network extended beyond a simple list of terms. The network specified links between terms that were often mentioned together, for example, *pricing* and *transparency* in cluster 4. Meaning extraction was further aided by the analysis of spatial proximity, cluster assignment, and cluster boundaries. For example, *preexisting condition*, as a term of interest, was directly and most strongly linked to *Obamacare* and *ACA*, which were mentioned with *preexisting condition* in multiple comments. This finding was consistent with a key ACA provision: insurance companies cannot use applicants' medical history to deny coverage or charge higher premiums based on their preexisting conditions [43]. Network structure's alignment with existing knowledge speaks to its structural relevance. *Preexisting condition* is located close to *premium,deductible, pricing,* market-related terms, and *government regulation* from cluster 4 about politics, as well as to *private health insurance* and *copay* on the far right of cluster 2, which is mostly dedicated to health care services. Therefore, when YouTubers discussed the US health care system, they used a noun phrase *preexisting condition* at the semantic intersection of health care politics and legislation, insurance pricing, and health services access.

In summary, the 323 networked terms, identified as most relevant by VOSviewer, unveiled discussions about illness and wellness; health services; ideology and society; the politics of health care agendas and reforms, market regulation, and health insurance; health care workforce; dental care; and concerns such as long wait times.

Before we removed 5575 duplicate comments, our original cluster map (Figure S1 in Multimedia Appendix 1) was quite similar to the cluster map in Figure 1. Our inquiry into the medical debt cluster comments uncovered repeated comments by a single YouTube user. After deletion, this cluster disappeared, but the network's overall structure largely remained intact, demonstrating its robustness.

Next, we examined clusters and nodes using overlays that reflected 2 aspects of the YouTube platform: the videos that elicited comments and the commentary itself. We assessed the usefulness of custom overlays as contextualization tools: Do they improve our understanding of nodes, node groups, and clusters? While we presented data on both video attributes and comment attributes, our analysis prioritized overlays depicting comment characteristics because they are more valuable for understanding digital publics' discussion of the US health care system.

Distribution of Video Groups Across Network Clusters

Thematic alignment between the video content that elicited the commentary and the commentary itself speaks to the content representativeness of the VOSviewer term co-occurrence network. The distribution of comments from 10 thematically diverse YouTube video groups across our term network is shown in overlays in Figure S2 in Multimedia Appendix 1. Our main findings are summarized in Table 1.

We observed substantial thematic congruence between video content and cluster terms. Nodes with above-average concentrations of comments related to the health care workforce were closely grouped in cluster 5, encompassing terms about nurses, staffing shortages, and management. Unlike most nodes in cluster 5, which were associated with health care workforce videos, the term *respect* had an above-average share of comments related to ACA and Obamacare reform videos. Our analysis of comments indicated that commenters mentioned respect for nurses, which explained the placement of *respect* in cluster 5. In addition, many comments on ACA and Obamacare reform videos expressed respect for Senator John McCain, which explained the connection between the term *respect* and *McCain*. *Respect*'s placement within cluster 5 but at its outer boundary, in the direction of node McCain, coupled with video overlay evidence, suggested semantic accuracy and structural relevance of our network.

Videos from 2 groups (health care policies, politics, ACA, and Obamacare reform) generated comments in cluster 4, which consisted of numerous political and reform-related terms. In addition, videos about health costs, one of which was titled "Dollars and Dentists," elicited discussions of dental care (cluster 6). Comments on videos about health care systems in different countries produced terms that appeared in multiple clusters but mostly in cluster 2 about health services and cluster 7 about long wait time concerns. At the same time, a Home Box Office video "Medicare for All" featuring John Oliver and a Netflix video featuring stand-up comedians making jokes about the US health care produced comments in nodes scattered across the map. The Netflix video showcased many comedians and topics, one of whom, Wanda Sykes, spoke about opioids from the perspective of racial and ethnic minority people. A commentary on this topic appeared in nodes *pain* and *prescription* (left side of the map) and *race/racism*, *Black person*, and *White person* (right side of the map), where commenters debated racial disparities in pain medicine access. For race-related nodes, the share of comments on the Netflix video (comedy on the US health care) varied between 1% and 8%, indicating that it was not the only video prompting the discussion. This finding is not unique; it was common for terms to represent commentaries to a wide variety of videos or video groups.

Across all video group overlay legends, the highest scale midpoint was 0.25 for videos about health care costs and financial issues. It means that, on average, 25% (SD 14%) of comments within a term come from that video group. Across 323 map terms and 10 video theme overlays, there were only 11 (0.34%) instances (out of 3230 possible instances) where terms represented >90% of comments from a single video group.

Table 1. Characteristics of videos that elicited comments related to cluster-specific terms.

Cluster number (color)	Topical areas	Cluster's 10 largest terms	Video groups that elicited comments related to most, some, or specific terms within a cluster
1 (red)	Illness and wellness, including mental health and end of life	*Cancer, death, pain, food, disease, diabetes, young generation/person, life expectancy, chemotherapy,* and *cure*	• Children's health care (some terms) • End-of-life health care (some terms) • Health care systems in different countries (*young generation/person* and *life expectancy*) • Comedy on the US health care (*pain*) • Medicare for All video by John Oliver (*pain*)
2 (green)	Health services	*Surgery, ambulance (ride), prescription, appointment, wait time, specialist, insulin, testing, copay,* and *emergency room*	• Health care systems in different countries (most terms) • Medicare for All video by John Oliver (most terms) • Comedy on the US health care (*prescription*)
3 (dark blue)	Ideology and society	*Socialism (socialist), capitalism (capitalist), economy, war, communism (communist), security, media, police, crime,* and *democracy*	• Single-payer health care (most terms) • Health care systems in different countries (some terms) • Medicare for All video by John Oliver (some terms) • Health care costs and financial issues (*capitalism*) • Comedy on the US health care (*race/racism, Black person,* and *White person*) • ACA[a]/Obamacare reform (*race/racism, Black person,* and *White person*)
4 (yellow)	Health care politics, reform, market regulation, and insurance	*Trump, Biden, Obamacare, Republican, Democrat, McCain, premium, voting, free market,* and *debate*	• Health care policies and politics (most terms) • ACA/Obamacare reform (most terms) • Medicare for All video by John Oliver (some terms) • Single-payer health care (some terms) • Health care costs and financial issues (market regulation terms)
5 (purple)	Health care workforce	*Nurse/nursing, staff, Covid-19, vaccine, pandemic, respect, shortage, management, CEO*[b], and *shift*	• Health care workforce (most terms) • Health care systems in different countries (vaccine) • ACA/Obamacare reform (*respect*)
6 (light blue)	Dental care	*Dentist, teeth, dental care, dentistry, implant, dental insurance, cleaning, cavity, filling,* and *brace*	• Health care costs and financial issues (most terms)
7 (orange)	Concerns	*Long wait time, cancer patient, DMV*[c], *urgent treatment,* and *death panel*	• Health care systems in different countries (most terms) • Single-payer health care (*DMV*)

[a]ACA: Affordable Care Act.

[b]CEO: chief executive officer.

[c]DMV: Department of Motor Vehicles.

Comment Date and Ongoing Discussions

When considering the timing of comments, the overall mean for all nodes was December 2020 (mean 2020.99, SD 0.81; range: from early 2018 for *repeal*, referring to the Trump administration and Republican lawmakers' efforts to repeal the ACA, to early 2023 for *do-not-resuscitate*). Clusters 1, 5, and 6 have terms with more recent comments than other clusters (Figure 2, left [42]), which is likely a function of when a video was uploaded on YouTube.

Also shown in Figure 2 are ongoing discussions, conceptualized at the term level as mean posting time since the first comment in the respective video. We calculated time for each comment, based on the video it came from, then averaged across all comments behind each term. The terms that scored above the midpoint of 0.49 years (approximately 6 months) highlighted areas on the map where YouTube users continued to contribute comments long after the videos were posted, serving as a proxy for ongoing interest and engagement. Comment scores were calculated in 2 ways: without standardization, expressed as a fraction of a year (Figure S3 in Multimedia Appendix 1), and with standardization, using the base-10 logarithm to adjust for skewed data. The standardized scores were then normalized so that the mean is 0 and the scale points represent SDs (Figure 2, right).

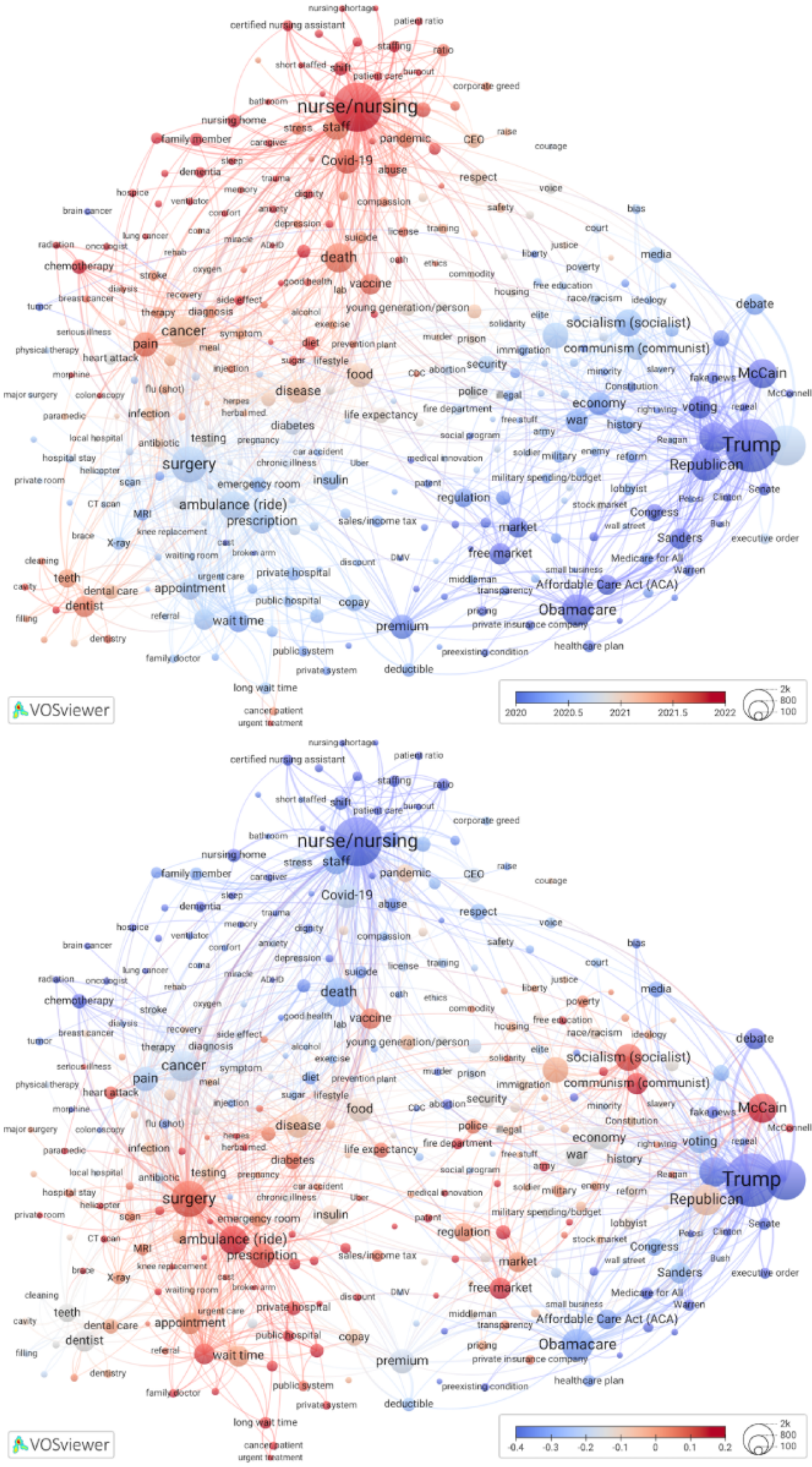
Ongoing discussions in cluster 1, “illness and wellness,” were about cure (*herbal medicine* and *herpes*), *diabetes*, and life expectancy, and young people persisted, on average, for 11 months. In cluster 2, “health services,” ongoing discussions revolved around ambulances, specialist care, prescriptions, appointment wait times, copays, and private (vs public) health insurance or services, roughly covering the same area as high-scoring nodes in an overlay for videos about health care systems in different countries. YouTube commenters demonstrated continued interest in these topics. On average, cluster 2 terms that scored above the mean came from comments posted approximately 9 months after the first comment on a given video.

In cluster 3, “ideology and society,” YouTube users’ comments on political ideologies, police, and military were typically added around the 8-month mark, on average. To better understand an unexpectedly salient group of ideological terms in our map, we analyzed hundreds of comments about communism, socialism, and capitalism. Our analysis confirmed node size and interconnectedness. The discussion of the US health care system was highly politicized, with ideological battles that revolved around dichotomies, such as socialism versus capitalism. Individuals who self-identified as capitalist, conservative, libertarian, or Republican outright rejected any government involvement in health care, calling it socialism, which was often equated with communism (thus confirming node proximity), social democracy, inefficiency, economic decline, and excessive control. Commenters who self-identified as progressive, liberal, social democrat, or left leaning pointed out that health care in the United States was already a mix of capitalism and socialism: publicly funded US police and army were essentially socialized law enforcement, similar to socialized medicine in other countries. They saw no logical reason to reject socialized medicine.

Moreover, several non-US commenters and US residents living abroad shared their positive experiences with health systems in Europe and elsewhere, pointing out that they were affordable to residents with low-income status. Commenters questioned the following: Why do Americans accept *GoFundMe* fundraising to cover medical expenses but not universal health care? Those who defended capitalism praised it for *medical innovation* and high quality of health care but often added that it must be properly regulated. Application of capitalist principles to the US health care system was also discussed in connection to greed, lack of access to health care services, inequities, and poor outcomes. Multiple comments suggested that every economy needed a mix of socialism (relating it to public good or public welfare) and regulated capitalism to counterbalance corporate interests.

Finally, in cluster 4, “health care politics, reform, market regulation, and insurance,” we observed ongoing discussions about market-related topics (*monopoly*, *regulation*, and *market*) and especially the role of John McCain during Obamacare repeal.

Figure 2. Overlays to Figure 1 for mean comment date (top) and ongoing discussions (standardized scores, bottom). High-resolution versions are available in Multimedia Appendix 1 (Figures S4 and S5).



Comment Likes

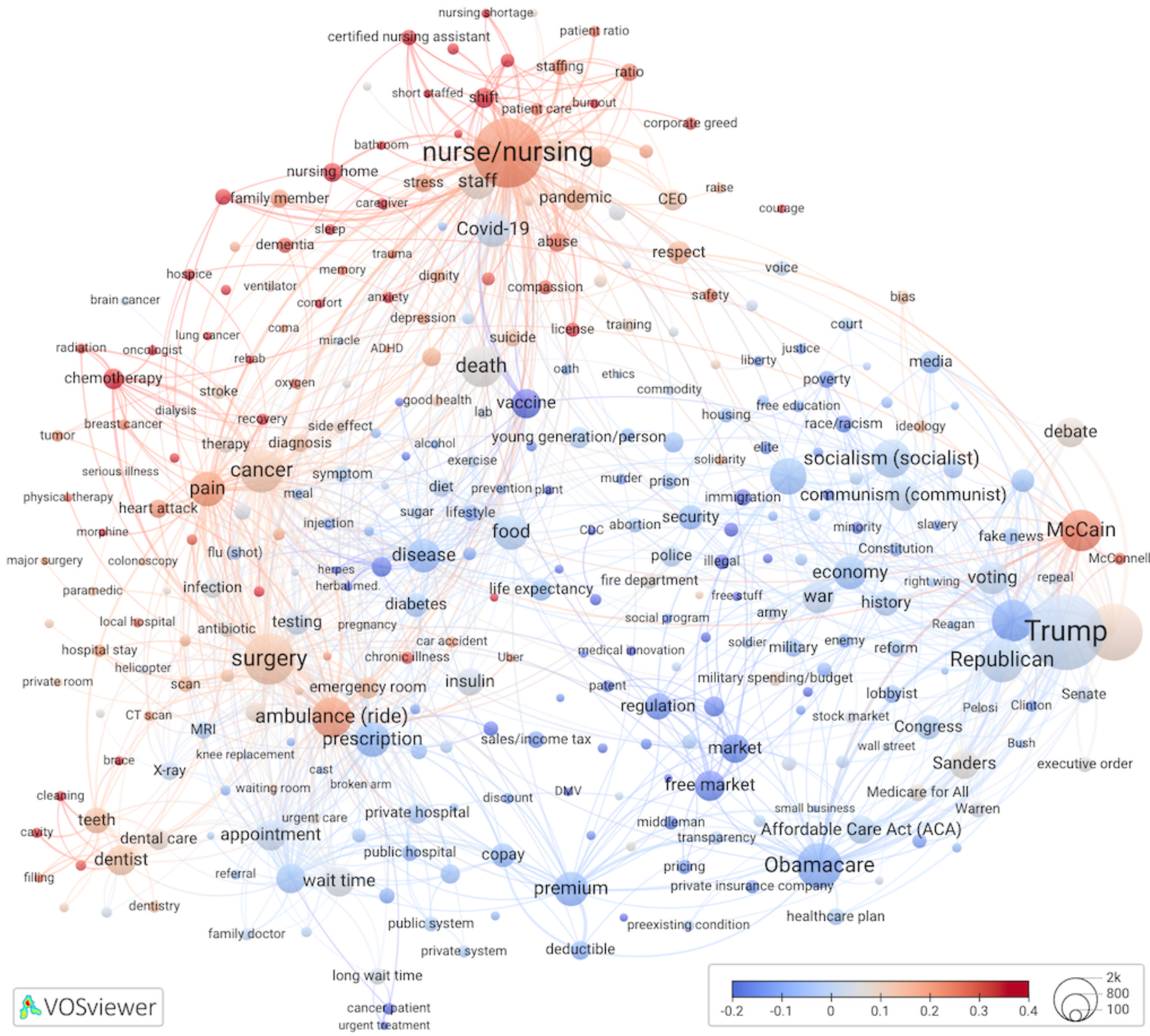
Comment likes were standardized using the same method as for ongoing discussions. We examined overlays for cluster-specific concentrations of terms that scored above the mean, identified dyads of linked terms that scored high, and summarized the most-liked comments from a specific cluster or term.

In Figure 3 [42], the largest concentrations of above-average liked comments were mostly cluster specific (clusters 1, 2, 5, and 6). Most-liked cluster 5 terms came from comments about shortage of nurses and nurse burnout as well as factors contributing to it (*shift*, *short staffed*, *corporate greed*, *patient ratios*, *abuse*, and *management*). We checked an unexpected connection between *shift* (0.58 SD above the mean for all terms) and *bathroom* (0.48 SD above the mean), which represented highly liked comments. A total of 20 unique commenters shared stories of extreme job demands, describing how nurses worked long shifts, endured heavy workloads, faced high patient-to-nurse ratios, and had to wait for breaks to address their physiological needs. All but 3 commenters self-revealed their profession. They were experienced nurses, practicing or retired, or nursing students on clinical rotations. Their detail-rich comments described burnout antecedents, such as profits over staffing, mistreatment of nurses, and mandatory overtime, and outcomes, for example, reduced patient care quality and medication errors.

In cluster 1, cancer-related terms, the term *sleep*, and end-of-life terms such as *do-not-resuscitate* were extracted from comments with many likes. Individuals who mentioned “do not resuscitate” (DNR; 0.42) expressed deeply personal desires for autonomy and the avoidance of prolonged distress at the end of life. The commenters identified themselves as older adults, patient advocates, veterans, or health care workers. They discussed the implications of DNR orders, sometimes expressing doubts that an overburdened health care system could handle their implementation in a patient-centered way. Nevertheless, some nurses who witnessed slow deaths of patients without DNR orders chose to create their own advance directives.

Comments about *sleep* were also well liked (0.43) but, unlike the DNR discussion, referred to many different contexts: caregivers, including nurses, experiencing stressors and sleeplessness; sleep as a precondition to wellness; and in the context of passing away peacefully in one’s sleep. The placement of *sleep* within our network, on the boundary of cluster 1 terms (*dementia*, *family member*, *nursing home*, and *caregiver*) and cluster 5 terms (*stress*, *trauma*, and a direct link to *nurse/nursing)*, matched these observations and provided evidence of semantic accuracy and structural relevance.

Figure 3. A mean comment likes (standardized) overlay to Figure 1.



Among dental treatment nodes in cluster 6, *cavity* scored the highest (0.48) on comments with likes. Cavity-related comments came from individuals who revealed the following self-identifications: residence (mostly the United States but also US residents living abroad and foreign nationals from multiple continents), low income (jobless or poor), and medical tourists (eg, US residents receiving dental treatments in Mexico). Commenters particularly liked quotes of low dental costs in Australia, France, Mexico, and other countries; stories of cost savings after buying airfare and paying for dental treatments abroad; personal accounts of dentists recommending unnecessary procedures; and oral health tips, such as reducing sugar intake. Comments specified systemic problems with US dental care: financial strains, even with dental insurance; potentially superfluous, according to second opinions, or unnecessarily extensive procedures (eg, on baby teeth); worsened conditions due to cost-related treatment delays; and processed sugar industry's influence on consumption of foods, leading to dental decay.

Other clusters also had node groups that were well liked. We explored 2 dyads of linked nodes that scored high on likes: *McCain–McConnell* (0.31-0.34, cluster 4) and *ambulance (ride)–Uber* (0.26 for both, cluster 2), with above-average likes. In first dyad comments, most commenters applauded McCain's vote that helped prevent the repeal of ACA and criticized McConnell and other Republicans. Comments from the second dyad, *ambulance* and *Uber,* were by YouTuber users who expressed concerns about the cost of US ambulances and Americans' reluctance to use specialized emergency transportation. To avoid unpredictable costs, some US commenters planned to use nonmedical transport, such as ride-sharing services like Uber, during health emergencies.

Comments With Select British Spellings

Figure S6 in Multimedia Appendix 1 displays an overlay that approximates contributions from commenters whose backgrounds are associated with regions where British spelling conventions are more common than in the United States. Such spelling was detected in multiple clusters, but the highest-scoring terms were in cluster 2 (*national insurance*, *government hospital*, and *private system*) and cluster 3 (*free education*, *unemployment*, and *justice*).

Commonly Mentioned Health Care Concepts: System Design Ideas

Our last set of overlays demonstrates the distribution of comments that mention policy-relevant ideas on health care system design: universal health care, Medicare for All, a single-payer system, and socialized medicine (Table 2). VOSviewer Online offers a modifiable legend with an option to normalize term scores by subtracting mean and dividing by SD. When term scores are normalized, we can directly compare multiple overlays (Figures 4 and 5 [42]) to identify map areas with terms that are extracted from a high (vs low) share of comments mentioning specific system design ideas. Unlike the standardization of comment scores, normalization is performed at the term level.

Table 2. Mentions of health care system design ideas.

Attributes	Design idea overlay[a]			
	Universal health care	Medicare for All	Single payer	Socialized medicine
Definition[b]	A system where all citizens have access to health care services without financial hardship	A proposed system to expand the US Medicare program to cover all individuals, eliminating private insurance	A system where a single entity (usually the government) pays for all health care costs	A system where the government not only funds but also provides the health care services
Comments, N	3638; "universal health" or "universal healthcare"	2909; M4A or "medicare for all"	1474; "single payer" or "single-payer"	716; "socialized medicine" or "socialised medicine"
Prevalence of comments that mention each design idea within a term-specific comment collection				
Highest-scoring term on a corresponding overlay	*Private room* (12/95, 12.6% of comments also mention universal health care)	*Warren* (116/276, 42% of comments also mention Medicare for All)	*Administrative cost* (16/108, 14.8% of comments also mention single payer)	*Medical innovation* (5/108, 4.6% of comments also mention socialized medicine)
Share of comments within ideological terms[c]				
Socialism/socialist	+1.44 SD	+0.04 SD	−0.16 SD	+0.64 SD
Communism/communist	+3.06 SD	−0.18 SD	−0.65 SD	+0.45 SD
Capitalism/capitalist	−0.33 SD	−0.28 SD	−0.49 SD	−0.53 SD

[a]Interactive overlays are available from the left panel (view>items>color >) [42].

[b]Commenters defined health system design ideas in different ways and sometimes used them interchangeably. For example, some commenters talked generally about a state-managed health care system in reference to both single payer and socialized medicine.

[c]Normalized health system design idea overlay scores for 3 ideology nodes are shown relative to all nodes' mean share of comments mentioning that specific health system design idea. Plus or minus signs refer to above or below all terms' mean share, expressed in SD units, within each health system design idea overlay.

Figure 4. Overlays to Figure 1 depicting the distributions of comments that mention “universal health” (top) and “Medicare for All” (bottom). High-resolution versions are available in Multimedia Appendix 1 (Figures S7 and S8).

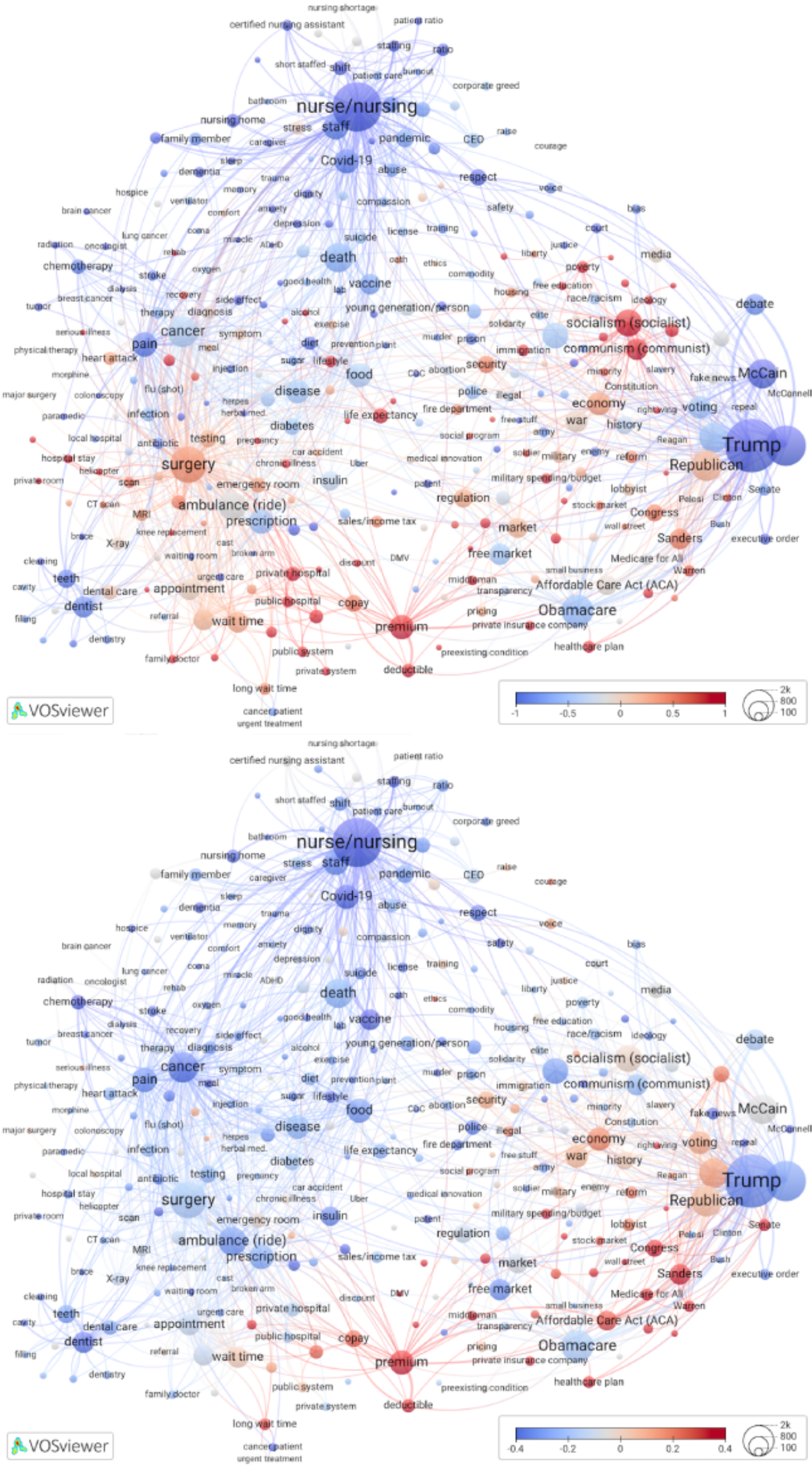
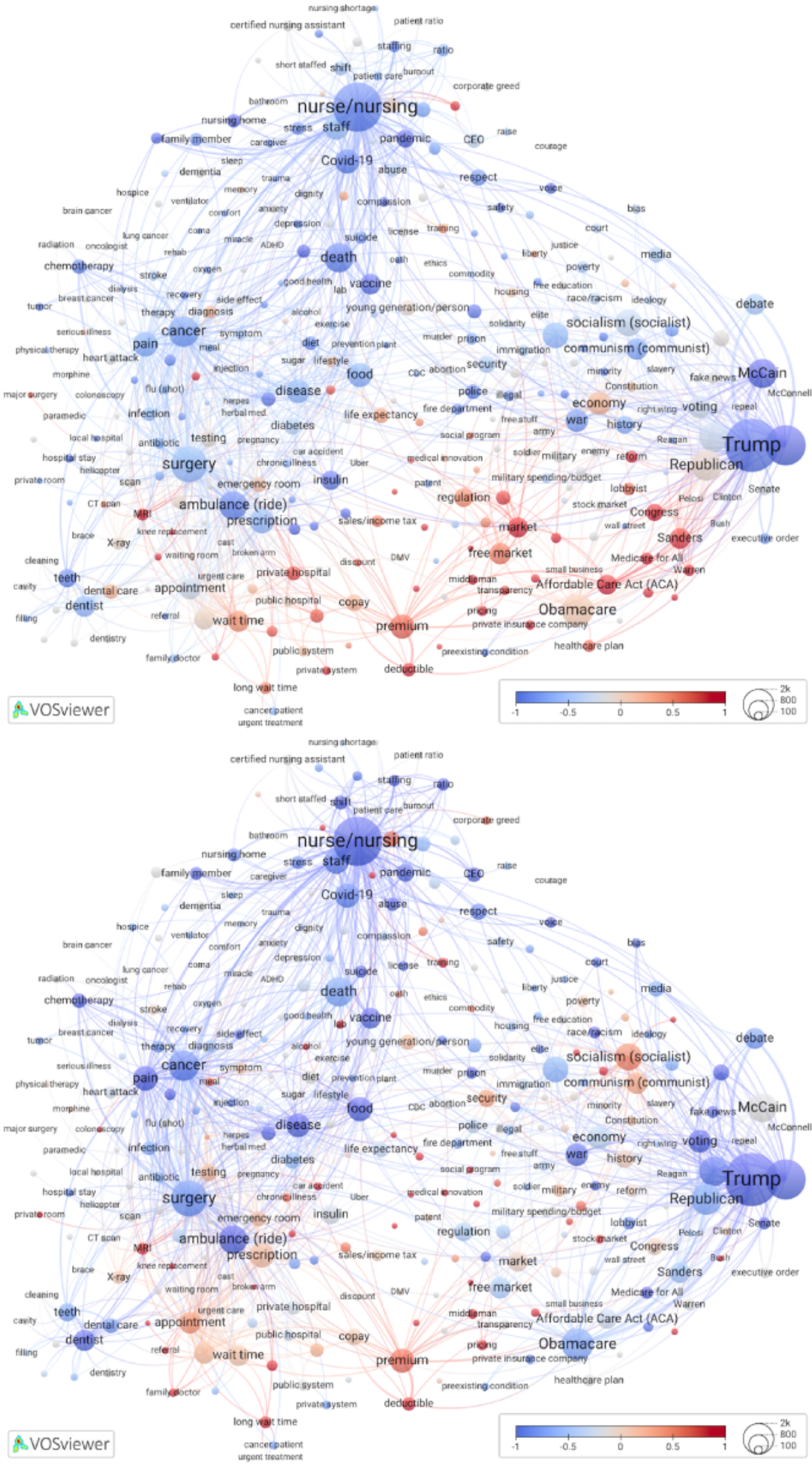


Figure 5. Overlays to Figure 1 depicting the distributions of comments that mention "single-payer" (top) and "socialized medicine" (bottom). High-resolution versions are available in Multimedia Appendix 1 (Figures S9 and S10).



As shown in Table 2, the most frequently mentioned health system design idea in our comments—universal health care—was discussed in connection to *private room*, the highest-scoring term on the universal health overlay. The term *private room* also had above-average share (3/95, 3%) of comments, with at least 1 (6%) of 18 British-spelled words. US residents and foreign nationals discussed semiprivate and private hospital rooms as a desirable high standard for hospital stays.

Commenters with experience in universal health systems explained that such systems serve everyone but may not provide extra luxuries unless a patient is also covered by private insurance or pays out of pocket. Several comments expressed preferences for universal health care systems with balanced public and private health care. Private rooms, marble floors, and hotel-like amenities in US hospitals were discussed as luxuries available to the rich, while care was being denied to the poor.

At the bottom of Table 2, we show 3 ideological terms and compare the extent to which they are linked to each health system design idea. For universal health overlay, the data address the following question: In node *socialism/socialist*, what is the share of comments that mentioned universal health and how far is this share, in SD units, away from the universal health care overlay's mean for all nodes? Compared to 3 other concepts (Medicare for All, a single-payer system, and socialized medicine), universal health care was most strongly linked to discussions of communism and socialism. Specifically, the share of universal health care comments in the node *socialism/socialist* was much greater than that in most other nodes (1.44 SD above all terms' mean). It was even higher for the node *communism/communist* (3.06 SD above the mean).

While discussing Medicare for All in early 2020, YouTube commenters were concerned that it was insufficiently supported by Elizabeth Warren, as compared to Bernie Sanders, which explains why *Warren* was the highest-scoring term in the Medicare for All overlay. In addition to questioning the political viability of Medicare for All, commenters expressed concerns about its funding and tax increases, possible loss of preferred private insurance, unemployment among health insurance workers, increased wait times, diminished quality of care, and fluctuating government or political control over reproductive health.

The highest-scoring term on the single-payer overlay, *administrative cost*, was often mentioned with a term *middleman*, an unnecessary intermediary, for example, private insurance companies and for-profit corporate interests. Discussions of single payer, administrative costs, and middlemen turned into debates. Advocates cited the potential for significant savings and increased efficiency by eliminating the profit-driven insurance model. They pointed to Medicare's low overhead as evidence that a single-payer system could reduce administrative costs. By cutting out middlemen, single-payer systems bring down administrative costs and simultaneously simplify system navigation and transactions for patients, restrain profiteering, reduce health care fraud, and open health care systems to cost control. Critics, however, expressed skepticism about the efficiency of government-run systems, cautioning that replacing one bureaucratic structure with another may not achieve the expected reductions in administrative costs.

Finally, the term *medical innovation* had the highest share of comments that mentioned socialized medicine. The comments often referred to the United States's top position in producing medical innovations. Several US commenters suggested that countries with socialized medicine rely upon US innovations without contributing comparable advancements in new treatments or medical technologies. US medical innovations, according to their comments, come at high cost but also contribute to high quality of care. Others expressed disagreement, saying the United States ranked fourth on medical innovation, behind Switzerland, Germany, and the Netherlands. In addition, hopes were expressed that rising costs of US health care could be controlled through medical innovations, especially in older adult care.

Of the 4 health system design ideas we analyzed, the concept of single-payer health system had the lowest use of ideological terms. The distribution of scores across the single-payer overlay shows that single-payer discussions were less prevalent in ideological terms (*socialism/socialist*, *communism/communist*, and *capitalism/capitalist)* than in other terms we mapped. In the *socialism/socialist* node, an above mean share of comments about Medicare for All (+0.04 SD), socialized medicine (+0.64 SD), and especially universal health care (+1.44 SD) indicated greater use of ideological terms, as compared to single-payer discussions (–0.16 SD). In addition, the universal health care discussion was much more centered around communism or communist (+3.06 SD) compared to the single-payer discussion (–0.65 SD).

Discussion

Overview

We discuss 2 sets of findings. First, we summarize our evaluation of the semantic network. We elaborate on the implications of repurposing VOSviewer to subsequent social media studies and anticipate scientific advances that may result from its broad application. Second, we summarize our US health system insights and discuss their policy implications, pointing out limitations.

VOSviewer Term Co-Occurrence Network as a Social Media Analysis Method

VOSviewer is one of several programs available to researchers for conducting semantic network analysis. For example, previous studies have used the Fruchterman-Reingold algorithm [44], Gephi [45], and R [46] to build semantic networks. At the same time, VOSviewer's user-friendly interface is suitable for users without advanced technical skills. Regardless of the tools used in their construction, semantic networks promise to represent knowledge, while their interconnected nodes likely capture meaning [12], as demonstrated by this analysis.

We used VOSviewer as a data visualization tool to respond to the critical need to decrypt chaotic and extensive social media discussions on a socially important topic. Our analysis suggests that VOSviewer produces visualizations with high information density, interactivity, and interpretive richness. In addition, we obtained evidence regarding the following characteristics of the VOSviewer-generated network: (1) robustness or resilience to variations in data, (2) content representativeness of the diversity of issues related to the US health system, (3) structural relevance defined as meaningful network relationships, and (4) semantic accuracy defined as accurate representation of comment meaning. Our evaluation of the network's decision support usefulness is discussed in the US Health System Insights and Their Policy Implications section.

First, our limited test of robustness confirmed the network's resilience to the removal of approximately 3% of repeated comments from our corpus. If such comments were retained, identical comments by just 1 social media user would have produced a user-specific map cluster about medical debt and bankruptcy. Striving to build a network reflective of broad conversations, we chose to remove it, but the comments we removed were relevant to the US health system. The person who posted them might have tried to express desperation or draw attention to the seriousness of medical debt.

Second, the network comprehensively covered 10 thematic video groups, representing the entire diversity of video content about the US health care system. In other words, comments from all video groups were represented within the network nodes. Third, we observed a meaningful cluster layout that, overall, could be intuitively interpreted. Structural relevance was confirmed by spatial arrangement of nodes in the network, where the proximity of nodes corresponded to the co-occurring nature of the semantic relationships observed in the text from which the nodes were derived. Moreover, the network's structure aligned with existing knowledge, for example, ACA provisions. Forth, multiple checks confirmed that the mapped terms, including unexpected or ambiguous ones, captured the meanings of posts as well as their context.

Anticipated Scientific Advances of the VOSviewer Application to Social Media Analyses

The VOSviewer's term co-occurrence mapping method and their custom overlays can advance computational social sciences through informative, contextualized semantic networks. Natural language processing enables unbiased extraction of relevant terms, with an option of manual term screening. Revealing large patterns in extensive source data, VOSviewer "visual narratives" [47] can guide researchers to efficiently allocate their analytical resources as they explore salient patterns of societal importance embedded in "context or domain-specific knowledge" [48]. As such patterns involve network terms—nouns and noun phrases that occur in comments—researchers can strategically focus on the most promising subsets of extant data. In addition, VOSviewer-enabled semantic networks bring to light the interdisciplinary nature of social media studies. According to our cluster map, an in-depth analysis of public perceptions of the US health system calls for input from scholars in fields such as communication, economics, health care management, medicine, political science, public health, and others.

Clusters model thematic structure at a macro scale; overlays provide interpretive richness. The method we demonstrated here offers a valuable way for researchers to experience relationships embedded in source data, some of which are hard to document using conventional analyses. Chronological overlays that show video dates, comment dates, and lags in time between the first and the nth comment offer clues on how the discussion progressed over time, enabling a study of unfolding discourses. This is particularly relevant for data from social media platforms, which are "inherently longitudinal" [48]. With additional automation, it would be possible to create dynamic network visualizations that are updated in near–real time as new comments are posted.

Another benefit of semantic map overlays is that they foster cluster exploration and hypothesis testing by combining different data sources. For the YouTube platform, overlays may reflect characteristics of comments, YouTube video channels, videos themselves, or social media users' channels. Therefore, visual overlays represent many opportunities for innovation and experimentation. For example, information excluded during term selection can be brought back in overlays. In this study, we removed geographical references from the cluster model's nodes but created an overlay to highlight discussions with British spelling.

The method we demonstrated in this study can also enhance the value of qualitative research. Resource-intensive qualitative methods can be deployed strategically, guided by the grasp of larger patterns evident in semantic networks. Semantic networks can be contextualized and nuanced through qualitative coding. The qualitative codes can then be incorporated into custom-designed overlays, leading to new hypotheses and qualitative analyses. This iterative approach enables visualization-assisted qualitative inquiry.

Given these methodological strengths, we believe that VOSviewer-enabled semantic network analyses of social media data can advance social science research in the digital era. Thinking even broader, the proposed method can be applied across a variety of contexts and data sources, not limited to social media, and across different disciplines, such as computational humanities.

US Health System Insights and Their Policy Implications

Overview

Health care debates unfold in both in real life and online spheres. We examined digital publics' discourse about the US health care system in response to YouTube videos from right, center, and left media outlets. The YouTube platform allows purposeful selection of videos by varied media outlets on different aspects of an issue. We provided evidence that thematic diversity of videos was passed on to the commentary, opening a door to the policy-relevant analysis of diverse viewpoints. The YouTube platform has emerged as a space for heated debates, thoughtful ideas, misconceptions, and personal narratives in response to the US health care system.

Understanding the viewpoints by social media users provides valuable input for policy makers, health care professionals, and advocates aiming to shape effective reforms. The insights gleaned from the VOSviewer semantic network carry significant implications, which we grouped into 3 categories (concerns about the health care system, domestic and global interconnections in health care discussions, and informing change through key health care discourse insights).

Concerns About the Health Care System

The clusters shed light on a wide range of areas of concern within the US health care system, including those that are likely to be voiced by the public when politicians mention universal health care, Medicare for All, a single-payer system, and socialized medicine. The network analysis was helpful in estimating the use of ideological terms in discussions of various health system design ideas and identifying related concerns, for instance, about continued medical innovation or patients' access to private hospital rooms. The ideology and society cluster terms, derived from politicized comments, reflect the entrenched ideological conflicts and capitalism-socialism dichotomies within the YouTube discourse about the US health care system.

We observed that comments in the health care workforce cluster, particularly those about staff shortages and burnout, received many likes. This pattern points to a widely shared perception of the urgent need to address challenges faced by nurses and other health professionals. If corroborated across time and other data sources, this sentiment may translate into public support for health care reforms that enhance workforce well-being, improve nurse-to-patient ratios, and support the essential role of health care workers in the system.

Online discussions also highlight ongoing debates about the balance between public and private health care services. Policy makers can use these insights to formulate strategies that optimize the strengths of both sectors, ensuring accessibility, affordability, and quality of care. In sum, a VOSviewer-generated semantic network with overlays shows promise as a decision support tool for policy makers.

Domestic and Global Interconnections in Health Care Discussions

Health care reforms should consider the broader societal and political context of the country to build sustainable and politically viable solutions. The health care discourse we described incorporated widespread debates about political ideologies, societal issues such as racism, and economic considerations. While many of these issues were domestic, there was also a significant international component. Terms such as *national insurance*, *government hospital*, *private system*, *free education*, *unemployment*, and *justice* represented 6% to 8% of comments with at least 1 British-spelled word from our list. In much smaller concentrations (2.5%-4%), British-spelled comments appeared in the wellness discussion (*nutrition*, *vegetable*, and *memory*) and conversations about tax break (or cut), social health care, and private insurance companies. Adding evidence in support of semantic accuracy, several terms extracted from a nonzero share of British-spelled comments (*national insurance* and *social health care*) described societies outside of the United States.

The presence of British-spelled words in our data indicated the global nature of US health care discussions, which is evident in international comparisons of prices and patient experiences. YouTube discussions offered opportunities for US social media users to learn about foreign health systems and explore their benefits, trade-offs, and foundational values. The information was conveyed not by experts or politicians but by laypeople who had encountered foreign systems as taxpayers and patients. Some informants lived in several countries and could compare multiple systems. Informed by global perspectives, the US public may shift its expectations, prompting politicians to incorporate best practices, for example, affordable drugs and predictable costs of emergency patient transportation, into reform initiatives. At the same time, both the public and policy makers stand to benefit from reexamining their own misconceptions and rigid ideological beliefs in light of successful health care models and practices in other countries.

Informing Change Through Key Health Care Discourse Insights

Our semantic network analysis provides insights into the topics that garner the most attention and engagement in ongoing discussions. Health care reforms can be supported by targeted public education and awareness campaigns addressing these key themes, fostering informed public discourse and encouraging active participation in the reform process. Accordingly, policy makers should continuously monitor public sentiments on platforms such as YouTube to inform dynamic, responsive health care policies that adapt to changing societal needs and concerns. Finally, leveraging user engagement patterns, particularly standardized likes and ongoing discussions, can establish effective feedback loops between policy makers and the public. Understanding which aspects of the discourse resonate most strongly with the public allows for the refinement of reform strategies. We provided empirical evidence of links between specific public opinions on health system designs and ideological discourse; comments about universal health care had a much higher use of ideological terms than discussions of single-payer health systems. Overall, the key takeaways drawn from the VOSviewer-generated semantic network analysis provide actionable insights for shaping reforms in health care, which are responsive, inclusive, and aligned with the diverse perspectives expressed by the public on digital platforms.

Finally, we share 2 observations on how VOSviewer maps may support evidence-based policy making and communicating with stakeholders. One consideration is the empirical rootedness of the information we mapped. Decision makers are more likely to accept and act upon information perceived as “evidence based” [48], for example, maps that display intuitively interpretable terms grounded in actual text. In the study by van der Voort et al [47] on big data, decision makers “wanted ‘stories to tell’ to feed public debate and highlight problems and opportunities,” favoring reports at higher resolutions. In our study, clusters communicated broad narratives about the public discourse of the US health system, while overlays enriched and contextualized interpretation of narratives, adding complexity and specificity.

How well decision makers with different levels of education can decode VOSviewer data visualizations remains to be tested. We anticipate that for most decision makers, the learning curve of interpreting maps will be less steep than that for statistical outputs with comparable informational value. While overlays provide a multidimensional understanding of the discourse, they may be harder to decode than clusters. At the same time, the interactive nature of VOSviewer Online is likely to add interest and user engagement, helping to translate research findings into informed decision-making and actionable policy measures.

Limitations

While VOSviewer offers a powerful tool for visualizing and analyzing co-occurrence networks, the algorithm’s effectiveness is contingent on the initial selection of terms. The manual screening of a list of terms introduces a potential bias. In addition, the study is limited to English language YouTube comments, which may not fully capture the broader public discourse on health care.

Further research is warranted to validate and expand upon our results. Future studies could use other advanced natural language processing techniques to enhance the accuracy of term selection and clustering. Moreover, a multiplatform analysis that includes other social media platforms and online forums would provide a more comprehensive understanding of public sentiment and discourse surrounding health care.

Acknowledgments

The authors sincerely thank Dr Rodina Bizri-Baryak for her valuable contributions to data collection, which greatly enhanced the depth and breadth of this study.

Data Availability

All data are available in the main text or Multimedia Appendix 1. Map files can be downloaded from map URLs provided in Multimedia Appendix 1. Original YouTube comments (initial comments and first-level replies) can be accessed through YouTube using the video descriptions provided in Multimedia Appendix 1.

Authors' Contributions

LVI conceptualized the study, curated the data, conducted the formal analysis, created the visualizations, provided supervision, and managed the project administration. LVI and EE collaborated on writing the original draft, methodology, investigation, validation of the findings, and contributed to the writing, review, and editing of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary information on video and comment analysis.
[DOC File , 50822 KB - infodemiology_v5i1e58227_app1.doc]

References

1. Blumenthal D, Gumas E, Shah A. The failing U.S. health system. N Engl J Med 2024 Oct 31;391(17):1566-1568. [doi: 10.1056/NEJMp2410855] [Medline: 39383455]
2. Obama B. United States health care reform: progress to date and next steps. JAMA 2016 Aug 02;316(5):525-532 [FREE Full text] [doi: 10.1001/jama.2016.9797] [Medline: 27400401]
3. Lazarus D. Column: The cure for healthcare? An end to selfishness. Los Angeles Times. 2020 Feb 21. URL: https://www.latimes.com/business/story/2020-02-21/column-healthcare-selfishness [accessed 2025-01-27]
4. AAPD responds to PBS frontline's "dollars and dentists". PR Newswire. 2012 Jun 28. URL: https://tinyurl.com/37vckk65 [accessed 2025-01-27]
5. Guo A. What to watch on Tuesday: ‘Frontline: The Healthcare Divide’ on PBS. The Washington Post. 2021 May 17. URL: https://tinyurl.com/mr34ex65 [accessed 2025-01-27]
6. Goldberg J. Who’s to blame for Trump’s failures? Must be Paul Ryan. Los Angeles Times. 2017 May 27. URL: https://www.latimes.com/opinion/op-ed/la-oe-goldberg-paul-ryan-blame-20170327-story.html [accessed 2025-01-27]
7. Gottfried J. Americans’ social media use. Pew Research Center. 2024 Jan 31. URL: https://www.pewresearch.org/internet/2024/01/31/americans-social-media-use/ [accessed 2025-01-27]
8. Matamoros-Fernández A, Farkas J. Racism, hate speech, and social media: a systematic review and critique. Telev New Media 2021 Jan 22;22(2):205-224 [FREE Full text] [doi: 10.1177/1527476420982230]
9. Uwalaka T. Evaluating military use of social media for political branding during online firestorms: an analysis of the Afghan troops withdrawal. J Polit Mark 2023 Jun 12:1-17 [FREE Full text] [doi: 10.1080/15377857.2023.2222683]
10. Johnen M, Jungblut M, Ziegele M. The digital outcry: what incites participation behavior in an online firestorm? New Media Soc 2017 Nov 29;20(9):3140-3160 [FREE Full text] [doi: 10.1177/1461444817741883]
11. Erzikova E, McLean C. Drowning out the message together: analysis of social media comments on a political sex scandal. J Soc Media Soc 2020;9(1):207-234 [FREE Full text]
12. Pereira HB, Grilo M, de Sousa Fadigas I, Souza Junior CT, do Vale Cunha M, Barreto RS, et al. Systematic review of the “semantic network” definitions. Expert Syst Appl 2022 Dec;210:118455. [doi: 10.1016/j.eswa.2022.118455]
13. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2010 Aug;84(2):523-538 [FREE Full text] [doi: 10.1007/s11192-009-0146-3] [Medline: 20585380]
14. Flis I, van Eck NJ. Framing psychology as a discipline (1950-1999): a large-scale term co-occurrence analysis of scientific literature in psychology. Hist Psychol 2018 Nov;21(4):334-362 [FREE Full text] [doi: 10.1037/hop0000067] [Medline: 28726441]

15. van Eck NJ, Waltman L. Citation-based clustering of publications using CitNetExplorer and VOSviewer. Scientometrics 2017 Feb 27;111(2):1053-1070 [FREE Full text] [doi: 10.1007/s11192-017-2300-7] [Medline: 28490825]
16. Waltman L, van Eck NJ, Noyons EC. A unified approach to mapping and clustering of bibliometric networks. J Informetr 2010 Oct;4(4):629-635 [FREE Full text] [doi: 10.1016/j.joi.2010.07.002]
17. van Eck NJ, Waltman L, Noyons EC, Buter RK. Automatic term identification for bibliometric mapping. Scientometrics 2010 Mar 11;82(3):581-596 [FREE Full text] [doi: 10.1007/s11192-010-0173-0] [Medline: 20234767]
18. van Eck NJ, Waltman L. Text mining and visualization using VOSviewer. arXiv Preprint posted online on September 9, 2011 [FREE Full text]
19. van Eck NJ, Waltman L. VOSviewer manual. Universiteit Leiden. 2023 Oct 31. URL: https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.20.pdf [accessed 2025-01-27]
20. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep 2019 Mar 26;9(1):5233 [FREE Full text] [doi: 10.1038/s41598-019-41695-z] [Medline: 30914743]
21. Vlieger E, Leydesdorff L. Content analysis and the measurement of meaning: the visualization of frames in collections of messages. Public J Semiotics 2011 Jun;III(1):28-50 [FREE Full text] [doi: 10.37693/pjos.2011.3.8830]
22. Holmberg K, Hellsten I. Integrating and differentiating meanings in tweeting about the fifth Intergovernmental Panel on Climate Change (IPCC) report. First Monday 2016;21(9) [FREE Full text] [doi: 10.5210/fm.v21i9.6603]
23. Eriksson-Backa K, Holmberg K, Ek S. Communicating diabetes and diets on Twitter - a semantic content analysis. Int J Netw Virtual Org 2016;16(1):8. [doi: 10.1504/ijnvo.2016.075133]
24. Nerghes A, Kerkhof P, Hellsten I. Early public responses to the Zika-Virus on YouTube: prevalence of and differences between conspiracy theory and informational videos. In: Proceedings of the 10th ACM Conference on Web Science. 2018 Presented at: WebSci '18; May 27-30, 2018; Amsterdam, The Netherlands. [doi: 10.1145/3201064.3201086]
25. Garcia-Rudolph A, Laxe S, Saurí J, Bernabeu Guitart M. Stroke survivors on Twitter: sentiment and topic analysis from a gender perspective. J Med Internet Res 2019 Aug 26;21(8):e14077 [FREE Full text] [doi: 10.2196/14077] [Medline: 31452514]
26. Puerta P, Laguna L, Vidal L, Ares G, Fiszman S, Tárrega A. Co-occurrence networks of Twitter content after manual or automatic processing. A case-study on “gluten-free”. Food Qual Prefer 2020 Dec;86:103993 [FREE Full text] [doi: 10.1016/j.foodqual.2020.103993]
27. Zhao Y, Cheng S, Yu X, Xu H. Chinese public's attention to the COVID-19 epidemic on social media: observational descriptive study. J Med Internet Res 2020 May 04;22(5):e18825 [FREE Full text] [doi: 10.2196/18825] [Medline: 32314976]
28. Bashir S, Bano S, Shueb S, Gul S, Mir AA, Ashraf R, Shakeela, et al. Twitter chirps for Syrian people: sentiment analysis of tweets related to Syria Chemical Attack. Int J Disaster Risk Reduct 2021 Aug;62:102397. [doi: 10.1016/j.ijdrr.2021.102397]
29. Tran HT, Lu SH, Tran HT, Nguyen BV. Social media insights during the COVID-19 pandemic: infodemiology study using big data. JMIR Med Inform 2021 Jul 16;9(7):e27116 [FREE Full text] [doi: 10.2196/27116] [Medline: 34152994]
30. Wonneberger A, Hellsten IR, Jacobs SH. Hashtag activism and the configuration of counterpublics: Dutch animal welfare debates on Twitter. Inf Commun Soc 2020 Feb 04;24(12):1694-1711 [FREE Full text] [doi: 10.1080/1369118x.2020.1720770]
31. Noor S, Guo Y, Shah SH, Fournier-Viger P, Nawaz MS. Analysis of public reactions to the novel Coronavirus (COVID-19) outbreak on Twitter. Kybernetes 2020 Nov 09;50(5):1633-1653 [FREE Full text] [doi: 10.1108/k-05-2020-0258]
32. Henson T, Baryak R, Merrill J, Ivanitskaya L. Term network visualization as a means of detecting attempts to sell illicit fentanyl to social media users. In: Proceedings of the American Public Health Association Annual Meeting and Expo. 2023 Presented at: APHA 2023; November 12-15, 2023; Atlanta, GA URL: https://apha.confex.com/apha/2023/meetingapp.cgi/Paper/536121
33. Henson T. Fentanyl, social media, and health policy: a mixed-methods health policy analysis of public comments on YouTube videos about fentanyl. Central Michigan University. 2024. URL: https://tinyurl.com/mwkfpnp6 [accessed 2025-01-27]
34. Nugroho LA, Wuryani E. Exploring public perceptions of ChatGPT in education through netnography and social media analysis. Gagasan Pendidikan Indonesia 2023 Dec;4(2):66 [FREE Full text] [doi: 10.30870/gpi.v4i2.23164]
35. Xiong Y, Pan Z, Yang L. Did online education exacerbate contract cheating during COVID19 in China? Evidence from Sina Weibo. JITE Res 2023;22:295-309 [FREE Full text] [doi: 10.28945/5181]
36. Frasco JM. A social media discussion of COVID-19 emergency use authorization: a mixed methods study of a YouTube influencer. Central Michigan University. 2023. URL: https://tinyurl.com/yeynd3t4 [accessed 2025-01-27]
37. Bizri-Baryak R. Analyzing the overturn of Roe v. Wade: a term co-occurrence network analysis of YouTube comments. Central Michigan University. 2024. URL: https://tinyurl.com/mp7jnnr9 [accessed 2025-01-24]
38. Ivanitskaya LV, Bacon LL, Erzikova E, Gandy L. Visualizing social media discussions of policy-relevant issues surrounding the opioid crisis in the United States. Popul Med 2023 Apr 26;5(Supplement). [doi: 10.18332/popmed/163845]
39. Media bias rating method. AllSides. URL: https://www.allsides.com/media-bias/media-bias-rating-methods [accessed 2024-03-09]
40. Eve MP. Close Reading with Computers: Textual Scholarship, Computational Formalism, and David Mitchell's Cloud Atlas. Redwood City, CA: Stanford University Press; 2019.

41. Grimmer J, Stewart BM. Text as data: the promise and pitfalls of automatic content analysis methods for political texts. Polit Anal 2017 Jan 04;21(3):267-297. [doi: 10.1093/pan/mps028]
42. VOSviewer online. VOSviewer. URL: http://tinyurl.com/4v6nafs5 [accessed 2025-01-28]
43. Chatterji P, Brandon P, Markowitz S. Job mobility among parents of children with chronic health conditions: early effects of the 2010 Affordable Care Act. J Health Econ 2016 Jul;48:26-43. [doi: 10.1016/j.jhealeco.2016.03.004] [Medline: 27060524]
44. Fruchterman TM, Reingold EM. Graph drawing by force - directed placement. Softw Pract Exp 2006 Oct 30;21(11):1129-1164. [doi: 10.1002/spe.4380211102]
45. Bastian M, Heymann S, Jacomy M. Gephi: an open source software for exploring and manipulating networks. Proc Int AAAI Conf Web Soc Media 2009 Mar 19;3(1):361-362. [doi: 10.1609/icwsm.v3i1.13937]
46. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2022. URL: https://www.R-project.org/ [accessed 2025-01-27]
47. van der Voort H, Klievink A, Arnaboldi M, Meijer A. Rationality and politics of algorithms. Will the promise of big data survive the dynamics of public decision making? Gov Inf Q 2019 Jan;36(1):27-38 [FREE Full text] [doi: 10.1016/j.giq.2018.10.011]
48. Kitchin R. Big Data, new epistemologies and paradigm shifts. Big Data Soc 2014 Apr 01;1(1) [FREE Full text] [doi: 10.1177/2053951714528481]

Abbreviations

ACA: Affordable Care Act
DMV: Department of Motor Vehicles
DNR: do not resuscitate

Edited by T Mackey; submitted 10.03.24; peer-reviewed by M Haupt, C Tong; comments to author 15.08.24; revised version received 10.10.24; accepted 08.12.24; published 11.02.25.

Please cite as:
Ivanitskaya LV, Erzikova E
Visualizing YouTube Commenters' Conceptions of the US Health Care System: Semantic Network Analysis Method for Evidence-Based Policy Making
JMIR Infodemiology 2025;5:e58227
URL: https://infodemiology.jmir.org/2025/1/e58227
doi:10.2196/58227
PMID:39932770

Original Paper

Geosocial Media's Early Warning Capabilities Across US County-Level Political Clusters: Observational Study

Dorian Arifi[1,2], MSc; Bernd Resch[1,2,3], Prof Dr; Mauricio Santillana[4,5], Prof Dr; Weihe Wendy Guan[3], PhD; Steffen Knoblauch[6,7,8], MSc; Sven Lautenbach[6,8], Prof Dr; Thomas Jaenisch[9,10], Prof Dr; Ivonne Morales[10,11], PhD; Clemens Havas[12], Prof Dr

[1]Department of Geoinformatics, University of Salzburg, Salzburg, Austria

[2]Interdisciplinary Transformation University Austria, Linz, Austria

[3]Center for Geographic Analysis (CGA), Harvard University, Cambridge, MA, United States

[4]Machine Intelligence Group for the Betterment of Health and the Environment, Northeastern University, Boston, MA, United States

[5]Department of Epidemiology, Harvard TH Chan School of Public Health, Boston, MA, United States

[6]GIScience Research Group, Heidelberg University, Heidelberg, Germany

[7]Interdisciplinary Centre of Scientific Computing (IWR), Heidelberg University, Heidelberg, Germany

[8]HeiGIT (Heidelberg Institute for Geoinformation Technology) GbmH, Heidelberg, Germany

[9]Center for Global Health, Colorado School of Public Health, Aurora, CO, United States

[10]Heidelberg Institute of Global Health, Heidelberg University Hospital, Heidelberg, Germany

[11]Department of Infectious Disease and Tropical Medicine, Heidelberg University Hospital, Heidelberg, Germany

[12]Salzburg University of Applied Sciences, Puch/Salzburg, Austria

Corresponding Author:
Dorian Arifi, MSc
Department of Geoinformatics
University of Salzburg
Kapitelgasse 4/6
Salzburg, 5020
Austria
Phone: 43 662 80440
Email: dorian.arifi@plus.ac.at

Abstract

Background: The novel coronavirus disease (COVID-19) sparked significant health concerns worldwide, prompting policy makers and health care experts to implement nonpharmaceutical public health interventions, such as stay-at-home orders and mask mandates, to slow the spread of the virus. While these interventions proved essential in controlling transmission, they also caused substantial economic and societal costs and should therefore be used strategically, particularly when disease activity is on the rise. In this context, geosocial media posts (posts with an explicit georeference) have been shown to provide a promising tool for anticipating moments of potential health care crises. However, previous studies on the early warning capabilities of geosocial media data have largely been constrained by coarse spatial resolutions or short temporal scopes, with limited understanding of how local political beliefs may influence these capabilities.

Objective: This study aimed to assess how the epidemiological early warning capabilities of geosocial media posts for COVID-19 vary over time and across US counties with differing political beliefs.

Methods: We classified US counties into 3 political clusters, democrat, republican, and swing counties, based on voting data from the last 6 federal election cycles. In these clusters, we analyzed the early warning capabilities of geosocial media posts across 6 consecutive COVID-19 waves (February 2020-April 2022). We specifically examined the temporal lag between geosocial media signals and surges in COVID-19 cases, measuring both the number of days by which the geosocial media signals preceded the surges in COVID-19 cases (temporal lag) and the correlation between their respective time series.

Results: The early warning capabilities of geosocial media data differed across political clusters and COVID-19 waves. On average, geosocial media posts preceded COVID-19 cases by 21 days in republican counties compared with 14.6 days in democrat counties and 24.2 days in swing counties. In general, geosocial media posts were preceding COVID-19 cases in 5 out of 6 waves across all political clusters. However, we observed a decrease over time in the number of days that posts preceded COVID-19 cases, particularly in democrat and republican counties. Furthermore, a decline in signal strength and the impact of trending topics presented challenges for the reliability of the early warning signals.

Conclusions: This study provides valuable insights into the strengths and limitations of geosocial media data as an epidemiological early warning tool, particularly highlighting how they can change across county-level political clusters. Thus, these findings indicate that future geosocial media based epidemiological early warning systems might benefit from accounting for political beliefs. In addition, the impact of declining geosocial media signal strength over time and the role of trending topics for signal reliability in early warning systems need to be assessed in future research.

(JMIR Infodemiology 2025;5:e58539) doi:10.2196/58539

KEYWORDS

spatiotemporal epidemiology; geo-social media data; digital disease surveillance; political polarization; epidemiological early warning; digital early warning

Introduction

On March 12, 2020, the World Health Organization (WHO) declared the novel coronavirus disease COVID-19 a pandemic [1]. Its high infectiousness and severity posed a great threat to large populations worldwide, ultimately causing an estimated 15.9 million pandemic-related deaths [2], challenging health care professionals, hospitals, and authorities alike. Thus, decision makers around the world sought to unravel and predict the spreading dynamics of this novel coronavirus. Consequently, researchers explored various ways of adjusting and improving existing epidemiological early warning systems, with complementary internet-based data sources being one such method to better monitor and anticipate how this new disease would affect different geographies around the world [3-5].

Multiple studies have already emphasized the role of geosocial media data in improving early warning of epidemiological phenomena. For instance, geosocial media data were used to improve real-time reporting on diseases like Zika and Ebola [6] or to enhance the prediction of dengue fever [7]. Accordingly, various recent examples further emphasize the ability of geosocial media data for real-time surveillance and early warning in the context of COVID-19 [8,9]. In this regard, Kogan et al [10] observed that in the beginning of the pandemic, increases in geosocial media activity, among other digital data sources, preceded surges in COVID-19 cases by 2 to 3 weeks on state level. Similarly, Zhang et al [11] used geosocial media posts in a linear regression model to predict COVID-19 signals on state-level. Yet, an increasing trend in epidemiological analysis focuses on ever finer spatial scales in the hopes of gaining a more distinct understanding of infection patterns. In this regard, Stolerman et al [12] investigated the value of X posts (formerly known as Twitter) for COVID-19 early warning on a representative subset of US counties. However, the authors only investigated a comparably small sample of counties (n=97), raising questions with respect to the generalizability of the presented results. Thus, in this study, we extended their investigation on the early warning capabilities of geosocial media data to all US counties.

Furthermore, geosocial media data garnered notable attention across various fields to answer research questions related to mental health or public attitudes, during the COVID-19 pandemic [13]. For instance, researchers investigated how language in Reddit posts reflected real-world pandemic-driven events like lockdowns, revealing significant psychological shifts among users which coincided with tendencies toward decreased analytical thinking [14]. Similarly, Swain et al [15] developed a machine learning model leveraging geosocial media data to predict disruptions in mental well-being caused by the COVID-19 pandemic. Beyond that, researchers explored geosocial media users' attitudes and concerns toward COVID-19 vaccines for the United States and the United Kingdom [16]. They observed that geosocial media derived results correlated broadly with nationwide surveys. In essence, the previous results suggest that geosocial media exchange during the COVID-19 pandemic was likely influenced by real-world public attitudes and even users' mental health. Similarly, a variety of studies indicate that the language used and the topics of interest of geosocial media users vary based on political beliefs [17-19]. This further supports our underlying assumption that differences in political beliefs are likely to be reflected in geosocial media behavior, which could, in turn, correspond to differences in geosocial media's early warning capabilities for COVID-19 cases.

However, even before the surge of the COVID-19 pandemic, researchers observed the emergence of echo chambers when analyzing pro and antivaccination attitudes on Facebook (Meta), which in their opinion might have caused further polarization [20]. In this regard, Howard et al [21] found that X was particularly prone to misinformation and polarizing content compared with professionally produced news during the 2016 presidential election. They even found more misinformation being prevalent in swing states. Such spread of misinformation and emerging political polarization on geosocial media should be of further concern for health experts and policy makers. In particular, since many researchers illustrated that diverging political beliefs can not only influence exchange on geosocial media [17-19], but also real-world individual behavior such as vaccine up-take [22] or the usage of nonpharmaceutical interventions such as mask wearing [23]. This is in line with previous findings [24], which highlight significant variation between individuals with different political beliefs with respect to self-estimated COVID-19 risks, self-reported adherence to COVID-19 health care measures, and expectations on the future course of the pandemic. In addition, researchers observed that US counties that voted in favor of the republican presidential candidate in the 2016 election, experienced up to 3 times higher mortality due to COVID-19 during the winter of 2020 [25].

Hence, in essence it can be assumed that individuals may respond differently on geosocial media to a swiftly politicized epidemic event like the COVID-19 pandemic [26], corresponding to their political beliefs. Evidence further suggests that differences in political beliefs do not only influence online and offline behavior, but they might indeed coincide with higher COVID-19 cases and death rates [25,27,28]. In summary, these results highlight the need to understand and adjust geosocial media based early warning systems with respect to political beliefs. Thus, within the scope of this paper, we seek to answer the following 2 research questions with a particular focus on geosocial media posts:

1. How do the early warning capabilities of geosocial media data change across consecutive epidemiological waves of COVID-19 cases?
2. What differences across US county-level political clusters can be observed with respect to geosocial media's early warning capabilities for COVID-19 cases?

To explore the early warning capabilities of geosocial media data, we determined the correlation between geosocial media posts and COVID-19 cases and the number of days by which signals in geosocial media data preceded actual COVID-19 cases (temporal lag). Furthermore, we specifically examined the temporal lag and the correlation in the context of political clusters based on US county voting data and over the course of 6 consecutive waves of COVID-19 cases.

Methods

Data Collection

We used 2 main data sources in this study. First, we gathered official data on confirmed COVID-19 cases in the United States and we obtained geolocated posts (Tweets) from the geosocial media network X. The time frame for which we collected our data ranges from February 28, 2020, the beginning of the pandemic in the United States, to April 27, 2022, which denotes the end of the first major Omicron wave that began in November 2021 [29]. This time frame covers the main COVID-19 waves, time periods before and after the availability of vaccines, and was selected based on retrospective knowledge on the course of the pandemic. The contiguous United States was chosen as our study area. Furthermore, to gain a more refined understanding of the underlying spatial patterns, we decided to use US counties as our finest spatial analysis resolution, on which we identified politically similar clusters, advancing previous research that was mostly performed on national or state levels.

COVID-19 Case Data

We downloaded officially confirmed COVID-19 cases for the United States in csv format from the not-for-profit public data aggregator USAFacts [30]. The COVID-19 cases csv file contained daily cumulated COVID-19 cases, which we transformed into daily incidence data. In addition, we applied a 14-day moving average to account for possible reporting delays and differing update cycles across states.

Geosocial Media Data

Furthermore, we collected geolocated posts from the geosocial media network X through their official application programming interfaces (APIs) during our investigation time frame [10,12], when academic access for researchers was still available. In particular, we used the Twitter REST and Streaming API access points to gather about 727 million geosocial media posts. The REST API allowed us to retrieve posts from the previous 7 days, with a limit of 450 requests per 15-minute window. In contrast, the Streaming API provided a continuous, real-time stream of posts. For both API endpoints we applied filters to capture only posts containing a geolocation. Thus, each collected geosocial media post includes a geolocation, which can either be the Global Navigation Satellite System position of the device through which the post was shared, or a user-defined location. Furthermore, locations can consist of polygons (eg, city, state level polygons) or point locations. We excluded geosocial media posts with polygon or point geometries that were not located within the county-level geometries, which left us with 242 million posts.

Next, to obtain geosocial media posts that are relevant to the analysis of COVID-19, we performed keyword filtering on the remaining 242 million posts located within county geometries. Therefore, we defined keywords based on the knowledge of geosocial media and health experts, with the goal to properly capture geosocial media trends relevant to the COVID-19 pandemic (Textbox 1). For some keywords only their word stem was used to allow for different variations of the word to be detected.

Textbox 1. Keywords used for relevant post extraction.

COVID-19 keywords:

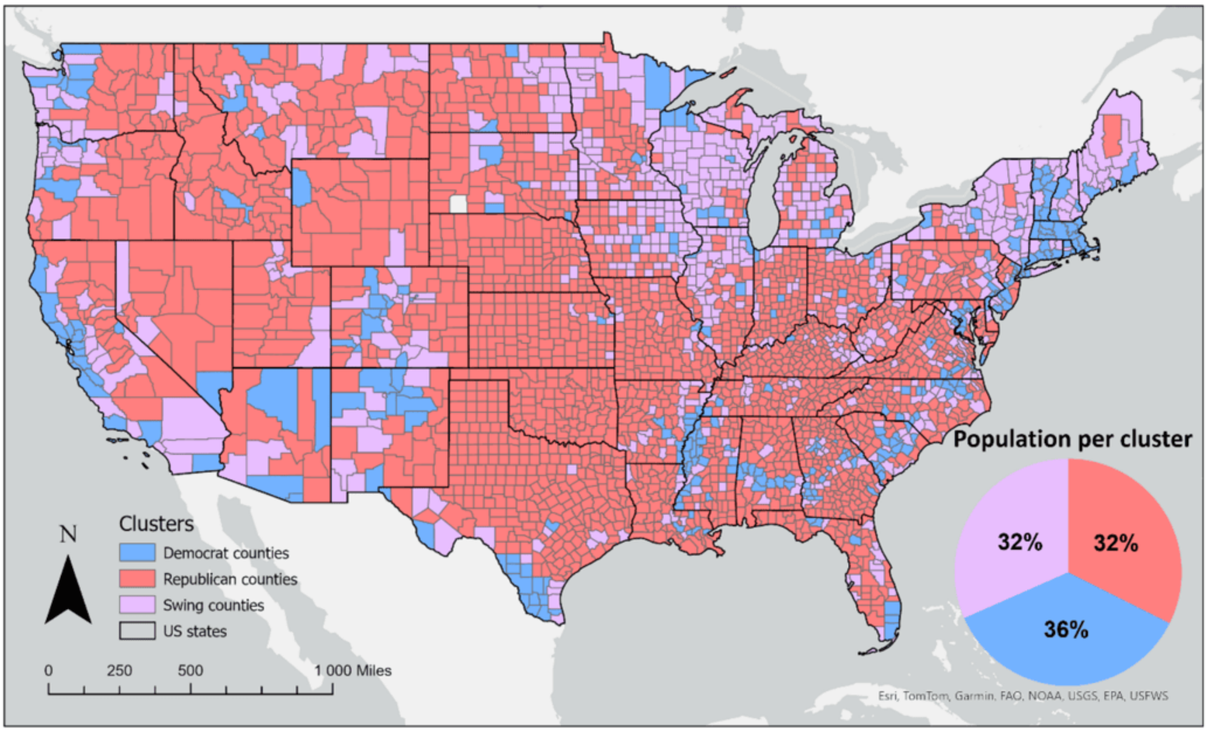
covid, corona, sarscov, sars-cov, sars, epidemic, pandemic, influenza, virus, viral, infect, spread, 2019-ncov, Delta variant, Omicron, H1N1, H3N2, Wuhan, sickness, transmission, contagio, Illness, outbreak, super spread, incubation, quarantine, lockdown, vaccin, fever, cough, headache, fatigue, body aches, loss of taste, loss of smell, no smell, no taste, respirator, face mask, masks.

After the keyword extraction, the posts were aggregated on US county-level and a 14-day moving average was applied. Finally, to cope with differing amounts of geosocial media posts over time and space, we normalized the amount of relevant filtered geosocial media posts over the amount of all geosocial media posts on county level. In the remainder of this study, we solely used this ratio, that is, the proportion of relevant posts over all posts per county. This allows us to account for spatially clustered population and post density. In total, the semantic filtering procedure left us with 3.3 million relevant posts.

Political Clusters

To examine the differences between the various political beliefs, we based our analysis on voting data from the last 6 US presidential elections. The voting data were obtained from the Harvard Dataverse [31]. We classified US counties into 3 different clusters depending on their historical vote share for either the republican or the democrat party. In the political sciences literature, swing states are traditionally defined through a variety of quantitative and qualitative indicators. However, most of these definitions such as the bellwether status of a state [32], or it being perceived as a battleground [32], are not directly transferable to county-level analysis. Thus, we decided to base the classification into republican, democrat, or swing county clusters, on the so-called flippability of a county [32]. We chose to assess the flippability of a county on its last 6 federal election cycles. Concretely, we classified a county as belonging to a specific party, if said party had won at least 5 consecutive elections in the last 6 elections cycles. All other counties were considered as flipping between political parties and thus classified as swing counties. This division yielded political clusters, each of which representing approximately one third of the US population (Figure 1).

Figure 1. Geospatial distribution of political belief clusters on county level based on the last 6 election cycles.

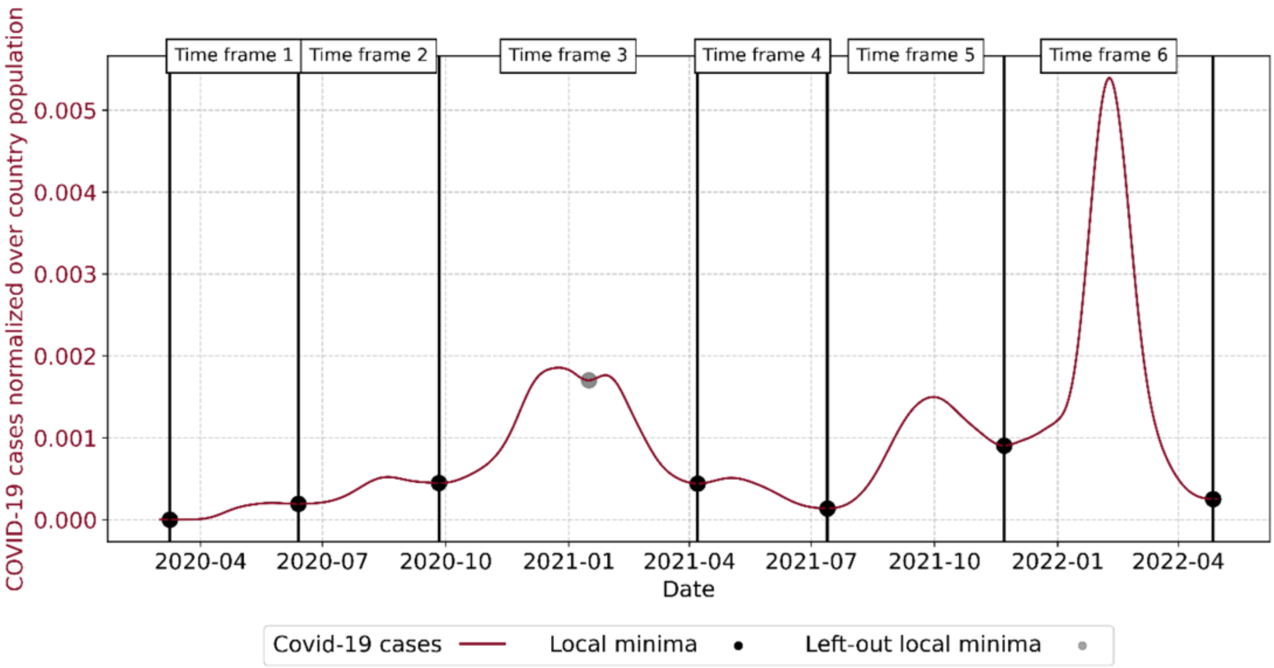


Defining COVID-19 Waves

We split the COVID-19 cases time series into smaller time frames, to capture individual epidemiological waves. However, there exist multiple approaches to define epidemic waves ranging from statistical methods using, for instance, exponential growth [10,33] or the effective reproduction number R [12,34]. In contrast, other authors tried to identify statistics and guiding principles on the duration of COVID-19 waves based on empirical data [35]. Nevertheless, all these approaches are based on strong assumptions and subjective definitions on what thresholds characterize an epidemic wave. Thus, similarly to [35], we based our definition of COVID-19 waves on a rule-based approach using the local minima on a 21-day moving average of the COVID-19 cases, which was informed through retrospective knowledge on the course of the pandemic.

We defined these time frames based on COVID-19 cases for the entire United States, rather than defining them individually for each political cluster. Furthermore, our procedure yielded 7 different time frames (Figure 2). Nonetheless, these 7 time frames did not accurately reflect all epidemic waves. In particular, the wave ranging roughly from October 2020 to April 2021, was split into 2. As a result, we decided to combine the original time frames 3 and 4 into 1 epidemic wave, which left us with 6 epidemic waves in total. This decision enabled us to capture the epidemic waves more accurately (Figure 2).

Figure 2. COVID-19 case waves for the entire US primarily defined through local minima.



Early Warning Capabilities

Finally, we quantified the early warning capabilities separately for each of the epidemic waves. We defined early warning capabilities twofold: (1) as the Pearson correlation between the time series of COVID-19 related geosocial media posts and COVID-19 cases, and (2) the number of days by which geosocial media posts preceded COVID-19 cases. However, the more important measure for early warning is the correlation between the 2 time series. Put differently, this means that if the temporal lag is high, however a correlation close to zero is present, it is obviously not reasonable to attribute any early warning capabilities to geosocial media data.

Furthermore, to identify the maximal correlation and the corresponding temporal lag, we shifted the geosocial media posts time series between 7 and 42 days into the future to determine the highest possible early warning capabilities. This procedure is repeated for each individual political cluster and epidemic wave, respectively. The decision to investigate a temporal lag between 7 and 42 days into the future was based on previous results [12], in which an early warning model, using, among others, geosocial media data, was able to predict COVID-19 cases between 1 and 6 weeks in advance.

Ethical Considerations

The study was carried out in accordance with the Declaration of Helsinki and with the ethical regulations in place at the Paris Lodron University of Salzburg, and complies with the General Data Protection Regulation legislation of the European Union. We only used publicly available data, which were collected in accordance with the terms of service of the respective geosocial media platform X at the time of data collection. Furthermore, no identifiable information was revealed in this study. Specifically, the user-provided geographic locations and semantic content were spatially aggregated to ensure user privacy and anonymity. Thus, we did not need to seek ethical approval from our institution for this study.

Results

Democrat Counties

Figure 3 depicts the Pearson correlation for different temporal lags between the time series of COVID-19 cases and geosocial media posts in democrat counties. In particular, the y-axis represents the individual waves of COVID-19 cases as introduced in Figure 2, while the x-axis denotes the number of days the posts time series was shifted into the future. The coloring of individual windows reflects the Pearson correlation between COVID-19 cases and the shifted posts time series. Furthermore, Figure 4 illustrates the corresponding COVID-19 cases, the post time series and the post time series shifted by the correlation maximizing temporal lag for each individual epidemic wave.

Figure 3. Depicting the Pearson correlation between COVID-19 cases and geosocial media post time series for each epidemiological wave when stepwise shifting the geosocial media post time series into the future for democrat counties.

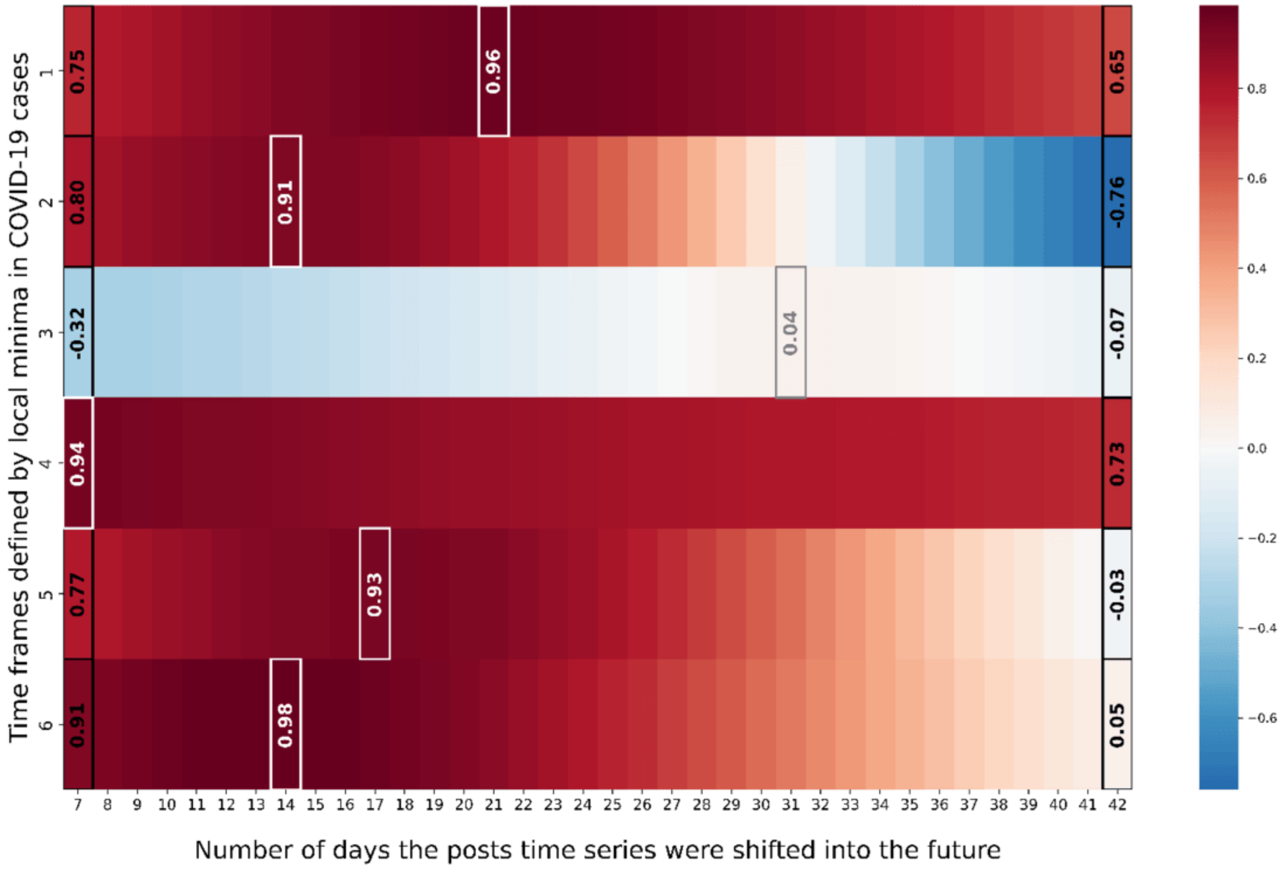
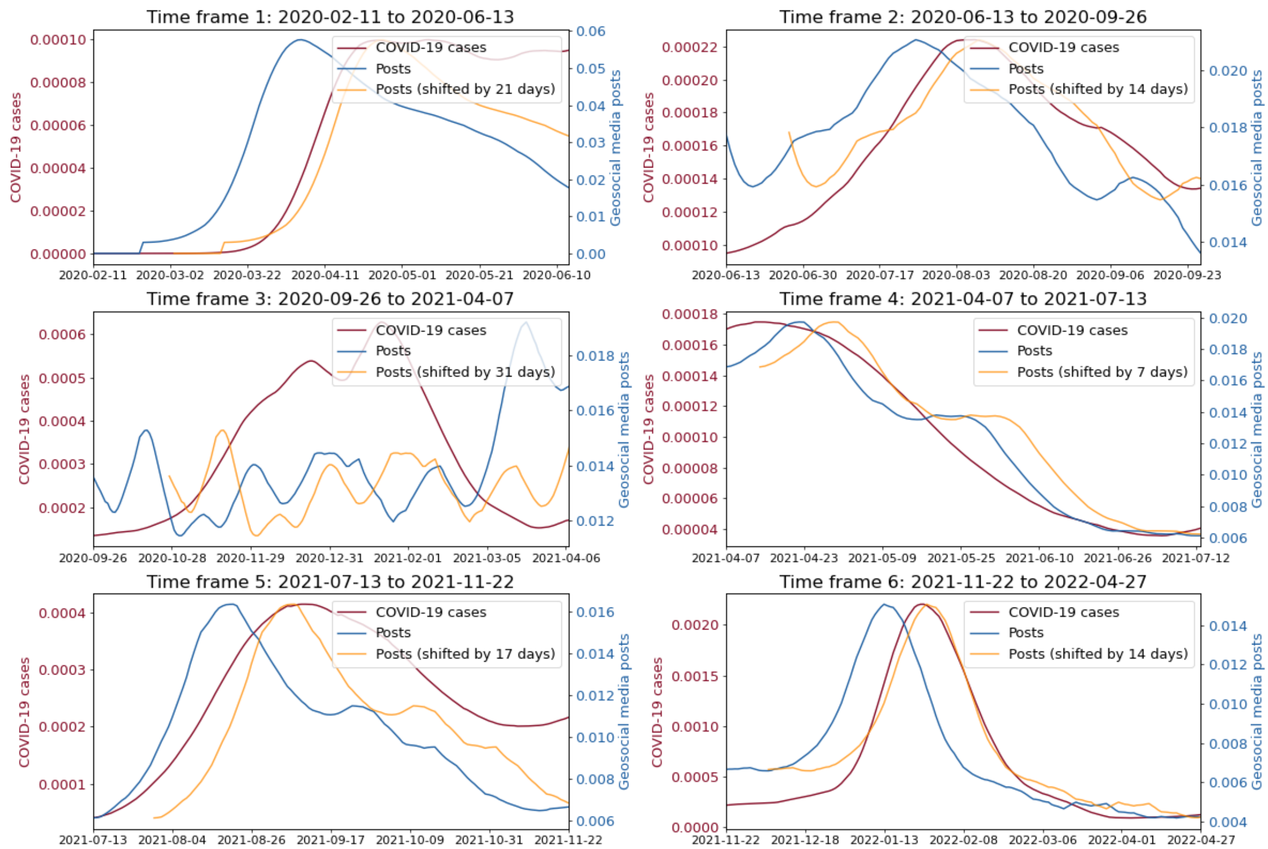


Figure 4. Depicting COVID-19 cases and the geosocial media posts time series and the shifted geosocial media posts time series across individual epidemic waves for democrat counties.



The results for democrat counties in Figure 3 indicate the highest Pearson correlations between posts and COVID-19 cases time series in 5 out of 6 epidemic waves, for a shift of 7 to 21 days (time frames 1, 2 and 4-6). For the same 5-time frames, the Pearson correlations ranged between 0.91 to 0.98. Furthermore, Figures 3 and 4 suggest that only for time frames 1, 2 and 4-6, geosocial media data exhibited actual early warning capabilities. For these time frames, signals in COVID-19 cases were clearly preceded by signals in X data, while for time frame 3 no clear early warning signal in geosocial media data was apparent. Nevertheless, in the beginning of the pandemic (time frames 1 and 2) geosocial media posts showcased a clear increase up to 21 (time frame 1) and 14 days (time frame 2) ahead increases in COVID-19 infections, with Pearson correlations of 0.96 and 0.91. In addition, the COVID-19 wave from mid of July 2021 to the end of November 2021 (time frame 5) was reflected in geosocial media posts up to 17 days earlier than an increase in COVID-19 cases, with a Pearson correlation of 0.93. Also, the Omicron wave (time frame 6) starting in mid of November 2021 [29] was accurately reflected 14 days in advance in the geosocial media time series (Pearson correlation of 0.98). Beyond that, Figure 4 clearly illustrates that the ratio of geosocial media posts related to COVID-19 decreased significantly over the course of the pandemic. Specifically, the percentage of relevant geosocial media posts gradually decreased from 5.7% at its peak in the first time frame, to 1.5% in the last time frame.

Republican Counties

Figure 5 illustrates for the republican counties that in 5 out of 6 time frames the post time series exhibited the highest Pearson correlation with the COVID-19 cases 7 to 38 days ahead of time (time frames 1, 2, and 4-6). Furthermore, for these time frames the Pearson correlations between posts shifted 7 to 38 days into the future and COVID-19 cases were between 0.74 and 0.97. Furthermore, Figure 6 showcases that for republican counties, early warning signals in geosocial media posts could be observed for time frames 1, 2 and 4-6. Similarly to the democrat county cluster, the COVID-19 cases wave in time frame 3 was not captured in advance by the geosocial media time series. The fact that all time frames besides time frame 3, lend themselves for early warning is also consistent with the results for the democrat counties. Furthermore, it appears that in the republican counties, geosocial media data preceded COVID-19 cases time series a few days more in advance. On average over all 5 time frames for which we attest early warning capabilities (time frames 1, 2, and 4-6), the mean temporal lag in democrat counties is 14.6 days (average correlation 0.94) and for 21 days republican counties (average correlation 0.9). Furthermore, it appears that the ratio of relevant posts decreased over time for republican counties from roughly about 5.3% to 0.9%.

Figure 5. Depicting the Pearson correlation between COVID-19 cases and geosocial media post time series for each epidemiological wave when stepwise shifting the geosocial media post time series into the future for republican counties.

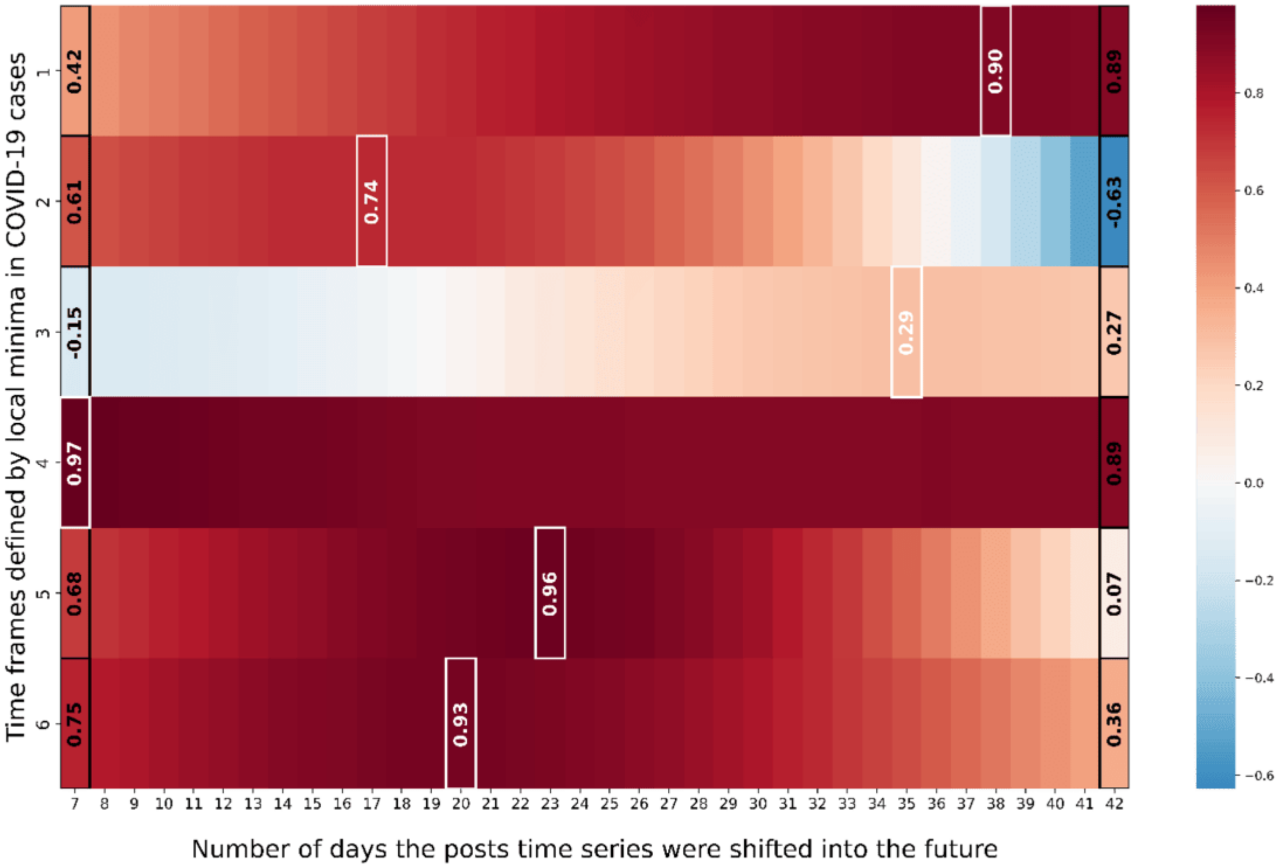
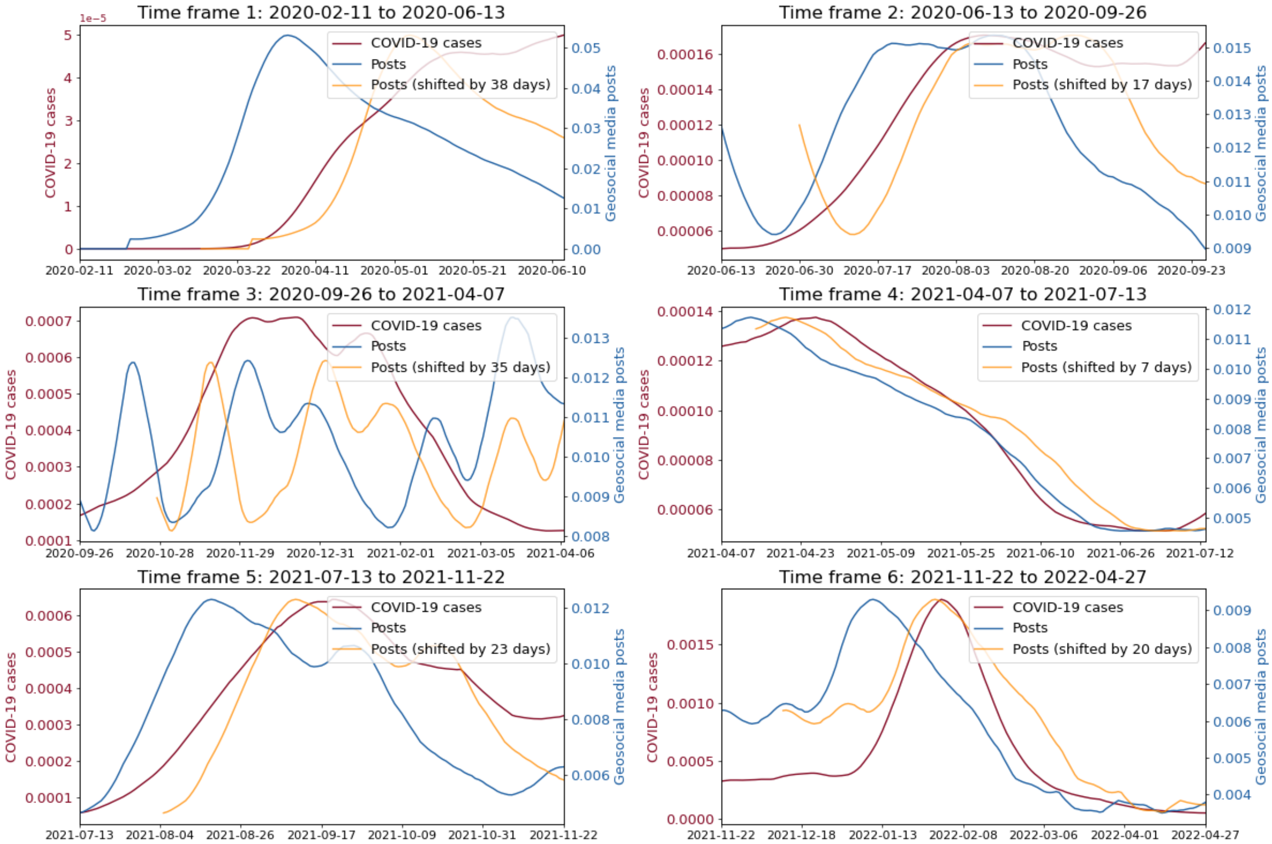


Figure 6. Depicting COVID-19 cases and the geosocial media posts time series and the shifted geosocial media posts time series across individual epidemic waves for republican counties.



Swing Counties

Figure 7 illustrates for swing counties that shifting the posts time series between 7 and 37 days into the future achieved the highest correlation for all time frames. Furthermore, for all time frames the maximal Pearson correlations between geosocial media posts and COVID-19 cases ranged between 0.52 and 0.96. Beyond that, Figure 8 shows that the time frames 1, 2 and 4-6 exhibited clear early warning signals in geosocial media data ahead increases in COVID-19 cases. Similarly to the republican and democrat counties, the COVID-19 wave in time frame 3 was not clearly captured in advance by geosocial media data. However, similar, to republican counties, Figure 8 showcases for swing counties that there actually existed a signal in geosocial media data which is in line with the COVID-19 data in time frame 3. Nevertheless, the actual early warning capabilities are still limited due to noise in the signal which coincides with the COVID-19 infection of former President Donald Trump. Overall, the posts time series preceded COVID-19 cases in swing counties across all time frames, excluding the third, on average by 24.2 days. Also, the intensity with which geosocial media data appears to precede COVID-19 waves clearly decreased for swing counties over the course of the pandemic (from 5.6% to 1.1%).

Figure 7. Depicting the Pearson correlation between COVID-19 cases and geosocial media post time series for each epidemiological wave when stepwise shifting the geosocial media post time series into the future for swing counties.

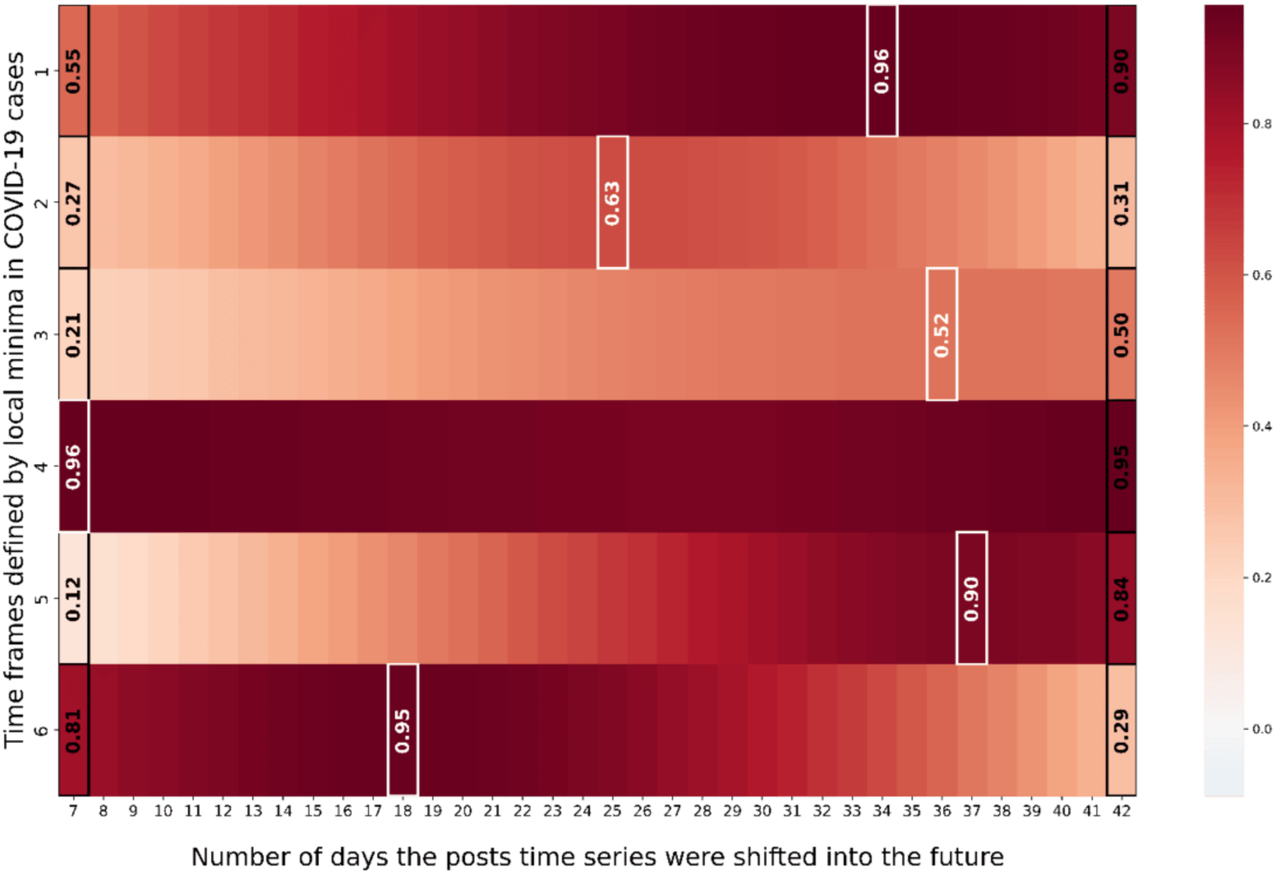
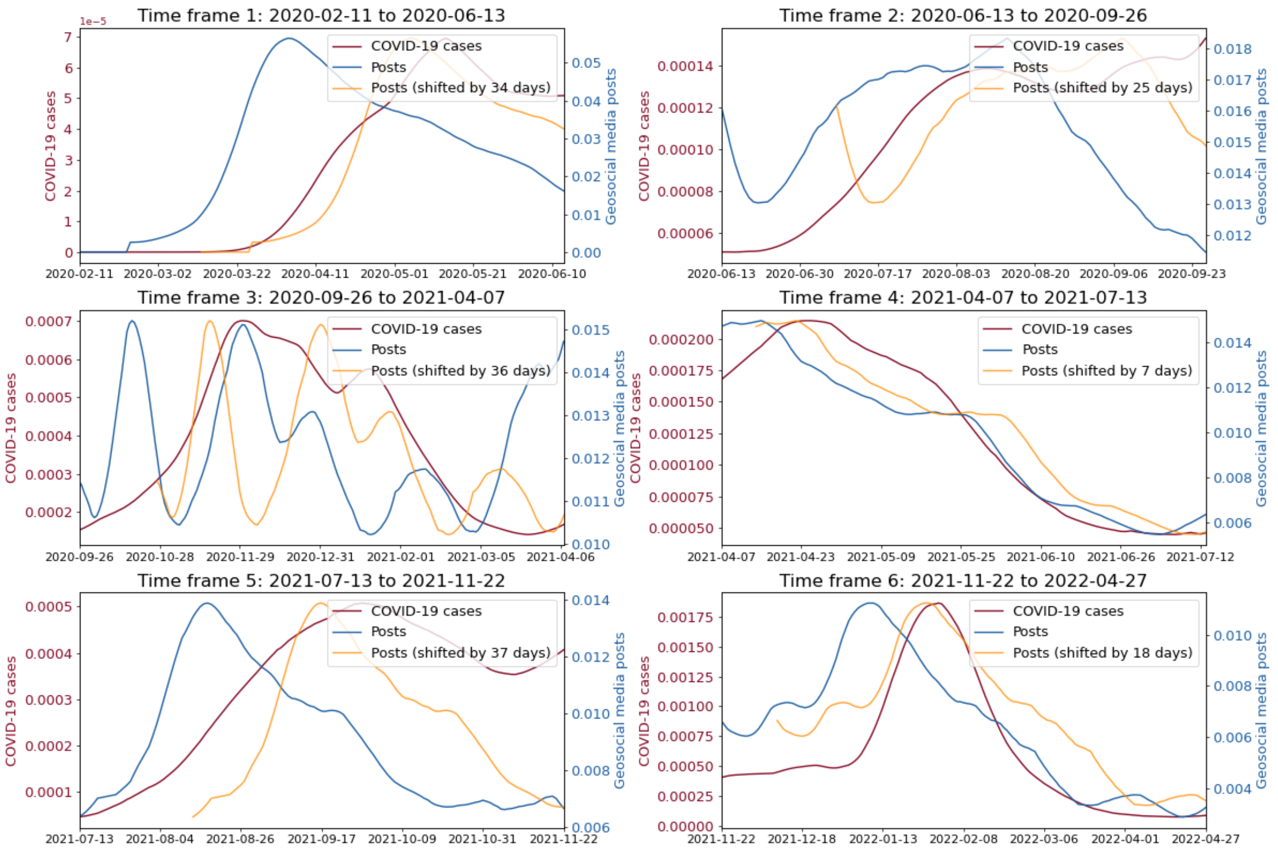


Figure 8. Depicting COVID-19 cases and the geosocial media posts time series and the shifted geosocial media posts time series across individual epidemic waves for swing counties.



Discussion

Principal Findings

The results of this study highlight how a deeper understanding of the relationship between COVID-19–related geosocial media data and confirmed COVID-19 cases, across politically distinct geographies, may help improve epidemiological early warning systems. Specifically, our analysis confirmed and expanded previous findings on the use of geosocial media posts as early indicators of disease activity [8-10,12]. However, we observed strong differences in the early warning capability of geosocial media data across different epidemiological waves. For example, geosocial media data were unable to reliably anticipate the third major COVID-19 wave, September 2020 to April 10, 2021 (time frame 3), across all political clusters. After significantly high COVID-19–related engagement on geosocial media in the first wave, it appears that the geosocial media signal lost some of its sensitivity in the third wave. The only event clearly detectable in COVID-19–related geosocial media posts in the third time frame is the COVID-19 infection of the former President Donald Trump in October 2020. The significance of this event might have reduced the sensitivity of the geosocial media users toward an increase in COVID-19 symptoms and infections. The reaction signal to this event was particularly visible in the republican and swing county clusters, while the democrat counties only registered a minor increase in geosocial media posts coinciding with the COVID-19 infection of President Trump. This further highlights how susceptible geosocial media data can be to politically charged trending topics and how these topics of interest might differ across political clusters. This is also in line with previous findings that the topics geosocial media users engage with and the language they use can differ depending on political beliefs [17-19]. Thus, we hypothesize that it might be key to identify different sets of keywords related to political beliefs and resulting trending topics, to capture geosocial media signals more accurately across political clusters. Therefore, future research should explore the influence of different geosocial media topics on early warning capabilities across political clusters and how such topics might change over time.

Furthermore, the findings of this study illustrate differences in the early warning capabilities of geosocial media posts for COVID-19 cases across counties with diverging political beliefs. This is particularly true for the number of days that geosocial media posts precede COVID-19 cases (temporal lag) and the Pearson correlation between these 2 time series for republican and democrat counties. For instance, geosocial media posts appear to anticipate COVID-19 cases in republican counties (21 days) on average 6.4 days earlier than in democrat counties (14.6 days). This difference in temporal lag might partly be caused by varying population densities between democrat and republican counties. In less densely populated republican counties [36], infection transmission might be slower [37], which could lead to a higher temporal lag between the onset of COVID-19 symptoms being observed and shared on geosocial media, to the eventual peak of infections in that region. However, it remains beyond the scope of this study to substantiate the actual underlying mechanisms which might cause these observed differences in early warning capability between political clusters. Despite that, the results of this study clearly emphasize the need to account for political beliefs in epidemiological early warning systems using geosocial media data. Nevertheless, the precise methodology to integrate political beliefs into real-time geosocial media-based early warning models remains the subject of future research.

The psychological effects of public health measures, such as lockdowns, might offer another explanation for the observed differences in early warning capabilities of geosocial media data across political clusters. These effects may be connected to the fact that public health measures were implemented and suspended at different points in time across political administrative areas. In this regard, Pettersen et al [38] associated more stringent public health and quarantine measures with increased mental distress in adults in Norway. Similarly, Ferwana and Varshney [39] observed a significant increase in visits to mental health facilities during the 2020 lockdown periods in the United States. While Ashokkumar and Pennebaker [14] even reported drops in analytical thinking and shifts in the emotional states of Reddit users coinciding with the start of lockdowns. Hence, it might be the case that the varying timing of public health measures across political regions caused various psychological effects, manifesting in changes of geosocial media behavior. However, our findings do not sufficiently verify this hypothesis. Although numerous studies have explored the psychological effects of public health measures, future research should focus on how these effects might influence the early warning capabilities of geosocial media data across the political spectrum.

In addition, we also found a clear decrease in the number of days with which geosocial media posts preceded COVID-19 cases and in the strength of the geosocial media post signal over time. Interestingly, yet to be explained, the decrease in temporal lag appears to be less pronounced in republican and swing counties. Nonetheless, this overall phenomenon might be caused by some sort of geosocial media and emotional COVID-19 fatigue. The association between self-reported depression symptoms and geosocial media usage [40], alongside potential factors contributing to social media fatigue [41-43] have already been explored in the context of the COVID-19 pandemic. For instance, recent findings by Li et al [43] indicate a direct relationship between social media overload during the COVID-19 pandemic and increased anxiety. Similarly, Sun and Lee [44] observe that COVID-19 information overload on social media directly contributes to fatigue toward pandemic related messages. Nevertheless, it remains beyond the scope of this study to substantiate whether the observed decreasing strength of the geosocial media post signal and temporal lag are robust and attributable to some form of geosocial media or COVID-19 fatigue. However, based on our observations, we advise caution, as the epidemiological early warning capabilities of geosocial media appear to change over time and depending on prevailing political beliefs. In this regard, it remains the task of future research to develop geosocial media-based early warning approaches, which can account for decreasing signal strength over time.

Furthermore, Howard et al [21] observed varying levels of misinformation and thus topics of interest, across states with different political beliefs. Interestingly, they found the highest rates of misinformation occurring in swing states. This is particularly noteworthy, as we found geosocial media data to be highly capable for early epidemiological warning in swing counties. Specifically, the average temporal lag of 24.2 days over all time frames in which we observed the highest early warning capabilities for swing counties, while mostly achieving high correlations (average correlation over all time frames with early warning capabilities 0.88). Thus, concluding from Howard et al [21] and our findings, it appears that it might not be the quality or factual correctness of the shared information on geosocial media that warrants its value for early warning purposes. Nevertheless, future research needs to further validate these findings in the context of different countries and their political ramifications as they might influence the relevance of shared information quality and factual correctness for epidemiological early warning capability.

Data and Methods

We acknowledge that using a simple linear correlation measure might not always reflect the actual similarity between time series accurately. However, in preliminary analysis we also used different nonlinear correlation measures, which yielded only neglectable differences in the actual results. In addition, we also tested more advanced time series matching algorithms such as dynamic time warping [45], the Fréchet distance [46], or mutual information [47]. Nevertheless, neither nonlinear correlation measures nor more advanced comparison algorithms outperformed conventional linear correlation measures for most of our analyses. We evaluated the performance of these different methods in their ability to match the peaks and onsets of geosocial media signals and COVID-19 cases. Nonetheless, we acknowledge that the alignment of peaks and onsets is not always feasible, as the time it takes from the onset to the peak may vary between geosocial media signals and COVID-19 cases. As a result, for some epidemic waves the determined temporal lag might not reflect the actual real-world early warning capabilities of geosocial media data. Despite that, our main objective in this study was not to assess the exact temporal lag and correlations, but rather to provide an algorithmic way to compare the early warning capabilities of geosocial media data across political clusters.

In addition, there is a need to discuss the definition of epidemiological waves based on COVID-19 cases of the entire United States as one might argue that this decision might potentially have caused the observed variations in the number of days and the correlation between the geosocial media and the COVID-19 cases time series. The reason for this is that the COVID-19 waves can have different starting points and intensities across states [48] and as our results show also across political clusters (Figures 4, 6, and 8). Therefore, it might appear reasonable to assume that variation in the starting points and intensities caused the underlying observed differences in temporal lag and correlation between geosocial media posts and COVID-19 cases across political clusters. However, upon testing this hypothesis by defining COVID-19 waves individually for each political cluster, the fundamental results of our study remained unchanged. Although minor discrepancies were present in the temporal lag (primarily ranging from 1-2 days) and the correlations between COVID-19 cases and geosocial media posts, their differences persisted across political clusters and time frames in a similar manner. For example, republican counties still exhibited on average a higher temporal lag than democratic counties and the decrease in geosocial media signals was also still prevalent across political clusters.

In addition, it is important to consider the choice of keywords used for our analysis, as they strongly influence the observed results. One might argue that some keywords relevant to the discourse related to the COVID-19 pandemic were left out. However, in this analysis we mainly focused on gathering less polarized keywords, topics, and hashtags. The reason for this is that certain words, topics and hashtags were predominantly used by 1 political faction [17,18], which might indeed introduce bias into the final comparison between early warning capabilities across political clusters from the start. Concretely, keywords used predominantly in republican counties and less in democrat counties might directly influence differences in early warning capability across political clusters. Therefore, we decided to use a condensed set of keywords, which was to our knowledge mostly not inherently politically charged or biased.

Furthermore, we acknowledge that some keywords which we used in the semantic filtering process of the geosocial media posts, might not be only COVID-19 specific. However, we argue that for most words there exists a baseline signal of how often these words are being used. Therefore, our underlying assumption is that a real-world epidemiological event causes a significant spike in the usage of relevant keywords. Indeed, our results confirmed this assumption. We observed a baseline fluctuation in geosocial media posts and significant spikes in filtered posts, which in most cases preceded COVID-19 cases.

We also tried to improve the semantic filtering by leveraging machine learning approaches such as BERTopic or Latent Dirichlet Allocation [49,50]. However, due to performance issues with our large dataset (600+ GB) and based on the insufficient results for subsample experiments, we decided to stick to traditional keyword filtering. Nevertheless, in future work large language models [51] might be a possibility to improve the process of identifying relevant geosocial media posts.

Limitations

The main limitation of this study stems from its retrospective nature. Our findings, while insightful for the past pandemic, may not be directly transferable to future epidemiological events. This limitation is partly due to the unpredictable nature of political polarization. Specifically, it is inherently difficult to predict whether a topic will become politically charged and, as a result, be discussed differently on social media across geographies with diverging political beliefs. In addition, social media behavior itself is influenced by various dynamic factors, for instance platform algorithms [52] or changing governance structures, which affect public engagement [53], all of which may differ significantly across social media platforms, future epidemiological events, and national borders. Although our study revealed differences in the epidemiological early warning capabilities of geosocial media data across US county-level political clusters, these results should be treated with caution when considering future-use cases.

Conclusion

Our results confirmed the findings of previous research [9,10,12], demonstrating that geosocial media data can improve epidemiological early warning for consecutive waves of COVID-19 cases. In addition, we expand the existing literature by showing that the early warning capabilities of geosocial media data vary across US county clusters with differing political beliefs. For instance, geosocial media posts in republican counties (21 days) tend to precede increases in COVID-19 cases on average about 6.4 days earlier than in democrat counties (14.6 days). We hypothesize that this discrepancy in temporal lag between the geosocial media signal and the COVID-19 cases may stem from differences in the adoption of public health measures or population density variations across regions. In addition, we observed that the early warning capabilities of geosocial media data can be mitigated due to its susceptibility to a shift in trending topics and a decrease in signal strength over time.

Based on our findings, we would recommend that policy makers and researchers enhance and further investigate real-time geosocial media monitoring capabilities to improve epidemiological early warning systems. In addition, our findings suggest that it could be particularly beneficial for such systems to account for political beliefs prevalent across finer spatial scales such as county-level, given their potential to impact the early warning capabilities of geosocial media signals. Furthermore, since our results clearly highlight the value of geosocial media data for epidemiological early warning, we strongly encourage social media companies to grant researchers access to their data. Furthermore, future research should examine the early warning capabilities of different geosocial media topics specific to regional political beliefs and assess the transferability of our findings to other countries with different political environments. Furthermore, investigating the role of political communication strategies and potential improvements to social media algorithms to mitigate political polarization could enhance our understanding of how geosocial media data can be leveraged for future epidemiological events.

Acknowledgments

This research was funded in part by the Austrian Science Fund (FWF) Grant-DOI: 10.55776/I5117. For open access purposes, the author has applied a CC BY public copyright license to any author accepted manuscript version arising from this submission.

MS has been funded (in part) by contract 200-2016-91779 and cooperative agreement CDC-RFA-FT-23-0069 with the Centers for Disease Control and Prevention (CDC). The findings, conclusions, and views expressed are those of the author(s) and do not necessarily represent the official position of the CDC. MS was also partially supported by the National Institute of General Medical Sciences of the National Institutes of Health under award number R01GM130668. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

WG has been sponsored by NSF award #1841403.

Conflicts of Interest

MS has received institutional research funds from the Johnson and Johnson foundation, from Janssen global public health, and Pfizer. All other authors declare no conflicts of interest.

References

1. Ciotti M, Ciccozzi M, Terrinoni A, Jiang WC, Wang CB, Bernardini S. The COVID-19 pandemic. Crit Rev Clin Lab Sci 2020;57(6):365-388. [doi: 10.1080/10408363.2020.1783198] [Medline: 32645276]
2. GBD 2021 Demographics Collaborators. Global age-sex-specific mortality, life expectancy, and population estimates in 204 countries and territories and 811 subnational locations, 1950-2021, and the impact of the COVID-19 pandemic: a comprehensive demographic analysis for the global burden of disease study 2021. Lancet 2024;403(10440):1989-2056 [FREE Full text] [doi: 10.1016/S0140-6736(24)00476-8] [Medline: 38484753]
3. Gahlot P, Alley KD, Arora S, Das S, Nag A, Tyagi VK. Wastewater surveillance could serve as a pandemic early warning system for COVID-19 and beyond. WIREs Water 2023;10(4):e1650. [doi: 10.1002/wat2.1650]
4. Kamalrathne T, Amaratunga D, Haigh R, Kodituwakku L. Need for effective detection and early warnings for epidemic and pandemic preparedness planning in the context of multi-hazards: Lessons from the COVID-19 pandemic. Int J Disaster Risk Reduct 2023;92:103724 [FREE Full text] [doi: 10.1016/j.ijdrr.2023.103724] [Medline: 37197332]
5. MacIntyre CR, Chen X, Kunasekaran M, Quigley A, Lim S, Stone H, et al. Artificial intelligence in public health: the potential of epidemic early warning systems. J Int Med Res 2023;51(3):3000605231159335 [FREE Full text] [doi: 10.1177/03000605231159335] [Medline: 36967669]
6. Al-Surimi K, Khalifa M, Bahkali S, EL-Metwally A, Househ M. The potential of social media and internet-based data in preventing and fighting infectious diseases: from internet to Twitter. In: Rezza G, Ippolito G, editors. Emerging and Re-Emerging Viral Infections: Advances in Microbiology, Infectious Diseases and Public Health Volume 6. Advances in Experimental Medicine and Biology. Midtown Manhattan, New York City: Springer International Publishing; 2017:131-139.

7. Albinati J, Meira W, Pappa G, Teixeira M, Marques-Toledo C. Enhancement of epidemiological models for dengue fever based on twitter data. : Association for Computing Machinery; 2017 Presented at: DH '17: International Conference on Digital Health; 2017 July 2-5; London United Kingdom p. 109-118. [doi: 10.1145/3079452.3079464]
8. Butt ZA. Early warning for emerging infectious disease outbreaks: digital disease surveillance for public health preparedness and response. In: Accelerating Strategic Changes for Digital Transformation in the Healthcare Industry. Amsterdam: Elsevier; 2023:309-320.
9. Comito C. How COVID-19 Information Spread in U.S.? The Role of Twitter as Early Indicator of Epidemics. IEEE Trans. Serv. Comput 2022;15(3):1193-1205. [doi: 10.1109/tsc.2021.3091281]
10. Kogan NE, Clemente L, Liautaud P, Kaashoek J, Link NB, Nguyen AT, et al. An early warning approach to monitor COVID-19 activity with multiple digital traces in near real time. Sci Adv 2021;7(10):eabd6989 [FREE Full text] [doi: 10.1126/sciadv.abd6989] [Medline: 33674304]
11. Zhang Y, Chen K, Weng Y, Chen Z, Zhang J, Hubbard R. An intelligent early warning system of analyzing twitter data using machine learning on COVID-19 surveillance in the US. Expert Syst Appl 2022;198:116882 [FREE Full text] [doi: 10.1016/j.eswa.2022.116882] [Medline: 35308584]
12. Stolerman LM, Clemente L, Poirier C, Parag KV, Majumder A, Masyn S, et al. Using digital traces to build prospective and real-time county-level early warning systems to anticipate COVID-19 outbreaks in the United States. Sci Adv 2023;9(3):eabq0199 [FREE Full text] [doi: 10.1126/sciadv.abq0199] [Medline: 36652520]
13. Tsao SF, Chen H, Tisseverasinghe T, Yang Y, Li L, Butt ZA. What social media told us in the time of COVID-19: a scoping review. Lancet Digit Health 2021;3(3):e175-e194 [FREE Full text] [doi: 10.1016/S2589-7500(20)30315-0] [Medline: 33518503]
14. Ashokkumar A, Pennebaker JW. Social media conversations reveal large psychological shifts caused by COVID-19's onset across U.S. cities. Sci Adv 2021;7(39):eabg7843 [FREE Full text] [doi: 10.1126/sciadv.abg7843] [Medline: 34550738]
15. Das Swain V, Ye J, Ramesh SK, Mondal A, Abowd GD, De Choudhury M. Leveraging social media to predict COVID-19-induced disruptions to mental well-being among university students: modeling study. JMIR Form Res 2024;8:e52316 [FREE Full text] [doi: 10.2196/52316] [Medline: 38916951]
16. 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 [FREE Full text] [doi: 10.2196/26627] [Medline: 33724919]
17. Guntuku SC, Purtle J, Meisel ZF, Merchant RM, Agarwal A. Partisan differences in twitter language among US legislators during the COVID-19 pandemic: cross-sectional study. J Med Internet Res 2021;23(6):e27300 [FREE Full text] [doi: 10.2196/27300] [Medline: 33939620]
18. Jing E, Ahn YY. Characterizing partisan political narrative frameworks about COVID-19 on twitter. EPJ Data Sci 2021;10(1):53 [FREE Full text] [doi: 10.1140/epjds/s13688-021-00308-4] [Medline: 34745825]
19. Sylwester K, Purver M. Twitter language use reflects psychological differences between democrats and republicans. PLoS One 2015;10(9):e0137422 [FREE Full text] [doi: 10.1371/journal.pone.0137422] [Medline: 26375581]
20. Schmidt AL, Zollo F, Scala A, Betsch C, Quattrociocchi W. Polarization of the vaccination debate on Facebook. Vaccine 2018;36(25):3606-3612. [doi: 10.1016/j.vaccine.2018.05.040] [Medline: 29773322]
21. Howard P, Kollanyi B, Bradshaw S, Neudert LM. Social media, news and political information during the US election: was polarizing content concentrated in swing States? arXiv:1802.03573 2018. [doi: 10.48550/arXiv.1802.03573]
22. Lyu H, Wang J, Wu W, Duong V, Zhang X, Dye TD, et al. Social media study of public opinions on potential COVID-19 vaccines: informing dissent, disparities, and dissemination. Intell Med 2022;2(1):1-12 [FREE Full text] [doi: 10.1016/j.imed.2021.08.001] [Medline: 34457371]
23. Kahane LH. Politicizing the mask: political, economic and demographic factors affecting mask wearing behavior in the USA. East Econ J 2021;47(2):163-183 [FREE Full text] [doi: 10.1057/s41302-020-00186-0] [Medline: 33424048]
24. Allcott H, Boxell L, Conway J, Gentzkow M, Thaler M, Yang D. Polarization and public health: partisan differences in social distancing during the coronavirus pandemic. Journal of Public Economics 2020;191:104254. [doi: 10.1016/j.jpubeco.2020.104254] [Medline: 32836504]
25. Kaashoek J, Testa C, Chen JT, Stolerman LM, Krieger N, Hanage WP, et al. The evolving roles of US political partisanship and social vulnerability in the COVID-19 pandemic from February 2020-February 2021. PLOS Glob Public Health 2022;2(12):e0000557 [FREE Full text] [doi: 10.1371/journal.pgph.0000557] [Medline: 36962752]
26. Hart PS, Chinn S, Soroka S. Politicization and polarization in COVID-19 news coverage. Sci Commun 2020;42(5):679-697 [FREE Full text] [doi: 10.1177/1075547020950735] [Medline: 38602988]
27. Neelon B, Mutiso F, Mueller NT, Pearce JL, Benjamin-Neelon SE. Associations between governor political affiliation and COVID-19 cases, deaths, and testing in the U.S. Am J Prev Med 2021;61(1):115-119 [FREE Full text] [doi: 10.1016/j.amepre.2021.01.034] [Medline: 33775513]
28. Gadarian SK, Goodman SW, Pepinsky TB. Partisanship, health behavior, and policy attitudes in the early stages of the COVID-19 pandemic. PLoS One 2021;16(4):e0249596 [FREE Full text] [doi: 10.1371/journal.pone.0249596] [Medline: 33826646]

29. Das S, Samanta S, Banerjee J, Pal A, Giri B, Kar SS, et al. Is Omicron the end of pandemic or start of a new innings? Travel Med Infect Dis 2022;48:102332 [FREE Full text] [doi: 10.1016/j.tmaid.2022.102332] [Medline: 35472451]
30. US COVID-19 cases and deaths by state.: USAFacts; 2020. URL: https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/ [accessed 2025-01-02]
31. County presidential election returns 2000-2020. MIT Election Data And Science Lab. 2018. URL: https://search.dataone.org/view/sha256:6c387220de1dcdaaf94c2e7ca85474badf51d3b7047902823946c8769acde2d5 [accessed 2024-12-31]
32. Schultz E, Jacob R, Melcher JP, Fried A. Presidential Swing States. Lanham: Lexington Books; 2018.
33. Kriston L. A statistical definition of epidemic waves. Epidemiologia (Basel) 2023;4(3):267-275 [FREE Full text] [doi: 10.3390/epidemiologia4030027] [Medline: 37489498]
34. Zhang SX, Arroyo Marioli F, Gao R, Wang S. A second wave? What do people mean by COVID waves? - A working definition of epidemic waves. Risk Manag Healthc Policy 2021;14:3775-3782 [FREE Full text] [doi: 10.2147/RMHP.S326051] [Medline: 34548826]
35. Ayala A, Villalobos Dintrans P, Elorrieta F, Castillo C, Vargas C, Maddaleno M. Identification of COVID-19 waves: considerations for research and policy. Int J Environ Res Public Health 2021;18(21):11058 [FREE Full text] [doi: 10.3390/ijerph182111058] [Medline: 34769577]
36. Gimpel JG, Lovin N, Moy B, Reeves A. The urban–rural gulf in American political behavior. Polit Behav 2020;42(4):1343-1368. [doi: 10.1007/s11109-020-09601-w]
37. Sy KTL, White LF, Nichols BE. Population density and basic reproductive number of COVID-19 across United States counties. PLoS One 2021;16(4):e0249271 [FREE Full text] [doi: 10.1371/journal.pone.0249271] [Medline: 33882054]
38. Pettersen JH, Hannigan LJ, Gustavson K, Lund IO, Pearson RM, Jensen P, et al. COVID-19 pandemic quarantines and mental health among adolescents in Norway. JAMA Netw Open 2024;7(7):e2422189. [doi: 10.1001/jamanetworkopen.2024.22189] [Medline: 38995642]
39. Ferwana I, Varshney LR. The impact of COVID-19 lockdowns on mental health patient populations in the United States. Sci Rep 2024;14(1):5689 [FREE Full text] [doi: 10.1038/s41598-024-55879-9] [Medline: 38454064]
40. Perlis RH, Green J, Simonson M, Ognyanova K, Santillana M, Lin J, et al. Association between social media use and self-reported symptoms of depression in US adults. JAMA Netw Open 2021;4(11):e2136113 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.36113] [Medline: 34812844]
41. Islam AKMN, Laato S, Talukder S, Sutinen E. Misinformation sharing and social media fatigue during COVID-19: an affordance and cognitive load perspective. Technol Forecast Soc Change 2020;159:120201 [FREE Full text] [doi: 10.1016/j.techfore.2020.120201] [Medline: 32834137]
42. Xiao H, Zhang Z, Zhang L. An investigation on information quality, media richness, and social media fatigue during the disruptions of COVID-19 pandemic. Curr Psychol 2023;42(3):2488-2499 [FREE Full text] [doi: 10.1007/s12144-021-02253-x] [Medline: 34511861]
43. Li K, Jiang S, Yan X, Li J. Mechanism study of social media overload on health self-efficacy and anxiety. Heliyon 2024;10(1):e23326 [FREE Full text] [doi: 10.1016/j.heliyon.2023.e23326] [Medline: 38163164]
44. Sun J, Lee SK. "No more COVID-19 messages via social media, please": the mediating role of COVID-19 message fatigue between information overload, message avoidance, and behavioral intention. Curr Psychol 2023;2:1-15 [FREE Full text] [doi: 10.1007/s12144-023-04726-7] [Medline: 37359620]
45. Wang K, Gasser T. Alignment of curves by dynamic time warping. Ann. Statist 1997;25(3):1251-1276. [doi: 10.1214/aos/1069362747]
46. Cleasby IR, Wakefield ED, Morrissey BJ, Bodey TW, Votier SC, Bearhop S, et al. Using time-series similarity measures to compare animal movement trajectories in ecology. Behav Ecol Sociobiol 2019;73(11). [doi: 10.1007/s00265-019-2761-1]
47. Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. Phys Rev E Stat Nonlin Soft Matter Phys 2004;69(6 Pt 2):066138 [FREE Full text] [doi: 10.1103/PhysRevE.69.066138] [Medline: 15244698]
48. Kumar A, Kalra Y, Guan WW, Tibrewal V, Batta R, Chen A. COVID-19 impact on excess deaths of various causes in the United States. Annals of GIS 2021;28(4):457-478. [doi: 10.1080/19475683.2021.1982001]
49. Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. J Mach Learn Res 2003;3(null):993-1022. [doi: 10.7551/mitpress/1120.003.0082]
50. Grootendorst M. BERTopic: neural topic modeling with a class-based TF-IDF procedure. arXiv:2203.05794 2022 [FREE Full text] [doi: 10.1002/asi.21166]
51. Pham CM, Hoyle A, Sun S, Iyyer M. TopicGPT: a prompt-based topic modeling framework. arXiv:2311.01449 2023. [doi: 10.18653/v1/2024.naacl-long.164]
52. Bellina A, Castellano C, Pineau P, Iannelli G, De Marzo G. Effect of collaborative-filtering-based recommendation algorithms on opinion polarization. Phys Rev E 2023;108(5-1):054304. [doi: 10.1103/PhysRevE.108.054304] [Medline: 38115540]
53. Schmidt S, Zorenböhmer C, Arifi D, Resch B. Polarity-based sentiment analysis of georeferenced tweets related to the 2022 Twitter acquisition. Information 2023;14(2):71. [doi: 10.3390/info14020071]

Abbreviations

API: application programming interface
WHO: World Health Organization

Edited by T Mackey; submitted 19.03.24; peer-reviewed by J Abbas, L Stolerman; comments to author 26.09.24; revised version received 17.11.24; accepted 23.11.24; published 30.01.25.

Please cite as:
Arifi D, Resch B, Santillana M, Guan WW, Knoblauch S, Lautenbach S, Jaenisch T, Morales I, Havas C
Geosocial Media's Early Warning Capabilities Across US County-Level Political Clusters: Observational Study
JMIR Infodemiology 2025;5:e58539
URL: https://infodemiology.jmir.org/2025/1/e58539
doi:10.2196/58539
PMID:

Original Paper

Unraveling the Use of Disinformation Hashtags by Social Bots During the COVID-19 Pandemic: Social Networks Analysis

Victor Suarez-Lledo[1,2*], PhD; Esther Ortega-Martin[1,3*], MSc; Jesus Carretero-Bravo[1,4*], PhD; Begoña Ramos-Fiol[1*], MSc; Javier Alvarez-Galvez[1,3*], PhD

[1]Computational Social Science DataLab, University Institute of Research for Sustainable Social Development (INDESS), University of Cadiz, Jerez de la Frontera, Spain

[2]Department of Sociology, University of Granada, Granada, Spain

[3]Department of General Economy (Sociology Area), Faculty of Nursing and Physiotherapy, University of Cadiz, Cadiz, Spain

[4]Department of Quantitative Methods, Universidad Loyola Andalucía, Seville, Spain

[*]all authors contributed equally

Corresponding Author:
Victor Suarez-Lledo, PhD
Computational Social Science DataLab
University Institute of Research for Sustainable Social Development (INDESS)
University of Cadiz
Avda. de la Universidad, 4
Jerez de la Frontera, 11406
Spain
Phone: 34 956167216
Email: victor.sanz@uca.es

Abstract

Background: During the COVID-19 pandemic, social media platforms have been a venue for the exchange of messages, including those related to fake news. There are also accounts programmed to disseminate and amplify specific messages, which can affect individual decision-making and present new challenges for public health.

Objective: This study aimed to analyze how social bots use hashtags compared to human users on topics related to misinformation during the outbreak of the COVID-19 pandemic.

Methods: We selected posts on specific topics related to infodemics such as vaccines, hydroxychloroquine, military, conspiracy, laboratory, Bill Gates, 5G, and UV. We built a network based on the co-occurrence of hashtags and classified the posts based on their source. Using network analysis and community detection algorithms, we identified hashtags that tend to appear together in messages. For each topic, we extracted the most relevant subtopic communities, which are groups of interconnected hashtags.

Results: The distribution of bots and nonbots in each of these communities was uneven, with some sets of hashtags being more common among accounts classified as bots or nonbots. Hashtags related to the Trump and QAnon social movements were common among bots, and specific hashtags with anti-Asian sentiments were also identified. In the subcommunities most populated by bots in the case of vaccines, the group of hashtags including #billgates, #pandemic, and #china was among the most common.

Conclusions: The use of certain hashtags varies depending on the source, and some hashtags are used for different purposes. Understanding these patterns may help address the spread of health misinformation on social media networks.

(JMIR Infodemiology 2025;5:e50021) doi:10.2196/50021

KEYWORDS

social media; misinformation; COVID-19; bot; hashtags; disinformation; network analysis; community detection; dissemination; decision-making; social bot; infodemics; tweets; social media network

Introduction

From the swine influenza (H1N1) pandemic in 2009 to the subsequent outbreak of the H7N9 virus, also known as bird flu, in 2013, Twitter (subsequently rebranded as X) has increasingly become a popular platform for sharing health information [1,2]. Using posts, users can express their thoughts and opinions on many health topics. That is why specific interaction tasks have attracted the attention of researchers. This research can inform public policy by encouraging governments and health care professionals to allocate necessary resources, act, and plan accordingly [3,4]. These social media platforms have played a crucial role in providing information to the public during the COVID-19 pandemic. However, there was an increase in low-quality information, as well as the infodemic phenomenon. The infodemic, defined as an excess of information that makes it difficult for people to find reliable sources [5], can have harmful consequences [6].

The COVID-19 pandemic triggered mandatory lockdowns, social distancing, quarantines, and SARS-CoV-2–protective measures that would give rise to all sorts of opinions and behaviors [7]. During the COVID-19 pandemic, mandatory lockouts drastically altered people's daily routines (work, travel, and leisure activities) to levels never before experienced by the populations of the different countries affected by the new disease [8]. The state of uncertainty in the face of an invisible threat would transform previously normal situations into situations of risk. Direct social interaction with people outside the nuclear family, attending a concert, meeting for dinner with friends and family, shaking hands with someone, and even hugging or kissing became exceptional situations during the most uncertain periods of the pandemic—situations that, as has been observed retrospectively, would have a significant impact on the mental health of the population [9]. Likewise, the health crisis gave rise to the infodemic that, through social media platforms, opened the door to fake news, misconceptions, hoaxes, and anecdotal evidence about the origin of the pandemic, the social agents to blame for the situation, and the possible measures to be taken at a time of maximum uncertainty [10].

To understand how during the new context of health emergency misinformation spreads on these platforms, studies analyzed different elements, including the quality of information sources through URL analysis; identification of topics that generate misinformation; and analysis of online communities that spread misinformation, such as the antivaccine movement [11-14]. Others focused on the use of hashtags to describe the organization of the debate around the COVID-19–related topics. Researchers examined the frequency of use and the topic analysis of hashtags, and emphasized their main role in certain conversations [15,16]. By analyzing specific hashtags, studies also demonstrated how antivaccine communities, the proliferation of racist sentiments, or the spread of conspiracy theories are articulated on social media [17-19]. Some studies paid particular attention to how hashtags were used or combined in online conversations about the COVID-19 pandemic, using clustering techniques to describe the themes and combining hashtags with semantic text analysis and natural language processing (NLP) methods to improve topic modeling [20-22]. In addition, social network analysis (SNA) became useful to examine the co-occurrence of hashtags [23]. These studies demonstrate how the combination of different approach is useful to analyze online conversations more thoroughly.

Recently, the role of social bots has contributed to the spread of misinformation on social media platforms in various ways [24]. This issue garnered more attention as fake news and misinformation were significant factors during the COVID-19 pandemic. In this sense, some studies analyzed the role of bots regarding the spread of misinformation in general, while others have focused specifically on topics such as vaccines, conspiracy theories, hate speech, or reactions to other political actions [25-31]. However, a small amount of research compared the behavior of bots and humans [32,33].

To better understand the influence of bots on social media conversations, a previous study used topic modeling to segment the Twitter conversation and compare differences between accounts [34]. Nevertheless, the analysis did not focus on the usage of hashtags, which is the primary focus of this study. We aim to identify patterns and trends in hashtag usage to describe how bots and nonbots differ in their use of hashtags.

Only a few studies analyzed how social media bots use hashtags compared to humans. Most studies in this field examine specific hashtags [17-19,35-37]. To address this knowledge gap, we explore how social bots use hashtags specifically in connection with certain infodemic topics, issues that contribute to the generation or spread of fake news, misinformation, or discriminatory narratives. By analyzing how frequently hashtags co-occur, we aim to understand how they appear in the conversation and how they are combined. Besides, we also considered the context in which hashtags are used. They can be used ironically or convey disagreement. Our goal is to address three key questions: (1) What are the most common hashtag co-occurrences? (2) What are the differences in hashtag usage between bots and nonbots? and (3) Do bots and nonbots use certain hashtags in different ways?

Methods

Data Collection

Data collection for this study took place from March 16 to June 15, 2020, using the Twitter Streaming application programming interface (API). The hashtags #covid_19, #covid19, #covid, and #coronavirus were used to capture conversations about the first wave of COVID-19 pandemic, and only English-language posts were selected.

Based on previous research, we created a list of topics that were commonly associated with fake news or misinformation. This list includes ozone, laboratory, 5G, conspiracy, Bill Gates, milk, military, and UV. Vaccines were also identified as a controversial topic in multiple studies, so we added them to the list [38-40].

Ethical Considerations

The present study was approved by the Ethics Committee of the University of Cadiz (005_2024).

Bot Classification

To identify whether accounts on Twitter were bots or not, we used Botometer by OsoMe (formerly known as BotOrNot) [41]. This publicly available application uses over a thousand criteria to determine how closely a Twitter account matches the typical characteristics of social bots.

To create a binary classification (bot or nonbot) and prioritize identifying true positives over true negatives, we set a threshold value of 0.8 [34]. Using this threshold, we classified approximately 14.8% of the accounts as bots, which is in line with the findings of other research that found bot levels to be between 9% and 15% of the total number of Twitter accounts [42].

Botometer also provides rankings for 6 main types of bots, including echo-chamber, fake follower, financial, self-declared, spammer, and others, in addition to the overall likelihood of being a bot. In this study, we focused on analyzing the behavior of social bot accounts, particularly those that were not identified as automated accounts. These types of accounts are often associated with press agencies, companies, newspapers, or journals, and their primary purpose is to automatically publish information about a specific topic. These accounts may indicate that they are automated, for example, by including the word “bot” in their screen name or being identified as bots on Botwiki [41]. Therefore, we chose to exclude self-declared bots from our analysis due to their different characteristics compared with other social bots [41].

For this study, we classified accounts as nonbots if their probability of being a bot was less than 0.8, as self-declared bots if their probability of being a self-declared bot was greater than 0.8, and as bots if their probability of being a bot was greater than 0.8 and their probability of being a self-declared bot was less than 0.8. We then filtered out self-declared bots and considered both bots and nonbots for analysis.

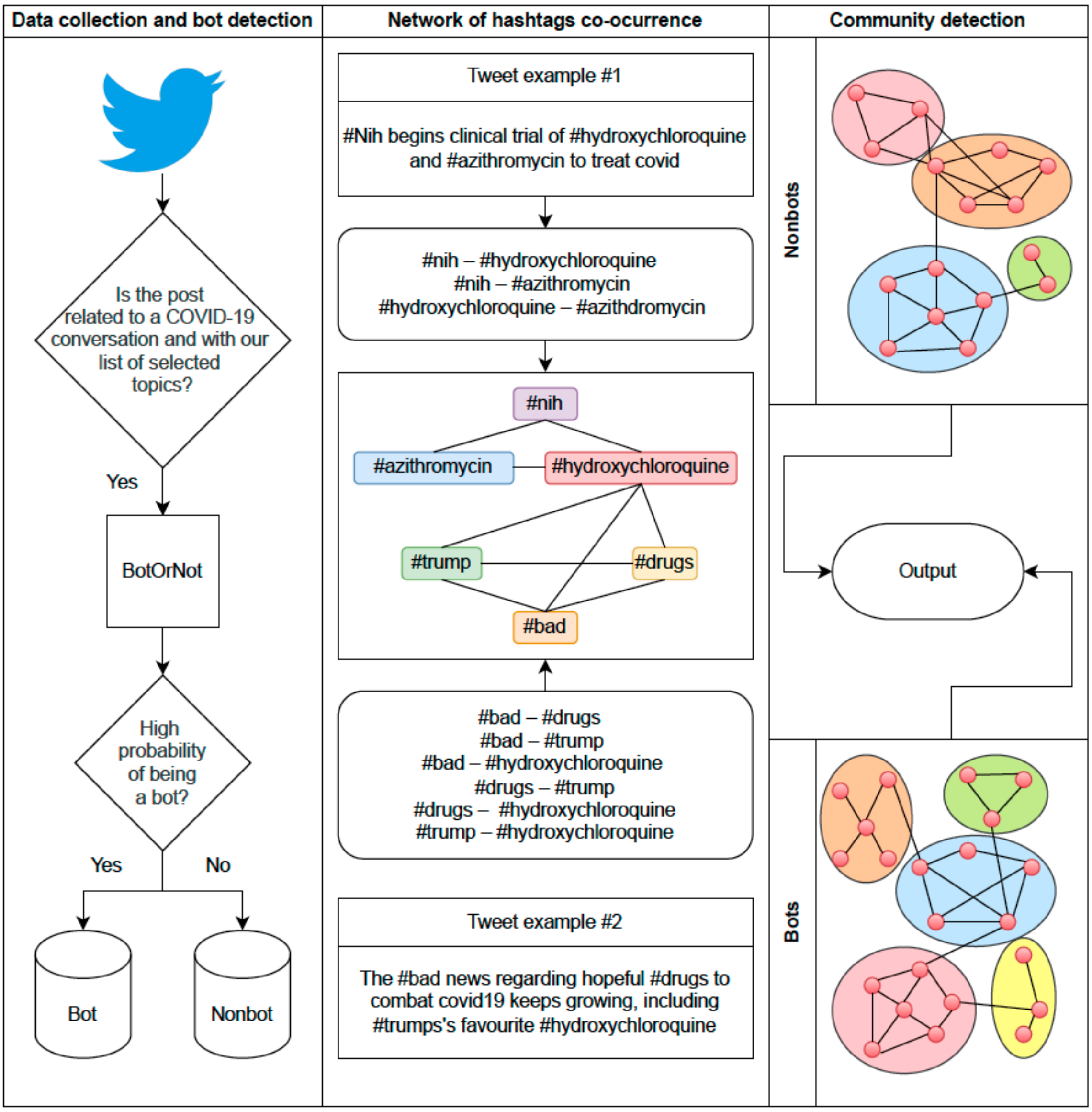
Network Analysis

To identify patterns in the usage of hashtags, we applied network analysis. We constructed a network by analyzing the co-occurrence of hashtags in posts and comparing the use of hashtags by bots and nonbots. In the network, hashtags were represented as nodes, and they were connected if they appeared in the same post. The weight of the connection between 2 hashtags was determined by the number of times they co-occurred.

We also calculated various metrics of connection, distribution, and segmentation of the hashtag network. We used the PageRank algorithm to identify the most important nodes in the network and the degree value, which represents the number of connections each hashtag has [43]. We also used the betweenness metric, which measures centrality [44]. In addition, we used the Louvain algorithm to detect the most important communities in the network. This algorithm maximizes a modularity score for each community, where the modularity measures the quality of the assignment of nodes to communities. This allowed us to identify hashtags that often co-occur together. We computed each metric separately considering whether the hashtags appear in posts posted by a bot or a nonbot. Figure 1 contains a flow diagram for the entire process.

In the following section, we first present the results for the entire network. In the following subsections, 1 for each topic, we segment the overall network of hashtag co-occurrences by extracting posts that specifically mention words related to each topic. For example, the network for vaccines will show the co-occurrences of all hashtags that appeared in posts about vaccines.

Figure 1. Flowchart from data collection to analysis.



Results

Overview

In total, we extracted around 107,173 posts from March to July 2020 that were related to the topics on our list. Most of these posts were about vaccines (59,090/107,173, 55.1%), hydroxychloroquine (17,731/107,173, 16.5%), or the military (12,548/107,173, 11.5%). Out of all the accounts analyzed, 85.2% (91,311/107,173) were identified with a low likely of being bots, that is, nonbots. Approximately 14.8% (15,862/107,173) of the posts were classified as likely being from bot accounts. As shown in Figure 2, the number of posts related to vaccines was consistently higher throughout the period, except for 2 specific moments. The first of these coincides with a message from US President Donald Trump recommending the use of hydroxychloroquine, an unproven drug. The second date also coincides with a message from Trump suggesting the injection of disinfectant to beat COVID-19 pandemic.

We created a graph of the full network of hashtags. For clarity, we selected a random sample from the entire collection of posts and depicted it in Figure 3. We also applied color to the Louvain communities and highlighted some hashtags that represent the topics analyzed in the study. This process is like the one we used for each topic in the list.

Figure 2. Bot and nonbot distribution by topic and date.

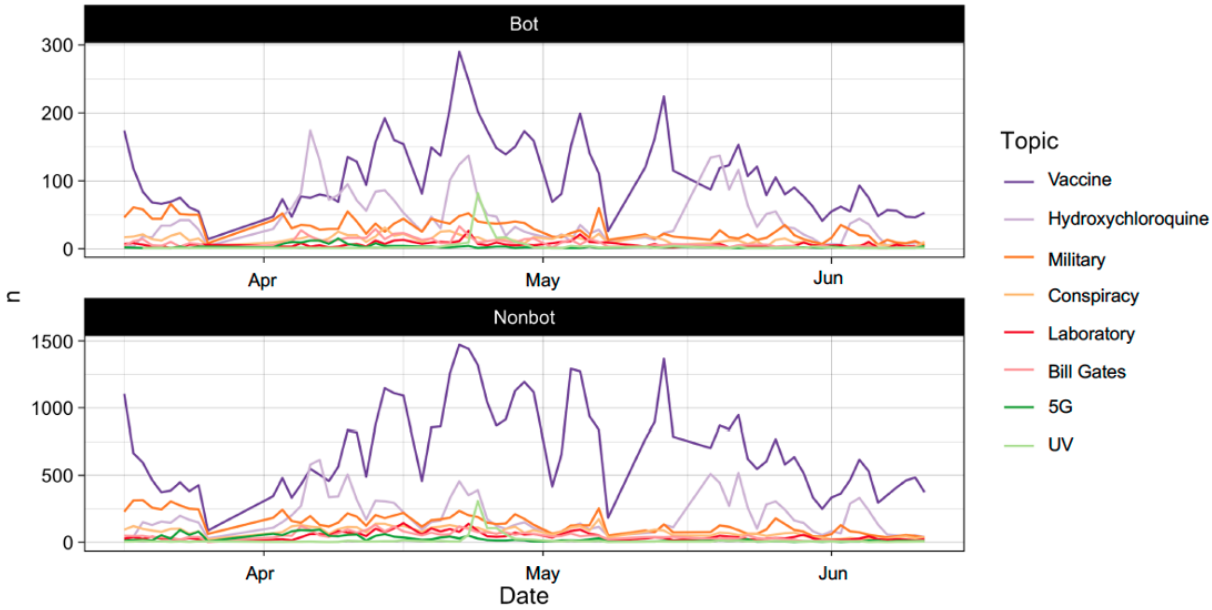
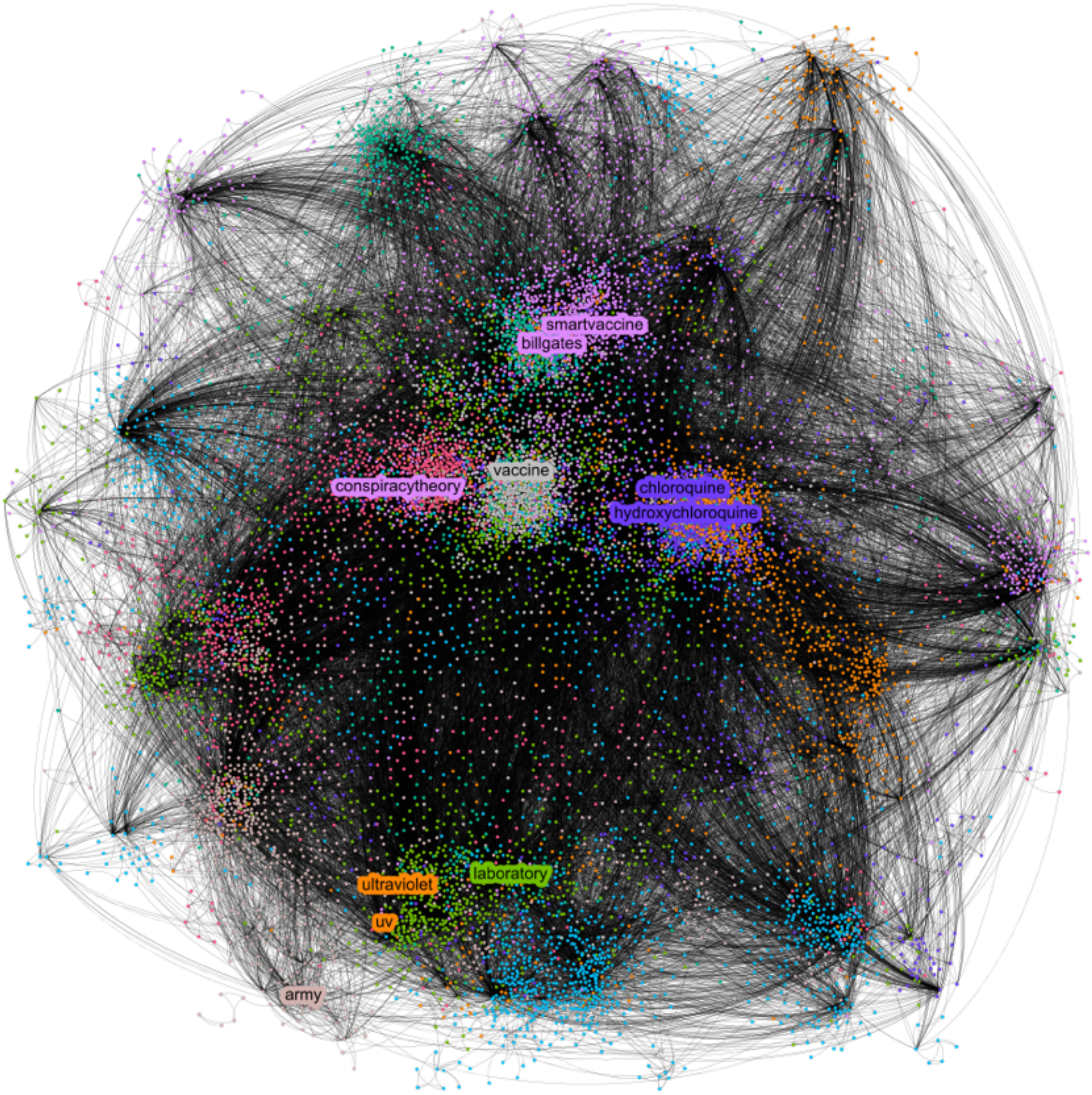


Figure 3. Hashtag network.



In Table 1, we present statistics for the overall hashtags network to provide a broad overview. As mentioned earlier, we calculated the metrics separately for each type of account. There are some differences in the most used hashtags between the 2 groups. For example, hashtags such as #Trump, #China, and #BillGates appear in both groups. However, the hashtag #vaccineswork is one of the most used by nonbots, while the hashtag #lka (which is the country code for Sri Lanka) is more frequently used by bots.

Table 1. Most common co-occurrences by bot and nonbot.

Hashtags	Posts, n (%)
Bots (n=3459)	
#chloroquine - #hydroxychloroquine	537 (15.52)
#hydroxychloroquine - #trump	490 (14.17)
#africaisnotalaboratory - #changeyourworld	437 (12.63)
#azithromycin - #hydroxychloroquine	345 (9.97)
#coronavirushoax - #prisonearth	280 (8.09)
#digitalvirus - #policestate	280 (8.09)
#digitalvirus - #prisonearth	280 (8.09)
#policestate - #prisonearth	280 (8.09)
#coronaviruslockdown - #lockdownextension	267 (7.72)
#changeyourworld - #coronacrisisuk	263 (7.6)
Nonbots (n=665)	
#hydroxychloroquine - #trump	133 (20)
#climatechange - #sustainability	106 (15.94)
#lka - #srilanka	86 (12.93)
#chloroquine - #hydroxychloroquine	84 (12.63)
#azithromycin - #hydroxychloroquine	72 (10.83)
#kag - #maga	53 (7.97)
#pandemic - #vaccine	35 (5.26)
#billgates - #vaccines	33 (4.96)
#kag - #qanon	33 (4.96)
#china - #vaccine	30 (4.51)

There are also some similarities in the co-occurrence of hashtags between the 2 groups. For example, hashtags #hydroxychloroquine and #trump appear in the same posts with higher frequency in both cases, at 14.17% (490/3459) and 20% (133/665), respectively. However, other hashtag pairs such as #kag-#maga, #billgates-#vaccines, or #kag-#qanon are common among bots. “KAG” stands for “Keep America Great,” which was President Trump’s campaign slogan in 2020, while “MAGA” stands for “Make America Great Again,” which was his campaign slogan in 2016. Both slogans have been associated with American nationalism, and the hashtag #MAGA has sometimes been used by white supremacist groups and Trump supporters.

On the other hand, nonbots tend to use other hashtag pairs such as #coronavirushoax-#prisionearth, #digitalvirus-#policestate, and #digitalvirus-#prisionearth. These hashtags, especially “#prisionearth,” were often used ironically to mock false rumors or exaggerations that were circulated online.

Vaccines

The most common co-occurrent hashtags used by nonbots regarding vaccines are #uk-#usa, #research-#science, #vaccineswork-#worldimmunizationweek. However, the most common hashtags in those posts posted by bots are #trump-#votebluetosaveamerica, #healthcare-#ppe, or even #healthcare-#ventilators. In addition, these last mentioned are exclusive of bots. That is, they only co-occur in posts from accounts classified as bots. Besides, it is worth mentioning that #billgates, along with #pandemic or #china, are the hashtags with the highest degree of connections, as seen in Table 2.

Table 2. Most important hashtags by topic.

Hashtags	Degree	PageRank	Betweenness
Vaccine			
billgates	44	0.025	22,728
pandemic	39	0.019	26,196
china	35	0.019	12,380
usa	30	0.013	7,375
vaccineswork	28	0.019	8,833
trump	28	0.015	15,704
stayhome	22	0.011	4,583
uk	21	0.010	2,703
science	21	0.011	5,048
france	19	0.008	2,064
Military			
trump	34	0.042	8,032
china	27	0.030	3,733
usa	22	0.026	5,561
italy	16	0.023	4,219
us	16	0.019	1,667
iran	15	0.020	1,938
russia	11	0.015	1,353
maga	10	0.012	620
wuhan	10	0.012	497
breaking	9	0.012	2,372
Laboratory			
wuhan	36	0.045	8,422
laboratory	26	0.033	11,660
africaisnotalaboratory	21	0.041	4,641
china	20	0.023	3,470
staysafe	11	0.017	7,566
stayhome	10	0.013	9,242
us	8	0.009	476
pandemic	8	0.009	8,614
coronaviruslockdown	7	0.011	1,676
healthcare	7	0.009	1,331
5G			
china	42	0.020	31,413
pandemic	27	0.012	25,136
wuhan	19	0.009	13,463
iot	18	0.008	11,045
qanon	17	0.008	6,437
bigdata	17	0.007	7,446
technology	17	0.008	8,731
ai	14	0.007	4,819

Hashtags	Degree	PageRank	Betweenness
tech	14	0.006	4,455
fakenews	14	0.007	8,353
Hydroxychloroquine			
trump	54	0.074	10,106
chloroquine	20	0.028	2,538
coronaviruspandemic	15	0.020	1,515
kag	14	0.017	897
maga	13	0.017	2,197
coronavirusoutbreak	12	0.016	1,089
india	12	0.017	855
hcq	12	0.020	1,468
usa	12	0.015	2,095
gop	11	0.014	636
Conspiracy			
conspiracy	35	0.084	1,872
conspiracytheory	25	0.054	2,111
conspiracytheories	16	0.037	686
pandemic	16	0.033	878
china	15	0.032	785
trump	12	0.030	732
disinformation	10	0.022	77
fakenews	10	0.023	321
usa	10	0.024	778
us	9	0.020	213
Bill Gates			
billgates	68	0.056	17,637
qanon	29	0.023	4,043
pandemic	27	0.024	7,341
maga	23	0.017	1,650
vaccines	19	0.016	5,232
stopbillgates	15	0.011	862
kag	13	0.009	104
trump	13	0.011	1,049
microsoft	13	0.010	1,978
usa	13	0.010	1,173
UV			
ai	14	0.041	839
trump	11	0.044	1,427
health	8	0.025	491
innovation	8	0.024	171
pandemic	8	0.029	428
uvlight	8	0.028	1,617
robots	7	0.023	754

Hashtags	Degree	PageRank	Betweenness
artificialintelligence	6	0.018	112
lysol	5	0.018	122
machinelearning	5	0.016	255

The algorithm extracted 5 different communities (Multimedia Appendix 1). We found significant differences in the hashtags that made up the Louvain communities. The first community contains hashtags related to news (#breaking, #usnews, and #breakingnews); countries (#canada, #france, #japan, #spain, and #africa); and others related to fake news like #wuhanvirus, #ccpvirus, #bioweapon, #hiddenhand, #psychopaths, #chinaisassho, and #madeinchina. This community is the most populated by bots, and the difference between bots and nonbots is the highest.

The second community contains hashtags related to famous people (#billgates, #anthonyfauci, and #georgesoros). These include people like Bill Gates and Anthony Fauci who played a leading role by holding provaccine positions. As in the previous case, we also found some hashtags related to fake news or conspiracy theories such as #billgatesisevil, #billgatesvaccine, #vaccinemafia, or #newworldorder. In this community, the quantity of nonbots is slightly higher than the number of bots.

On the other hand, the number of bots is also higher in the third community. In this case, the hashtags mention politics, such as #trump, #biden, and #borisjohnson. In addition, there were also some hashtags related to measures to curb the pandemic, such as #stayhome, #socialdistancing, or #lockdown. Only a few infodemic-related hashtags were found: #methanemouth, #pussygrabber, or #bananarepublic. The number of nonbots is higher in the other 2 communities. The fourth and fifth communities contain hashtags related to research and vaccines (#research, #health, and #medicine) or diseases and public health campaigns (#vaccineswork, #measles, #endpolio, and #healthforall), respectively. In particular, #vaccineswork is a hashtag used by health institutions such as the World Health Organization. Conversations on these hashtags were related to second waves and the importance of vaccines to fight against the COVID-19 pandemic.

Hydroxychloroquine

Hashtags related to Trump and the Republican movement were common in the case of hydroxychloroquine. These hashtags, such as #kag, #maga, #gop, #qanon, and #tcot, were more common in bot posts. Although #trump also appears in the case of nonbots, there were other hashtags related to news: #breaking-#breakingnew and #chinavirus-#wuhanvirus. Consequently, #trump has the highest degree of connection and the one with the highest betweenness. This hashtag, along with #chloroquine or #coronaviruspandemic, is the hashtag with the highest number of connections. There is a big difference between the first and the rest of the hashtags shown in Table 2. This difference indicates the leading role that #trump plays in the conversation about hydroxychloroquine.

We identified 8 different communities (Multimedia Appendix 1). Regarding the composition of the communities, it is worth mentioning the difference between the 2 most important ones. On the one hand, the first contains hashtags related to drugs, vaccines, or the pharmaceutical industry: #azithromycin, #biotech, #chloroquine, #lupus, #malaria, #cdc, or #hcq. In the same line, in the fourth community, the predominance of nonbots is noticeable. This time the hashtags mention countries (#uk, #us, #coronavirusuk, #france, #italy, and #germany), news (#worldnews and #usnews), TV series (#greysanatomy and #littlefireseverywhere), and supporting hashtags (#inthistogether).

On the other hand, in the second community, most of the hashtags are related to Trump or social movements related to him (#trump, #gop, #maga, and #donaldtrump). Nonetheless, some are against him (#notaleader, #worstpresidentinhistory, and #putinpuppet). In addition, the number of bots is higher than the number of nonbots, contrary to what happens in the first one.

Military

In this case, hashtags are related to specific countries that were mentioned during the pandemic. For nonbots, those most mentioned are #china-#us, #italy-#russia, and #lka-#srilanka. The latter is the most common among bots, followed in fourth place by #italy-#russia. Among the sets that do not mention countries, we find hashtags related to Trump (#gop-#trump, #kag-#maga, and #kag-#qanon).

These hashtags have similarities to those of hydroxychloroquine. The bots' unique hashtags are related to the Trump movement or Republican movements (#gop, #kag, and #qanon). In addition, #trump has the highest degree of connectivity and betweenness. This situation is also present in the communities (Multimedia Appendix 1). The first community detected contains hashtags related to Trump, and the second is related to military and veterans (#usmc, #veterans, or #usairforce). In both cases, these relationships take place in posts posted by bots.

Conspiracy

In this group, we found some hashtags related to conspiracy theories (or misinformation) and others related to countries. Regarding bots, the most common hashtags are #fakenews-#technology, #conspiracytheories-#socialmedia, and #donthecon-#trumplies. In line with this, for the nonbots, the most common hashtags are #conspiracytheory-#woke. The hashtags used only by bots are also related to racism (#racism-#sinophobia) or the economic system (#capitalismfails-#socialismworks).

Of the 6 most prominent communities (Multimedia Appendix 1), 3 of them have only nonbots. Topics in these communities are about minority groups (#blackpeople, #lgbt, and #amerikkka), about Trump (#maga, #bananarepublic, and #qanon), and about the pandemic (#coronavirusoutbreak,

#coronaviruspandemic, and #pandemictech). Of the other 3, in the first one, the number of nonbots is slightly higher than the number of bots. Some of the hashtags have to do with conspiracy theories (#conspiracytheory, #disinformation, and #propaganda), media (#qanonnfoxnews, #propaganda, and #fakenews), and others in a derogatory tone (#covidiot, #plandemic, and #plandemicdocumentary). On the other hand, in the second and fifth communities, the numbers of bots are higher. In this case, the most common hashtags are related to countries (#china, #us, and #iran), Iran specifically (#irancovidtruth and #iranregimechange), or against right-wing political parties (#rightwingignorance).

Laboratory

In this case, there are apparent differences in the geographical areas of the most used hashtags. On the one hand, nonbots mostly use #africaisnotalaboratory, while bots use #srilanka and #lka (country code for Sri Lanka). The hashtag #indiafightscorona is also common for bots. The hashtags #china-#wuhan are very common in both cases. This explains why #wuhan is the hashtag with the highest PageRank value and the highest degree of connection (Table 2), followed by #laboratory in second place and #africaisnotalaboratory in third place.

The differences between hashtags and the type of account that wrote the message were very clear in this case. On the one hand, in the first and fourth communities, the presence of bots is higher than nonbots (Multimedia Appendix 1). The first is focused on China, with some examples such as #ccpvirus, #chinamustexplain, or #chinaliedpeopedied, and the second is focused on Southeast Asia, such as #armenia, #abudhabi, or #masdarcity.

Bill Gates

The data from the Bill Gates conversation are similar to those obtained in the case of hydroxychloroquine. Trump-related hashtags were very common (#kag, #maga, and #qanon) in both bots and nonbots. The centrality and degree values are among the highest, as can be seen in Table 2. There were also new hashtags related to this type of political movement that only appears in this conversation, such as #crimesagainsthumanity, #gatesofhell, or #greatawakening. In addition, hashtags disparaging the figure of Bill Gates are also common, such as #saynotobillgates or #billgatesisevil.

We identified 5 communities of hashtags (Multimedia Appendix 1). Among the 3 largest communities, the number of bots is higher than the number of nonbots in the second one. In this community, the most frequent hashtags are #trump, #depopulationagenda, #eugenetics, #repubicans, #auspol, #qanon, and #americafirst. The hashtags, as mentioned above, are related to Trump or against some figures who have publicly supported vaccines. Examples are #trump, #americafirst, or #faucifraud. These hashtags can also be found in the first community, where the percentage of both account types is similar. However, in this community, the number of bots is not higher than that of nonbots. In the third community, the number of nonbots is higher than bots. Most hashtags in this community mention COVID-19 (#coronaviruschallenge, #coronavirusbill, #coronaviruschina, and #coronavirusnewyork), but other hashtags such as #hoaxvirus, #tedconnnect, #freedomovefear, or #trumpisevil also appear.

5G

Regarding 5G, hashtags related to technology or news were the predominant ones in the case of nonbots, such as #techwar-#tradewar or #bbcaq-#itvnews. On the other hand, in the case of bots, the hashtags continue to mention geographical areas: #america-#china and #america-#lka. There are other hashtags with higher intensity, for example, #chinesecoronavirus-#democratshateamerica or #conspiracytheories-#technology. As can be seen in Table 2, the #china hashtag gets the highest PageRank value, followed by #pandemic and #wuhan. In addition, #china has 42 degrees of connectivity, doubling the value of the second, which is #pandemic with 27 connections. But above all, these values indicate the central place these hashtags take in the conversation. On the one hand, the high degree indicates they co-occur with many different hashtags. On the other hand, a high betweenness value indicates a central place in the network.

This time, the algorithm found 5 different communities of hashtags (Multimedia Appendix 1). The presence of bots is higher than nonbots in the first 3. The first is related to #tech, #bigdata, #cibersecurity, and so on. The second one is focused on #conspiracytheories, #digitalskynet, and #misinformation. The third is focused on China, with hashtags such as #batflu, #chinesevirus, and #huaweithis. The last 2 communities, where the level of nonbots is higher, are formed by varied hashtags. The fourth community is formed by hashtags such as #kag or #maga. The fifth one contains hashtags mentioning rumors or disinformation: #fakenews, #disinformation, and #democrathoax. In this community, it is worth mentioning the appearance of hashtags related to #blacklivesmatter, such as #racism, #blacklivesmatteraustralia, or #policebrutality.

UV

In this case, the appearance of technology-related hashtags (#ai and #healthtech) is even more noticeable, especially in the case of bots (Table 2). On the other hand, the most common hashtags are #batflu-#quarantine in the case of nonbots. Concerning the 6 communities we found (Multimedia Appendix 1), in the first 3, the number of nonbots is higher. The subject matter of these communities is related to politicians (#trump, #joebiden, and #berniesanders), technology (#artificialintelligence, #bioinformatics, and #machinelearning), or more specifically to technological innovation (#health, #innovation, #coronavirusnewyorkty, and #smartcities).

Discussion

Principal Findings

This study examined the use of hashtags by social bots on Twitter during the early stages of the COVID-19 pandemic. By analyzing the co-occurrence of hashtags, we were able to identify differences between accounts classified as bots and nonbots. We used Louvain communities to further classify these co-occurrences and found consistent differences in hashtag usage between the 2 groups. We used social network analysis based on the co-occurrence of hashtags to take advantage of hashtags as key elements of online texts and understand how different users tag posts.

The analysis of hashtags provided several key insights into attitudes toward the COVID-19 pandemic and related behaviors. We consistently observed differences between bots and nonbots. In the case of bots, it was more common to find co-occurrences of hashtags related to political movements, particularly those on the right wing and related to Trump. This is consistent with findings in the literature showing a higher presence of conservatives in topics related to misinformation about COVID-19 pandemic [45].

In the conversation about vaccines, we observed that bots used hashtags related to fake news, such as #billgates and #china, more frequently. This analysis also identified specific uninformative hashtags (#ccpvirus and #chinesevirus) associated with anti-Asian sentiment [18]. Other hashtags expressed different opinions, such as criticism (#billgateisevil) or hate (#chinaliedpeopledied). It is worth noting that most of the tweets posted by nonbot users came from official accounts of institutions such as the World Health Organization, ministries of health, or entities related to public health. These messages focused on reporting on the evolution of the pandemic; the number of deaths; infection rates; and the health measures implemented, such as lockdowns and vaccination campaigns to contain the spread of the virus.

In our analysis of the conversation related to hydroxychloroquine, we identified 2 distinct communities of hashtags. One group was related to public health or medicine, while the other group was related to political movements and associated with Trump. Other studies have also found that Trump was involved in this conversation [46,47]. However, we also found that some of the hashtags in the conversation about hydroxychloroquine related to scientific facts. These differences suggest a highly polarized conversation with scientific arguments pitted against controversial political campaigns.

According to one of these studies [47], accounts with a higher impact on topics related to hydroxychloroquine disinformation were more likely to support President Trump. In addition, these types of content had a higher volume of tweets, longer duration in time, and greater echo. Our findings on the number of bots in these communities with politicized hashtags would partly explain the permanence over time and high echo values. Bots amplify these debates and increase the impact of the messages they disseminate [29,48,49]. However, our results also identify communities with anti–President Trump hashtags and higher numbers of bots. Liberals also engage in these conversations, although to a lesser extent than Conservatives [45].

These findings are extensible to topics such as the military or Bill Gates, where the conversation has been highly politicized and permeated with fake news. According to the results obtained, Trump occupied a leading role in the Twitter conversations during the period analyzed. This fact has also been noted in other previous works. Trump publicly supported the use of hydroxychloroquine and other drugs to combat the advance of the COVID-19 pandemic, with its corresponding impact on increased searches [50]. In addition, Bill Gates is often the protagonist in conspiracy theories [51].

Limitations and Strengths

There are several factors to consider when categorizing accounts as nonbot or bot. Botometer is backed by a large volume of research, but its effectiveness has been debated. It is important to remember that Botometer only provides a probability that an account is a bot, not a definitive classification. To get the most accurate results, it is recommended to compare probability distribution. However, in some cases it may be necessary to establish a binary classification for research purposes. In such cases, previous research has shown that using a cutoff value and comparing the results is a successful strategy [52].

It is important to consider the language constraint of this study. Only selecting tweets written in English may limit the focus to actors and events from English-speaking countries. In addition, no geographic limitations were placed on the collection of tweets, which allows for a larger volume of data but may also make it difficult to interpret results. It is also worth noting that the tweets analyzed in this study were from the early stages of the pandemic, and conversations and topics may have evolved over time.

Conclusion

Our analysis of hashtag usage on Twitter showed that there were differences in the patterns of use between bot and nonbot accounts. By grouping hashtags based on co-occurrence, we were able to identify distinct patterns in the usage of hashtags. On controversial or highly polarized issues, the hashtags used often pertained to the campaign or movement being promoted, with a significant portion related to Trump. In some cases, hashtags opposing these movements were also identified. On less polarized topics, hashtag usage was more diverse and included references to specific geographic locations or social groups. This analysis method can be useful in detecting hashtags that may be linked to fake news or misinformation, or in tracing the spread of such content on social media platforms.

Acknowledgments

We would like to acknowledge the support of the University Research Institute for Sustainable Social Development and the University of Cádiz. The publication is part of project NETDYNAMIC (CNS2022-135907), funded by MCIN/AEI/10.13039/501100011033 and by the European Union “Next Generation EU”/PRTR. The present study has also been supported by the project DCODES (PID2020-118589RB-I00), granted by the Spanish Ministry of Science and Innovation and financed by MCIN/AEI/10.13039/501100011033.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Bot distribution by topic.
[PNG File , 104 KB - infodemiology_v5i1e50021_app1.png]

References

1. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of tweets during the 2009 H1N1 outbreak. PLoS One 2010;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
2. Vos SC, Buckner MM. Social media messages in an emerging health crisis: tweeting bird Flu. J Health Commun 2016;21(3):301-308. [doi: 10.1080/10810730.2015.1064495] [Medline: 26192209]
3. Page BI, Shapiro RY. Effects of public opinion on policy. Am Polit Sci Rev 2014;77(1):175-190. [doi: 10.2307/1956018]
4. Han X, Wang J, Zhang M, Wang X. Using social media to mine and analyze public opinion related to COVID-19 in China. Int J Environ Res Public Health 2020;17(8):2788 [FREE Full text] [doi: 10.3390/ijerph17082788] [Medline: 32316647]
5. EPI-WIN: WHO information network for epidemics. World Health Organization. URL: https://www.who.int/teams/epi-win [accessed 2022-09-23]
6. Sylvia Chou WY, Gaysynsky A, Cappella JN. Where we go from here: health misinformation on social media. Am J Public Health 2020;110(S3):S273-S275. [doi: 10.2105/AJPH.2020.305905] [Medline: 33001722]
7. Kyprianidou M, Konstantinou P, Alvarez-Galvez J, Ceccarelli L, Gruszczyńska E, Mierzejewska-Floreani D, et al. Profiling hesitancy to COVID-19 vaccinations in six European countries: behavioral, attitudinal and demographic determinants. Behav Med 2023;49(4):362-373. [doi: 10.1080/08964289.2022.2071198] [Medline: 35546258]
8. Alvarez-Galvez J, Anastasiou A, Lamnisos D, Constantinou M, Nicolaou C, Papacostas S, et al. The impact of government actions and risk perception on the promotion of self-protective behaviors during the COVID-19 pandemic. PLoS One 2023;18(4 April). [doi: 10.21203/rs.3.rs-1270359/v1]
9. Gloster AT, Lamnisos D, Lubenko J, Presti G, Squatrito V, Constantinou M, et al. Impact of COVID-19 pandemic on mental health: an international study. PLoS One 2020;15(12):e0244809 [FREE Full text] [doi: 10.1371/journal.pone.0244809] [Medline: 33382859]
10. Balakrishnan V, Ng WZ, Soo MC, Han GJ, Lee CJ. Infodemic and fake news - A comprehensive overview of its global magnitude during the COVID-19 pandemic in 2021: a scoping review. Int J Disaster Risk Reduct 2022;78:103144 [FREE Full text] [doi: 10.1016/j.ijdrr.2022.103144] [Medline: 35791376]
11. Yang KC, Pierri F, Hui PM, Axelrod D, Torres-Lugo C, Bryden J, et al. The COVID-19 infodemic: Twitter versus Facebook. Big Data & Society 2021;8(1):205395172110138. [doi: 10.1177/20539517211013861]
12. Singh L, Bode L, Budak C, Kawintiranon K, Padden C, Vraga E. Understanding high- and low-quality URL sharing on COVID-19 Twitter streams. J Comput Soc Sci 2020;3(2):343-366. [doi: 10.1007/s42001-020-00093-6] [Medline: 33263092]
13. Rovetta A, Bhagavathula AS. Global infodemiology of COVID-19: analysis of Google web searches and Instagram hashtags. J Med Internet Res 2020;22(8):e20673 [FREE Full text] [doi: 10.2196/20673] [Medline: 32748790]
14. Blane JT, Bellutta D, Carley KM. Social-cyber maneuvers during the COVID-19 vaccine initial rollout: content analysis of tweets. J Med Internet Res 2022;24(3):e34040 [FREE Full text] [doi: 10.2196/34040] [Medline: 35044302]
15. Petersen K, Gerken JM. #COVID-19: an exploratory investigation of hashtag usage on Twitter. Health Policy 2021;125(4):541-547. [doi: 10.1016/j.healthpol.2021.01.001] [Medline: 33487479]
16. Xi W, Xu W, Zhang X, Ayalon L. A thematic analysis of Weibo topics (Chinese Twitter hashtags) regarding older adults during the COVID-19 outbreak. J Gerontol B Psychol Sci Soc Sci 2021;76(7):e306-e312. [doi: 10.1093/geronb/gbaa148] [Medline: 32882029]
17. Khadafi R, Nurmandi A, Qodir Z, Misran. Hashtag as a new weapon to resist the COVID-19 vaccination policy: a qualitative study of the anti-vaccine movement in Brazil, USA, and Indonesia. Hum Vaccin Immunother 2022;18(1):2042135 [FREE Full text] [doi: 10.1080/21645515.2022.2042135] [Medline: 35240923]
18. Hswen Y, Xu X, Hing A, Hawkins JB, Brownstein JS, Gee GC. Association of "#covid19" versus "#chinesevirus" with anti-Asian sentiments on Twitter: March 9-23, 2020. Am J Public Health 2021;111(5):956-964. [doi: 10.2105/AJPH.2021.306154] [Medline: 33734838]
19. Gruzd A, Mai P. Going viral: how a single tweet spawned a COVID-19 conspiracy theory on Twitter. Big Data & Society 2020;7(2):205395172093840. [doi: 10.1177/2053951720938405]
20. Tahamtan I, Potnis D, Mohammadi E, Miller LE, Singh V. Framing of and attention to COVID-19 on Twitter: thematic analysis of hashtags. J Med Internet Res 2021;23(9):e30800 [FREE Full text] [doi: 10.2196/30800] [Medline: 34406961]
21. Vicient C, Moreno A. Unsupervised topic discovery in micro-blogging networks. Expert Systems with Applications 2015;42(17-18):6472-6485. [doi: 10.1016/j.eswa.2015.04.014]
22. Alkhodair SA, Fung BCM, Rahman O, Hung PCK. Improving interpretations of topic modeling in microblogs. Asso for Info Science & Tech 2017;69(4):528-540. [doi: 10.1002/asi.23980]

23. Eriksson Krutrök M, Lindgren S. Continued contexts of terror: analyzing temporal patterns of hashtag co-occurrence as discursive articulations. Social Media + Society 2018;4(4):205630511881364. [doi: 10.1177/2056305118813649]
24. Caldarelli G, de Nicola R, del Vigna F, Petrocchi M, Saracco F. The role of bot squads in the political propaganda on Twitter. Commun Phys 2020;3(1):1-15. [doi: 10.1038/s42005-020-0340-4]
25. Ruiz-Núñez C, Segado-Fernández S, Jiménez-Gómez B, Hidalgo PJJ, Magdalena CSR, Pollo MDCA, et al. Bots' activity on COVID-19 pro and anti-vaccination networks: analysis of Spanish-written messages on Twitter. Vaccines (Basel) 2022;10(8):1240 [FREE Full text] [doi: 10.3390/vaccines10081240] [Medline: 36016126]
26. Yousefinaghani S, Dara R, Mubareka S, Papadopoulos A, Sharif S. An analysis of COVID-19 vaccine sentiments and opinions on Twitter. Int J Infect Dis 2021;108:256-262 [FREE Full text] [doi: 10.1016/j.ijid.2021.05.059] [Medline: 34052407]
27. Scannell D, Desens L, Guadagno M, Tra Y, Acker E, Sheridan K, et al. COVID-19 vaccine discourse on Twitter: a content analysis of persuasion techniques, sentiment and mis/disinformation. J Health Commun 2021;26(7):443-459. [doi: 10.1080/10810730.2021.1955050] [Medline: 34346288]
28. 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 [FREE Full text] [doi: 10.2196/22374] [Medline: 32936771]
29. Uyheng J, Carley KM. Bots and online hate during the COVID-19 pandemic: case studies in the United States and the Philippines. J Comput Soc Sci 2020;3(2):445-468. [doi: 10.1007/s42001-020-00087-4] [Medline: 33102925]
30. Uyheng J, Bellutta D, Carley KM. Bots amplify and redirect hate speech in online discourse about racism during the COVID-19 pandemic. Social Media + Society 2022;8(3):205630512211047. [doi: 10.1177/20563051221104749]
31. Green M, Musi E, Rowe F, Charles D, Pollock FD, Kypridemos C, et al. Identifying how COVID-19-related misinformation reacts to the announcement of the UK national lockdown: an interrupted time-series study. Big Data & Society 2021;8(1):205395172110138. [doi: 10.1177/20539517211013869]
32. Chang HCH, Ferrara E. Comparative analysis of social bots and humans during the COVID-19 pandemic. J Comput Soc Sci 2022;5(2):1409-1425 [FREE Full text] [doi: 10.1007/s42001-022-00173-9] [Medline: 35789937]
33. Zhang M, Chen Z, Qi X, Liu J. Could social bots' sentiment engagement shape humans' sentiment on COVID-19 vaccine discussion on Twitter? Sustainability 2022;14(9):5566. [doi: 10.3390/su14095566]
34. Suarez-Lledo V, Alvarez-Galvez J. Assessing the role of social bots during the COVID-19 pandemic: infodemic, disagreement, and criticism. J Med Internet Res 2022;24(8):e36085 [FREE Full text] [doi: 10.2196/36085] [Medline: 35839385]
35. Graham T, Bruns A, Angus D, Hurcombe E, Hames S. #IStandWithDan versus #DictatorDan: the polarised dynamics of Twitter discussions about Victoria's COVID-19 restrictions. Media International Australia 2020;179(1):1329878X2098178. [doi: 10.1177/1329878X20981780]
36. Ruffer N, Knitza J, Krusche M. #Covid4Rheum: an analytical Twitter study in the time of the COVID-19 pandemic. Rheumatol Int 2020;40(12):2031-2037. [doi: 10.1007/s00296-020-04710-5] [Medline: 32995894]
37. Liu AKC, Ophir Y, Tsai SA, Walter D, Himelboim I. Hashtag activism in a politicized pandemic: framing the campaign to include Taiwan in the World Health Organization's efforts to combat COVID-19. New Media & Society 2022;26(6):146144482210991. [doi: 10.1177/14614448221099173]
38. Sharevski F, Alsaadi R, Jachim P, Pieroni E. Misinformation warnings: Twitter's soft moderation effects on COVID-19 vaccine belief echoes. Comput Secur 2022;114:102577 [FREE Full text] [doi: 10.1016/j.cose.2021.102577] [Medline: 34934255]
39. Calac AJ, Haupt MR, Li Z, Mackey T. Spread of COVID-19 vaccine misinformation in the ninth inning: retrospective observational infodemic study. JMIR Infodemiology 2022;2(1):e33587 [FREE Full text] [doi: 10.2196/33587] [Medline: 35320982]
40. Herrera-Peco I, Jiménez-Gómez B, Romero Magdalena CS, Deudero JJ, García-Puente M, Benítez de Gracia E, et al. Antivaccine movement and COVID-19 negationism: a content analysis of Spanish-written messages on Twitter. Vaccines (Basel) 2021;9(6):656 [FREE Full text] [doi: 10.3390/vaccines9060656] [Medline: 34203946]
41. Sayyadiharikandeh M, Varol O, Yang KC, Flammini A, Menczer F. Detection of novel social bots by ensembles of specialized classifiers. 2020 Presented at: CIKM '20: The 29th ACM International Conference on Information and Knowledge Management; October 19-23, 2020; Virtual event, Ireland. [doi: 10.1145/3340531.3412698]
42. Varol O, Ferrara E, Davis CA, Menczer F, Flammini A. Online human-bot interactions: detection, estimation, and characterization. 2017 Presented at: Proceedings of the International AAAI Conference on Web and Social Media; May 15-18, 2017; Montreal, QC p. 280-289 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14871 [doi: 10.1609/icwsm.v11i1.14871]
43. Page L, Brin S, Motwani R, Winograd T. The pageRank citation ranking: bringing order to the web. WWW Internet And Web Information Systems 1998;54(1999-66).
44. Freeman LC. A set of measures of centrality based on betweenness. Sociometry 1977;40(1):35-41. [doi: 10.2307/3033543]
45. Havey NF. Partisan public health: how does political ideology influence support for COVID-19 related misinformation? J Comput Soc Sci 2020;3(2):319-342. [doi: 10.1007/s42001-020-00089-2] [Medline: 33163686]

46. Blevins JL, Edgerton E, Jason DP, Lee JJ. Shouting into the wind: medical science versus "B.S." in the Twitter maelstrom of politics and misinformation about hydroxychloroquine. Social Media + Society 2021;7(2):205630512110249. [doi: 10.1177/20563051211024977]
47. Haupt MR, Li J, Mackey TK. Identifying and characterizing scientific authority-related misinformation discourse about hydroxychloroquine on Twitter using unsupervised machine learning. Big Data & Society 2021;8(1). [doi: 10.1177/20539517211013843]
48. Xu W, Sasahara K. Characterizing the roles of bots on Twitter during the COVID-19 infodemic. J Comput Soc Sci 2022;5(1):591-609 [FREE Full text] [doi: 10.1007/s42001-021-00139-3] [Medline: 34485752]
49. Stella M, Ferrara E, de Domenico M. Bots increase exposure to negative and inflammatory content in online social systems. Proc Natl Acad Sci U S A 2018;115(49):12435-12440 [FREE Full text] [doi: 10.1073/pnas.1803470115] [Medline: 30459270]
50. Niburski K, Niburski O. Impact of trump's promotion of unproven COVID-19 treatments and subsequent internet trends: observational study. J Med Internet Res 2020;22(11):e20044 [FREE Full text] [doi: 10.2196/20044] [Medline: 33151895]
51. Shahsavari S, Holur P, Wang T, Tangherlini TR, Roychowdhury V. Conspiracy in the time of corona: automatic detection of emerging COVID-19 conspiracy theories in social media and the news. J Comput Soc Sci 2020;3(2):279-317. [doi: 10.1007/s42001-020-00086-5] [Medline: 33134595]
52. Shao C, Ciampaglia GL, Varol O, Yang KC, Flammini A, Menczer F. The spread of low-credibility content by social bots. Nat Commun 2018;9(1):4787 [FREE Full text] [doi: 10.1038/s41467-018-06930-7] [Medline: 30459415]

Abbreviations

API: application programming interface
NLP: natural language processing
SNA: social network analysis

Edited by T Purnat; submitted 16.06.23; peer-reviewed by I Herrera-Peco, A Rovetta; comments to author 21.10.23; revised version received 08.02.24; accepted 15.05.24; published 09.01.25.

Please cite as:
Suarez-Lledo V, Ortega-Martin E, Carretero-Bravo J, Ramos-Fiol B, Alvarez-Galvez J
Unraveling the Use of Disinformation Hashtags by Social Bots During the COVID-19 Pandemic: Social Networks Analysis
JMIR Infodemiology 2025;5:e50021
URL: https://infodemiology.jmir.org/2025/1/e50021
doi:10.2196/50021
PMID:

Original Paper

Identifying Misinformation About Unproven Cancer Treatments on Social Media Using User-Friendly Linguistic Characteristics: Content Analysis

Ilona Fridman[1], PhD; Dahlia Boyles[2], MA; Ria Chheda[3], BS; Carrie Baldwin-SoRelle[4], MLiS; Angela B Smith[1], MD; Jennifer Elston Lafata[1,5], PhD

[1]Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC, United States
[2]Department of Communication, University of North Carolina, Chapel Hill, NC, United States
[3]Computer Science Department, University of North Carolina, Chapel Hill, NC, United States
[4]Health Sciences Library, University of North Carolina, Chapel Hill, NC, United States
[5]Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, United States

Corresponding Author:
Ilona Fridman, PhD
Lineberger Comprehensive Cancer Center
University of North Carolina
450 West Dr
Chapel Hill, NC, 27599
United States
Phone: 1 6469028137
Email: ilona_fridman@med.unc.edu

Abstract

Background: Health misinformation, prevalent in social media, poses a significant threat to individuals, particularly those dealing with serious illnesses such as cancer. The current recommendations for users on how to avoid cancer misinformation are challenging because they require users to have research skills.

Objective: This study addresses this problem by identifying user-friendly characteristics of misinformation that could be easily observed by users to help them flag misinformation on social media.

Methods: Using a structured review of the literature on algorithmic misinformation detection across political, social, and computer science, we assembled linguistic characteristics associated with misinformation. We then collected datasets by mining X (previously known as Twitter) posts using keywords related to unproven cancer therapies and cancer center usernames. This search, coupled with manual labeling, allowed us to create a dataset with misinformation and 2 control datasets. We used natural language processing to model linguistic characteristics within these datasets. Two experiments with 2 control datasets used predictive modeling and Lasso regression to evaluate the effectiveness of linguistic characteristics in identifying misinformation.

Results: User-friendly linguistic characteristics were extracted from 88 papers. The short-listed characteristics did not yield optimal results in the first experiment but predicted misinformation with an accuracy of 73% in the second experiment, in which posts with misinformation were compared with posts from health care systems. The linguistic characteristics that consistently negatively predicted misinformation included tentative language, location, URLs, and hashtags, while numbers, absolute language, and certainty expressions consistently predicted misinformation positively.

Conclusions: This analysis resulted in user-friendly recommendations, such as exercising caution when encountering social media posts featuring unwavering assurances or specific numbers lacking references. Future studies should test the efficacy of the recommendations among information users.

(JMIR Infodemiology 2025;5:e62703) doi:10.2196/62703

KEYWORDS

linguistic characteristics; linguistic features; cancer; Linguistic Inquiry and Word Count; misinformation; X; Twitter; cancer; alternative therapy; oncology; social media; natural language processing; machine learning; synthesis; review methodology; search; literature review

Introduction

Approximately 16% of people reported using social media to inform their medical decisions [1]. This percentage, based on estimates from the National Cancer Center, equates to 37 million adults in the United States. A recent systematic review estimated that up to 40% of health-related social media posts contain misinformation [2]. Misinformation could cause more harm to individuals with serious conditions such as cancer. Patients who believe in misinformation and use unproven therapies in parallel or in place of cancer treatment tend to be less adherent to evidence-based treatment [3-5]. Moreover, patients with cancer might choose to delay or reject evidence-based treatment and instead pursue unproven and potentially toxic therapies, which, for some patients, results in up to 2.5 times shorter life expectancy [6]. Approximately 30% of cancer-related social media posts on Facebook, Reddit, Pinterest, and X (previously known as Twitter) contain misinformation, and a staggering 77% of these posts have the potential to encourage patients to pursue futile and toxic therapies, resulting in physical, psychological, and logistical burdens [7]. Cancer misinformation persists across various cancer types and is more pervasive in more prevalent cancers. Across various social media platforms, two-thirds of the most shared posts about prostate cancer contain misinformation [8]. Researchers identified misinformation in 59% of posts related to breast cancer prevention and treatment [9] and 30% of posts related to gynecological cancer [10]. When surveyed, 70% of patients with cancer reported encountering misinformation about cancer on social media, with 71% believing that some of this misinformation was accurate [11].

There is a growing need to protect health information users from misinformation, especially those who are affected by serious conditions such as cancer. Multiple recommendations have been developed to assist individuals in their search for reliable health information [12-14]. However, many of the recommendations are complex, as they require individuals to possess a certain level of scientific knowledge and skills. For instance, recommendations frequently suggest taking steps such as identifying authors and their credentials, evaluating potential conflicts of interest, understanding funding sources, and assessing the original sources of scientific information. Considering the time and expertise required, expecting individuals to perform these tasks routinely is unrealistic. Moreover, these guidelines often fall short when it comes to addressing the challenges posed by social media platforms. Those who post may not disclose their real names or sources of findings, which makes some recommended steps not possible.

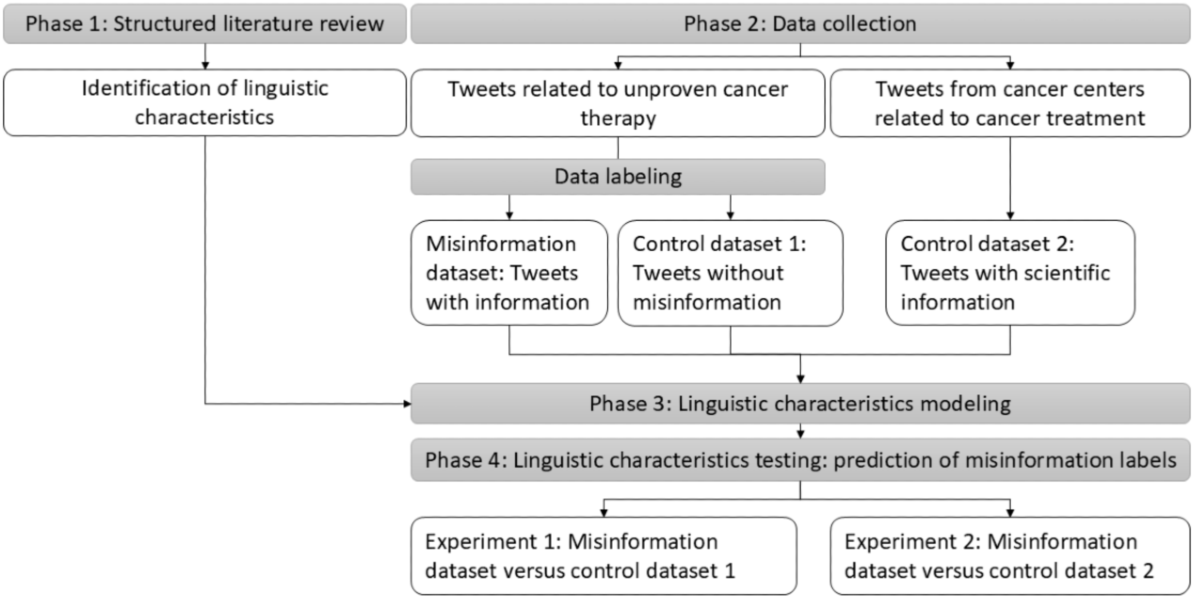
In this work, our goal is to identify user-friendly recommendations for addressing the high rate of misinformation on social media. We began by exploring literature on the algorithmic detection of misinformation. The algorithmic approach often involves the analysis of linguistic characteristics that differentiate between factual information and misinformation [15]. Linguistic characteristics describe a body of text in an abstract manner regardless of context and may include counts of words and word parts such as nouns, verbs, adjectives, and negations, as well as specific symbols such as URLs, hashtags, and question marks. An additional category of linguistic characteristics includes words associated with the psychological state of an author [16], which includes words related to emotions, expressions of certainty, tentativeness, insight, persuasion, and gratitude. To date, linguistic characteristics have been used by algorithms only. However, some of these characteristics are observable and could be used by individuals when they need to evaluate the credibility of the text. While individuals are unlikely to count words in social media posts regularly, they may routinely note other linguistic characteristics, such as emotions, URLs, and a strong degree of certainty. Linguistic characteristics have been shown to be effective in distinguishing misinformation from factual information across multiple contexts. However, it is unknown (1) whether the linguistic characteristics are effective in cancer-related context and (2) which subset of user-friendly linguistic characteristics could effectively distinguish misinformation. In this work, we identify the linguistic characteristics specific to the context of cancer. These characteristics will be recommended as guidelines for health information users when browsing social media.

Methods

Study Design

The main sequence of study procedures is illustrated in Figure 1 and includes (1) a structured literature review, in which we assemble linguistic characteristics that were used in algorithms for distinguishing factual information and misinformation (phase 1); (2) data collection, which encompasses assembling cancer-related posts using the X application programming interface (API) and labeling them as misinformation and non-misinformation (phase 2); (3) identification of the linguistic characteristics in collected datasets using natural language processing tools (phase 3); and (4) conducting predictive modeling analysis to evaluate the effectiveness of linguistic characteristics in distinguishing social media posts with cancer misinformation (phase 4).

Figure 1. Summary of the study procedures.



Ethical Considerations

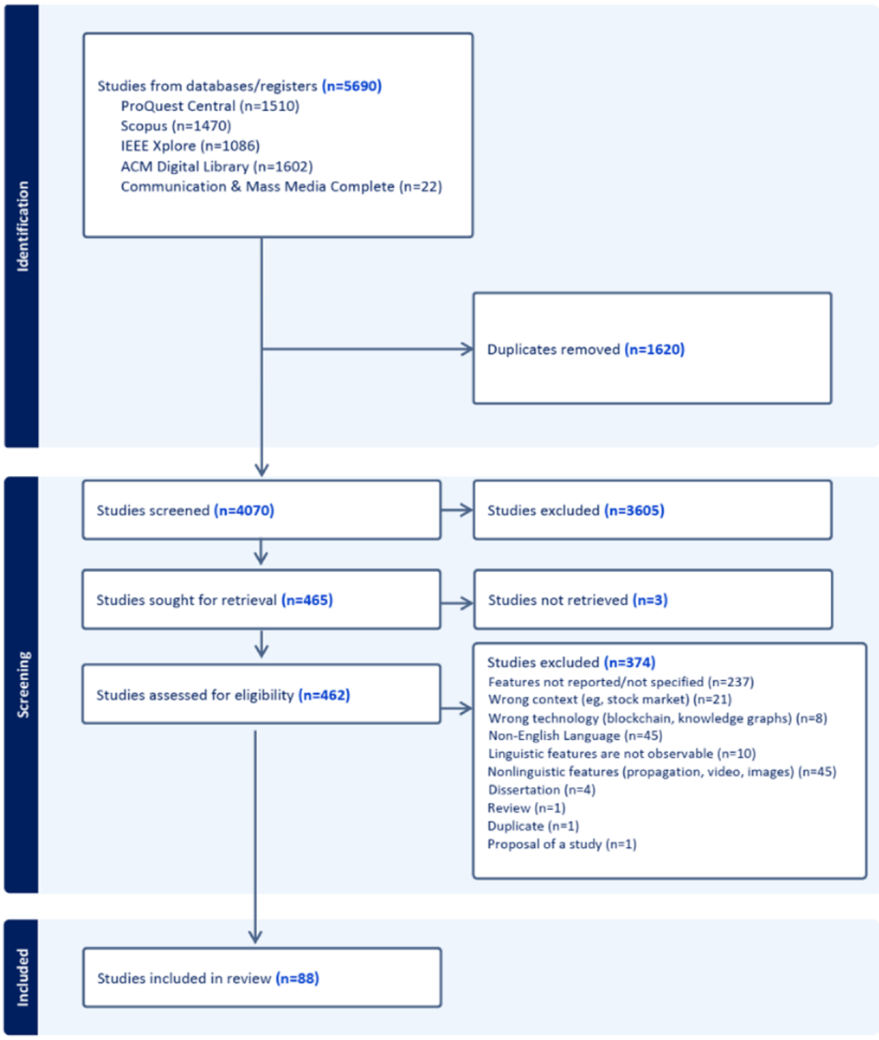
The study was institutional review board–approved by the University of North Carolina (IRB#21-2861). This was an analysis of publicly available data. As such, participants were not compensated and did not need to provide consent for the study, because the study did not involve any prospective data collection. To protect the confidentiality and anonymity of participants in this secondary data analysis, we reworded reported posts from X.

Structured Literature Review

To identify linguistic characteristics, we developed a literature review protocol that included the search strategy and keywords. This process was informed by a collaboration with a health sciences librarian (CBS), who suggested an initial set of keywords referenced in several relevant reviews [17-21]. She also created an expanded title, abstract, and keyword search strategies for each of the following concepts: (1) text as a unit of analysis, (2) misinformation, (3) algorithms, (4) internet, and (5) linguistic features or characteristics. After the search was peer reviewed by a second health sciences librarian (CB), 5 databases were searched: ProQuest Central (ProQuest), which includes the arXiv repository; Scopus (Elsevier); IEEE Xplore (Institute of Electrical and Electronics Engineers); ACM Digital Library (Association for Computing Machinery); and Communication & Mass Media Complete (EBSCOhost). The keywords and search strategies are reported in Multimedia Appendix 1. Results were limited to citations published between January 2012 and December 2022. Within databases, results were limited to journal papers, conference proceedings, working papers, and book chapters.

Two reviewers (IF and DB) independently coded titles and abstracts in Covidence software (Veritas Health Innovation) [22] and resolved conflict in codes during research meetings. Papers were included if they focused on detecting misinformation and contained a "Methods" section describing an approach for algorithmically detecting misinformation (eg, reviews and viewpoints were excluded). Examples of the algorithms included supervised and semisupervised machine learning (eg, Bidirectional Encoder Representations from Transformers [BERT] classification) that was built on linguistic characteristics. Papers were excluded if they did not report specific linguistic characteristics, focused on misinformation in any language other than English, or used human coding but not algorithms. The detailed inclusion-exclusion criteria and PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram are reported in Figure 2.

Figure 2. Flowchart of paper identification and extraction.



Identification of Linguistic Characteristics

Upon identifying eligible papers, 2 team members (IF and DB) reviewed the full text and extracted the linguistic characteristics. Around 11% (10/90) of papers underwent double-coding. After reviewers reached an agreement, we continued with single coding. The linguistic characteristics were extracted based on the following criteria: observability, applicability, and generalizability. The observability criterion was related to whether readers could easily observe the linguistic characteristics within the text; for example, positive emotions could be easily observed while morale or cognitive language styles may be difficult to distinguish. The applicability criterion distinguished linguistic characteristics that readers could easily apply while reading the text. For instance, common characteristics such as the number of words required substantial effort from readers to evaluate and, therefore, were deemed nonapplicable. In contrast, readers could easily use citations and hashtags in their post evaluations as the mere presence of these characteristics was determined to be helpful in identifying misinformation. The third criterion, generalizability, was chosen to ensure that linguistic characteristics were not related to a specific context but could be applied across various contexts. Thus, characteristics that were based on specific words such as "COVID-19," or "cure" were excluded.

Data Collection: Unproven Therapy

Overview

To test how extracted linguistic characteristics could distinguish social media posts from misinformation and factual information, we collected social media posts from X. Misinformation was operationalized here *as information that promoted cancer treatment that was known as ineffective or information that suggested cancer causes not supported by current scientific evidence* [23]. This definition focused our investigation on misinformation that could be harmful to patients with cancer or cancer survivors. Based on this operationalization, we searched existing resources that summarized unproven cancer therapy, such as "List of unproven cancer therapy" [24], a list of "Illegally sold cancer drugs" [25], and previous literature [23,26]. We extracted keywords and constructed 176 queries associated with unproven cancer treatments (Multimedia Appendix 2). Using these queries, we randomly selected up to 500 posts per query from social media. We used R software (R Foundation for Statistical Computing) to access the Academic X API. The data were manually evaluated to determine their relevance to the cancer context and unproven therapies. Queries were edited to ensure relevance. Upon corrections, the data collection was implemented on a schedule every other week between July 2022 and August 2023. After data collection was completed, the duplicate posts were removed.

Data Labeling

To distinguish posts with misinformation from other discussions, 2 reviewers (IF and CR) double-coded a randomly chosen subset of 1064 posts, achieving an acceptable interrater agreement of 0.68 measured with Krippendorff α [27]. Since the agreement was rather on a lower bound, we followed the current recommendations [28] and resolved disagreements between coders during research meetings, reaching consensus case by case. The initial criterion for coding misinformation was developed deductively based on the definition of misinformation used in this study. A post was coded as containing misinformation if it promoted an unproven therapy as a cancer-directed treatment. For example, a post claiming that an alkaline diet can eliminate cancer would be classified as misinformation: "Cure for cancer is an alkaline diet and lots of alkaline water." As reviewers worked with the data, they developed additional criteria based on observed patterns. Specifically, posts were labeled as containing misinformation if they discussed unproven approaches to prevent cancer, for example, "Pygeum Bark is nature's defense against prostate cancer." Furthermore, if a post contained a combination of factual and false information it was labeled as "misinformation."

Posts that were labeled as non-misinformation fell into 4 distinct categories. First, posts mentioned complementary and alternative medicine but did not promote it as a cancer treatment, for example, "Acupuncture and acupressure seem to be helpful in reducing pain and anxiety in patients having surgery." Second, posts that used sarcasm and actively debunked misinformation related to cancer were in the non-misinformation category, for example, "If what you stated is true, then Gerson treatment for cancer is false." The third category included posts that discussed complementary and alternative therapies but not in the context of promotion of cancer treatment, for instance, "Grapes can help protect you from the sun! Who knew?" Finally, posts that presented information with ambiguity, lack of clarity, or insufficient context were categorized as non-misinformation, for instance, "As a pancreatic cancer patient providing myself with all the additional holistic care practices made all the difference." The author did not specify whether his symptoms were alleviated or cancer progression was slowed down because of holistic practices. Therefore, the post was coded as non-misinformation.

Once a subset of the database was labeled by 2 reviewers (IF and RC), we applied an algorithm to populate labels to the entire database. We worked with BERT [29], a machine learning model for natural language processing. The BERT model was chosen because it (1) worked well with short, informal text [30]; (2) was shown to be applicable to medical text extracted from X [31]; and (3) was successfully used in previous research to identify misinformation on X [32]. The BERT model was implemented with the programming language Python (Python Software Foundation). The manually prelabeled subset served as training data for the BERT model. Such semisupervised approaches are commonly used in similar classification tasks [33]. After training, BERT used its understanding of the language and context learned from the large corpus it was originally trained on and the specific examples from the manually prelabeled dataset. BERT predicted labels for each post in the rest of the data (unlabeled dataset), determining whether each was likely to contain misinformation or not based on the patterns and features it learned from the manually coded dataset.

After BERT algorithm assigned labels to the posts, a researcher (IF), blinded to the model's results, manually coded a random subset of the posts (n=960) using the same "misinformation" and "non-misinformation" labels, adhering to the same criteria that were used to prelabel the data. When compared with manual coding, the algorithm identified misinformation with an accuracy of 83%, with a higher 86% specificity, and a slightly lower sensitivity of 82%. Upon labeling, 2 datasets were created and used in the first experiment: the misinformation dataset included only posts with misinformation, and control BERT dataset 1 included only posts with non-misinformation (Figure 1).

Data Collection: Posts From Cancer Centers

Following the definition of misinformation as "information not supported by scientific evidence or expert consensus" [34] and the definition used for this research, we assumed that posts originating from cancer centers reflect scientific evidence and expert consensus. To collect posts with factual information, we retrieved X data posted by cancer centers. Cancer centers often shared internal announcements and organizational news on X. To make posts comparable between the dataset with misinformation and control datasets, we used the keywords "cancer," "treatment," "chemotherapy," "healing," and other words related to treating cancer or controlling cancer progress. With the help of R software, we sampled 300 posts per cancer center between June 2011 and November 2022. A researcher (IF) manually checked randomly chosen (n=100) posts. As expected, no misinformation was found in the posts originating from cancer centers. The dataset, therefore, was assumed to consist of non-misinformation posts from cancer centers and was designated as control dataset 2, which was used in the second experiment alongside the misinformation dataset.

Linguistic Characteristics Modeling

Upon data collection and labeling, we used algorithmic approaches to model linguistic characteristics. First, we used an automated text search using regular expressions in Python [35] to capture digital numbers, hashtags, and URLs in the text.

Second, we used the Linguistic Inquiry and Word Count (LIWC) software [36]. LIWC calculates the proportion of the words in the posts associated with distinct psychological dimensions [37]. In this study, LIWC identified when authors of posts used certain, absolute, or tentative language.

Third, we leveraged the natural language processing tool, Name Entity Recognition [38], which was trained on human-labeled datasets to extract names from unstructured text. Using Name Entity Recognition, we were able to identify which posts contained personal names, organizational names, or locations identified from text.

Fourth, we experimented with several models for sentiment analysis and identified the DistilBERT algorithm as an optimal approach for its accuracy in handling health-related X data [39]. DistilBERT is a black-box algorithm that is trained on a large corpus of data and is based on multiple deep stack layers. The DistilBERT algorithm identified positive, negative, and neutral tones present in the posts. To evaluate the algorithm's performance, we manually labeled 300 posts across the databases. On average, the DistilBERT algorithm achieved an 83% accuracy (82% for misinformation and 84% for the control database) in detecting the emotional tone within the posts.

Linguistic Characteristics Testing: Prediction of Misinformation Labels

Identified linguistic characteristics were used in an algorithm to test whether these could distinguish misinformation in posts. As shown in Figure 1, we conducted 2 experiments using *tidymodels* package in R software [40]. Using linguistic characteristics as predictors, we forecast the "misinformation" and "non-misinformation" labels in the datasets semimanually coded by researchers and BERT classification algorithm. Data were split 60:40 to enable evaluation of the predictive power of the model and reported the accuracy as a ratio of correctly classified posts to the total number of posts. We also reported area under the curve (AUC), which accounted for both false-positive and false-negative predictions. AUC value ranged from 0 to 1, where 0.5 indicated that the model performs no better than a random chance, and 1 was a perfect prediction. The model was built on the basis of Lasso ("Least Absolute Shrinkage and Selection Operator") regression, which allowed variable selection by shrinking the coefficients of less important predictors to zero [41]. Bootstrapping procedure was applied to optimize and stabilize the selection of variables [42]. Lasso was chosen to address multicollinearity and overfitting issues in the regression model. More importantly, Lasso regression helped identify a set of linguistic characteristics that effectively distinguished posts containing misinformation. To evaluate the significance of specific linguistic characteristics, we computed importance scores, with higher scores indicating greater relevance in distinguishing posts containing misinformation. Importance scores, a common measure in predictive modeling, indicates to what extent individual predictors contribute to the overall model performance. The assessment involves permutating the characteristic values through shuffling and measuring the subsequent decline in model performance, effectively revealing the critical factors influencing predictions. Finally, we conducted a permutation statistical test (with 1000 permutations) to determine whether models with linguistic characteristics significantly outperformed random chance.

Results

Structured Literature Review

A total of 5677 citations were initially identified across all databases. After removing 1598 duplicates, we screened 4070 unique citations in Covidence. Subsequently, 3605 were excluded during the title and abstract review phase, leaving 464 papers for full-text review. Ultimately, we extracted linguistic characteristics from 88 full-text papers. These papers featured algorithmic approaches for identifying misinformation through automated text analysis, spanning various contexts, including politics, social issues, and computer science. Exclusion reasons are detailed in Figure 2, and additional information about the included papers can be found in Multimedia Appendix 3.

Identified Linguistic Characteristics

The extracted linguistic characteristics and corresponding literature are detailed in Table 1. Representative examples that contain each linguistic characteristic were chosen by selecting posts from the misinformation dataset. We used results from linguistic characteristic modeling to identify such posts. The first category of characteristics pertains to the sentiment and emotional expression in the text and includes positive, negative emotions, and neutral sentiments (absence of either). Some papers delved into more nuanced emotions such as anger, fear, surprise, and others. We excluded these emotions due to the potential difficulty for readers to detect nuanced emotions reliably in the text.

The next category comprises linguistic characteristics that pertain to psychological concepts. It is worth noting that some psychological concepts consist of a combination of linguistic characteristics, such as social processes including references to family, friends, other people, and verbs indicating interactions. Although algorithms frequently use such combinations, we decided to exclude the following psychological concepts that consisted of combinations of linguistic characteristics such as cognitive, perceptual, social processes, and morality or deception. The rationale behind this exclusion is that users are unlikely able to observe and combine linguistic characteristics for evaluations of the posts. We also excluded characteristics mentioned in fewer than 4 studies, such as gratitude, insight, causation, and persuasion. Following our 3 criteria, we included negations, tentativeness, profanity (as a proxy of informality), and words associated with absolutes and certainty.

Other categories that met our inclusion criteria were linguistic characteristics such as names of individuals, locations, and organizations, as well as categories related to the presence of URLs, hashtags, personal pronouns, and numbers. Readers can identify these characteristics without additional efforts (observability criterion) and use them for evaluation of the text (applicability) because the presence of these characteristics in social media has historically been a distinguishable factor in detecting misinformation. Furthermore, these characteristics were not context-dependent and, therefore, satisfy the generalizability criterion.

Table 1. Linguistic characteristics and examples of misinformation.

Characteristics	Examples of linguistic characteristics and posts with misinformation[a]	Studies using characteristics for misinformation detection
Sentiment[b]		[43-93]
Negative emotions	• Chemo is costly and very painful. It seems to worsen illness and hasten life’s end. • Sad this happened, to overcome cancer, consider utilizing cannabis oil in combination with vitamin B17. • Feeling frustrated that insurance doesn’t cover certain treatments I believe in. Wish there were more options beyond the conventional cut, burn, and poison approach.	
Positive emotions	• Cure for cancer that works holistically, Vitamin B17, very good! • Please do some heavy doses of medical organic marijuana if possible let it eat that cancer. Wishing you healing and joy and comfort. • Wonderful treatment! Discover the incredible benefits of ProstateRelax, a natural herbal treatment for prostate cancer. ProstateRelax effectively treats and prevents the progression of prostate cancer.	
Neutral emotions	• Anyone with cancer. Check your body’s pH level. Drink alkaline water, eat alkaline foods, and avoid acidic sugary treats and dairy. • Cancer cells thrive in low oxygen environments. B17, found in apricot seeds, can help. • Antineoplastons, a protein suppressed by cancer, could hold the key to a potential cure.	
Psycholinguistic		
Negation	• Unlock the potential of Acupuncture to modulate immunity and create an environment where cancer cannot thrive. Discover the holistic power of this ancient practice in bolstering your body's defenses against cancer. • I wonder why aren’t we utilizing hyperbaric chambers for Cancer? Ask your doctor about the incredible potential of pure oxygen in rejuvenating and generating new cells to combat this disease. • Don’t consume sugar (as cancer thrives on it), minimize or eliminate carb-rich foods like bread and pasta, and limit alcohol intake. Embrace the power of fasting to allow your body to heal itself.	[46,49,53,70,79,81,94-96]
Tentativeness	• 3 women with similar cancer, undergoing comparable treatments—2 passed away, but 1 is thriving Possible factor? She incorporated mistletoe & other non-pharma medicines into her regimen. • Concerns about [standard treatment] as a cancer solution persist, with claims of it being a harmful creation backed by influential medical forces. If it truly worked, wouldn’t it have been banned long ago like Laetrile? • Listen or not: Vitamin B17, found in Apricot seeds and sold online as a “health supplement,” has caught my attention as a potential cancer cure.	[49,51,59,61,62,66,81,94,96-100]
Absolute language or certainty	• I take sea buckthorn pills! They are an absolute lifesaver. • Vitamin B17 has definitely prevented my cancer from spreading. It's been a while, and there has been no growth. • During my time in a chemo clinic, alternative treatments were never allowed to be discussed or promoted. I left and started studying herbal medicine.	[43,51,59,61,94,97-101]
Profanity	• Create an alkaline environment that cancer can’t thrive in! Incorporate herbs, vitamins, and minerals to support your healing journey. You are going to heal and beat that s*** • Go to a poor country and you get real tea with real ginger. Go to a rich country and you will get chemical b**** that will give you cancer • It damages healthy cells, no surefire cancer cure. It's like a c*** shoot for survival & recurrence. But I choose a different path: starving cancerous cells with therapeutic fasting & lifestyle shifts.	[48,57,62,63,66,69,81,89,96,98,102]

Characteristics	Examples of linguistic characteristics and posts with misinformation[a]	Studies using characteristics for misinformation detection
Named entities		[44,49,51,60,64,69,79,93,103-109]
Names	• I watched the documentary of Dr. B [name] on YouTube. He cured stage 4 cancer with no chemotherapy and no radiation.	
Location	• Fascinating, study from M [name of State]! Certain sound frequencies may aid the body in fighting cancer. Pair this with an alkaline diet - and the world is cured!	
Organization	• Must-watch documentary on YouTube! Unveiling a shocking cancer cure cover-up for over 40 years! B [name]: The Cancer Cure Cover-Up—Full documentary available now!	
URL	• Insights from Dr. N [name]! Learn how to transform the cancer terrain, boost immunity, and create an inhospitable environment for cancer using Acupuncture, Chinese herbal medicines, and food therapies. Check out the discussion here: [link provided].	[45,51,52,54,55,62,69,78,79 ,86-88,92,93,98,99,101,104,107-117]
Numeric data	• Cancer is nearly 100% curable but beware of certain hospital treatments. Explore alternative options for better outcomes.	[44,49,51,57,65,67,70,72,73,79,81,94,98,101,105]
Pronouns	• I love your positivity and your fight against cancer. Keep up the fight and adhere to Alkaline Diet for a healthier journey. • Your cancer can be cured by #fasting paired with no sugar alkaline diet. • A pro basketball player revealed how organic Wheatgrass healed his close friend from blood cancer. A testament to the power of natural remedies!	[61,66,68,72,78,79,93,97,99,103,106,108,112,118-121]
Hashtag	• #TualangHoney helps against skin Cancer with no side effects.	[43,44,47,52-55,59,64,66,77-79,82,87, 92,96,98,101,104,107,108,111,115,119,122,123]

[a]All posts were paraphrased to protect the author's anonymity.

[b]In sentiment analysis, emotions are identified by a "black box" model (DistilBERT). While we report here examples and highlight "negative/positive" words in the sentence, we must acknowledge that the algorithm may or may not use these words for detecting emotions.

Collected Data From X

We collected a total of 45,791 posts related to unproven cancer therapies. Among these, 13,046 posts were labeled as misinformation (forming the misinformation dataset), while 32,745 posts were categorized as non-misinformation (comprising control dataset 1). Furthermore, we gathered 6782 posts from the profiles of comprehensive cancer centers, which were used as control dataset 2, as shown in Figure 1. The content description of both the misinformation dataset and the control dataset 1 is shown in Table 2. To illustrate the dataset in this study, we categorized the X posts into 9 distinct categories. The examples of the posts with misinformation are shown in Table 1.

Table 2. Relevant prevalence of therapy categories within posts about unproven cancer therapy.

Categories of therapies	Total posts, n	Posts with misinformation, n (%)[a]	Examples of unproven cancer therapy
Diet based	5179	3069 (59)	Antioxidant, fasting, and alkaline diet
Alternative health system	7036	2250 (32)	Herbal therapy and ayurveda
Plant- and fungus-based	13,851	4386 (32)	Mushrooms
Synthetic substances	8471	2637 (31)	Antineoplastic Brudzinski and vitamin C
Spiritual and mental healing	2347	272 (12)	Meditation, praying, and tai chi
Electromagnetic and energy-based	2825	283 (10)	Polarity therapy and magnetic
Physical procedures	1144	49 (4)	Acupuncture
Other	4938	100 (2)	N/A[b]
Total	45,791	13,046 (28)	N/A

[a]Out of the total number of posts.

[b]N/A: not applicable.

Linguistic Characteristics Testing: Prediction of Misinformation Labels

As shown in Table 3, experiment 1 demonstrated that linguistic characteristics predicted misinformation with 60% accuracy. In experiment 2, they exhibited even stronger predictive power, achieving an accuracy of 77%. The importance scores for each linguistic characteristic are shown in Table 4.

Next, we selected linguistic characteristics with an impact score 0.05 and consistent predictive performance across experiments 1 and 2. These short-listed characteristics underwent further testing within the same experiments. In experiment 1, the short-listed characteristics achieved an accuracy rate of 50%, which did not significantly differ from random chance (P>.90). However, in experiment 2, these characteristics predicted misinformation with an accuracy rate of 73% and an AUC of 83. This performance was significantly better than random chance (McNemar $^2{}_1$=5.7 $\times 10^7$; P<.001). The importance scores for the short-listed characteristics are shown in Table 4. For a more detailed breakdown of the importance scores, we have summarized the percentage of posts containing these short-listed characteristics by dataset in Table 4 and the complete list in Multimedia Appendix 4.

Table 3. Lasso regression performance.

Name of the dataset	Total posts, n	Posts with misinformation, n	Accuracy, %
Experiment 1: misinformation dataset and control dataset 1	45,791	13,046	60
Experiment 2: misinformation dataset and control dataset 2	19,828	13,046	77

Table 4. Importance scores.

Linguistic characteristics	Experiment with control group 1		Experiment with control group 2		Experiment with short-listed characteristics (control group 2)	
	Predictors		Predictors		Predictors	
	Negative	Positive	Negative	Positive	Negative	Positive
Absolute language	—[a]	*0.11*[b]	—	*0.69*	—	*0.84*
Certainty	—	*0.21*	—	*1.13*	—	*1.02*
First-person pronoun	0.27	—	—	1.31	—	—
Hashtags	*0.56*	—	*1.55*	—	*1.6*	—
Location	*0.27*	—	*0.27*	—	*0.46*	—
Name	—	0.08	0.91	—	—	—
Negation	0.53	—	—	0.73	—	—
Negative emotions	0.24	—	0	—	—	—
Neutral emotions	0	—	—	0.07	—	—
Number	—	*0.17*	—	*0.29*	—	*0.28*
Organization	—	0.02	0.63	—	—	—
Positive emotions	—	0.31	0.46	—	—	—
Profanity	0.92	—	—	1.99	—	—
Second-person pronoun	0.02	—	0.45	—	—	—
Tentativeness	*0.08*	—	*0.16*	—	*0.08*	—
Third-person pronoun	0	—	0.23	—	—	—
URL	*0.3*	—	*2.28*	—	*2.47*	—

[a]Not applicable.

[b]Italicized values represent short-listed characteristics.

Table 5. The percentage of posts with short-listed linguistic characteristics.

Linguistic characteristics	Misinformation dataset (n=13,046), n (%)	Control dataset 1 (n=32,745), n (%)	Control dataset 2 (n=6782), n (%)
Positive predictors			
Certainty	1579 (12)	3044 (9)	208 (3)
Absolute	2741 (21)	7294 (22)[a]	630 (9)
Number	6358 (49)	14,360 (44)	2497 (37)
Negative predictors			
URL	6978 (53)	19,591 (60)	6560 (97)
Hashtags	2296 (18)	8512 (26)	4343 (64)
Location	1212 (9)	3373 (12)	975 (14)
Tentativeness	4154 (32)	11,171 (34)	1835 (27)[a]

[a]Valence of predictions is inferred from the model, which includes all characteristics simultaneously.

Discussion

Principal Findings

We have identified linguistic characteristics that can help people affected by cancer detect cancer misinformation on social media platforms such as X. Linguistic characteristics that were *likely* to be present in posts with misinformation were related to certain, absolute language, and numbers. Certain language included phrases that reflected a "degree of bravado" or "boasting of certainty." Examples of certain languages could be "I really believe," "it is definitely helpful," and similar others [36]. The absolute language referred to phrases that reflect black-and-white thinking and included words such as "none," "all," "never," and others [36]. The number category encompassed any information reported with digits such as percentages, count of any units, years, and priorities. Notably, all 3 linguistic characteristics could be united under the umbrella of definite, confident language. Linguistic characteristics that were *unlikely* to be present in posts with misinformation encompassed URLs, hashtags, and location mentions. Each of these attributes could be considered as a form of citation or reference. URLs offered direct links to the original source or further information, hashtags connected posts to broader relevant discussions, while locations mentioned in posts provided context and a sense of origin to the information shared. Our findings are consistent with some of the suggestions provided by previous guidelines for identifying misinformation. For instance, the Food and Drug Administration recommends being vigilant if patients read confident statements such as a drug definitely "cures cancer" or "guarantees results" [124]. Other guidelines encouraged users to search for references and original sources of health-related information [12-14].

While consistent with previous recommendations, our findings make a unique contribution. Previous work has based the guidelines on theoretical assumptions, while our study is one of the first to provide some empirical evidence based on a large dataset to support the recommendations for users. Another contribution is that we outlined ineffective linguistic characteristics for detecting cancer misinformation. Despite a substantial body of research showing that social media posts with sentiments predicted fake news, we did not find these relationships. A potential explanation could be the algorithm's limited efficiency in identifying emotions within cancer-related contexts. Furthermore, it is possible that authors express a limited range of emotions in cancer-related conversations, typically negative emotions toward cancer and both positive and negative emotions toward various treatments, including those that are unproven. These emotions may vary little across posts containing valid and nonvalid information, making emotions an unreliable factor for distinguishing misinformation.

Our work accumulates knowledge about misinformation detection from the literature covering a wide range of contexts—including political, social, and computer science—and translates this knowledge to the cancer context. The findings highlighted promising avenues for future research and could expedite the development of automated and augmented methods for identifying and verifying cancer-related misinformation on social media platforms. Finally, the robust labeled datasets developed by our research team are available to other researchers upon request to the corresponding author, thereby further supporting research on misinformation within the context of cancer and social media.

In practice, our work is at the forefront of customizing recommendations and contextualizing them for social network users. Our exploratory findings suggest a promising direction for studying linguistic characteristics that information users might apply when making quick judgments while scrolling through X feeds. Empowering users to stay vigilant in their initial evaluations could help reduce the spread of misinformation and the formation of erroneous beliefs. This is a crucial area for future research, which should explore how these findings apply in different cancer-related contexts and across various social networks.

Limitations

All the studies included in our analysis exclusively originate from peer-reviewed journals and conference proceedings; however, we must exercise caution when considering the potential for publication bias. Furthermore, in accordance with our selection criteria for linguistic characteristics, we included only those papers that focused on text and excluded other forms of social media content, such as videos and images. We recommend that future research comprehensively explore social media, including multimedia content, as it could potentially provide additional insights for user-friendly recommendations.

In selecting linguistic characteristics, we prioritized observability, applicability, and generalizability. However, alternative criteria may be considered when users are open to a more thorough exploration of a post's validity. For example, future research should explore the use of metadata, link content analysis, and hashtag meanings. As misinformation evolves and its authors adjust to societal changes, the linguistic characteristics that identify misinformation may also shift. A longitudinal analysis is necessary to understand how linguistic characteristics perform in predicting misinformation over time.

Algorithms used in our analysis operate with a certain level of accuracy. Specifically, the accuracy of label identification in the dataset reached 83%, indicating that approximately 17% of posts were labeled incorrectly. This means that in experiment 1 some proportion of misinformation is included in the non-misinformation group and vice versa, making further exploration less accurate in experiment 1. This degree of uncertainty is common in algorithmic performance. Therefore, it is important to interpret our results in light of the inherent imperfections in algorithmic performance.

Furthermore, we encountered that the short-listed linguistic characteristics did not significantly outperform random chance in identifying misinformation in experiment 1. This outcome underscores a potential boundary condition of the effectiveness of the linguistic characteristics. Notably, experiment 1 encompassed more homogeneous data in contrast to experiment 2. Based on these findings, it becomes plausible to speculate that linguistic characteristics might provide limited help when a reader assesses posts within a closely knit community.

In experiment 2, the control dataset 2 consisted of posts shared by cancer centers and was compared with the misinformation dataset comprising random posts. To address this limitation, we collected posts from cancer centers that contain words related to cancer therapies. This step was taken to ensure a similar context of discussion as the posts with unproven therapy. Next, we exclude linguistic characteristics that are likely displayed differences between datasets due to the distinct nature of the information within control dataset 2. For example, linguistic traits such as “the use of profanity” or “first-person pronouns” were discarded. Furthermore, we decided to focus our analysis solely on the text within the posts and omitted other accompanying metainformation that users might observe, such as the user's name, location of the author, and posting time. This approach allowed us to assume that posts shared by cancer centers might be perceived more broadly, for instance, as posts shared by researchers, physicians, administrators, and patient advocates. Because of these measures, we anticipate that the linguistic characteristics identified in this research may help differentiate between health misinformation and factual posts on social media, irrespective of their sources. Despite our precautionary measures, we cannot fully guarantee that identified linguistics characteristics certainly distinguish between posts with misinformation and non-misinformation versus posts produced by the general public and posts by health experts from health care systems. However, there are factors that support the first conclusion more than the second. First, our findings are consistent with the previous theoretical and practical recommendations for identifying misinformation [12-14]. Second, the associated with misinformation linguistic characteristics, such as numbers and assertive language, are expected to be used by health experts. For instance, providers use numbers more confidently than the general public [125]. Professional guidelines for health providers encourage them to use numbers over verbal descriptions [126] as well as the use of assertive language in communication with patients [127,128]. Yet, our study associated these characteristics with misinformation shared by the general public on social media, which suggests that we might be finding more than just a mere distinction between the general public language and the health professional language. One study in and of itself is not yet a comprehensive body of evidence. Our findings will need to be validated and built upon via additional studies—including those that use posts from other types of entities and comparison groups.

Finally, our data were collected only on a single social network X. Many characteristics and customs of X are transferable to other social networks and our recommendations are likely to go beyond application on X, as demonstrated by the consistency of our recommendations with the recommendations of other researchers [12-14]. Given this limitation, our results need to be generalized cautiously, and further similar research is needed for different platforms (eg, Facebook, Pinterest, etc).

Conclusions

Our structured review synthesized knowledge from studies that used algorithmic approaches for text analysis to detect misinformation in social media. From this literature, we identified user-friendly linguistic characteristics that can assist individuals in distinguishing misinformation when they seek health-related information on social media. The linguistic characteristics, such as certainty, absolute language, and numbers, were positively associated with misinformation, while characteristics such as URLs, hashtags, and location mentions were negatively predictive of misinformation. Based on these findings, we suggested that users should be cautious of social media posts containing confident promises or specific numbers without proper references to the original information. According to our analysis, we expect that this approach will allow users to filter out two-thirds of posts with cancer-related misinformation. Yet, before drawing a definitive conclusion, further testing with different datasets is required.

Acknowledgments

ChatGPT 3.5 (OpenAI) [129] was used to assist with the professional editing of the manuscript. This study was supported by North Carolina Translational Research and Clinical Science Institute, Pilot Award Spring 2022.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Strategy for literature review.
[DOCX File , 29 KB - infodemiology_v5i1e62703_app1.docx]

Multimedia Appendix 2
List of unproven therapy.
[DOCX File , 25 KB - infodemiology_v5i1e62703_app2.docx]

Multimedia Appendix 3
Summary of the literature.
[DOCX File , 47 KB - infodemiology_v5i1e62703_app3.docx]

Multimedia Appendix 4
Summary of linguistic characteristics.
[XLSX File (Microsoft Excel File), 10 KB - infodemiology_v5i1e62703_app4.xlsx]

References

1. How much do you agree or disagree with the following statements? I use information from social media to make decisions about my health. HINTS. 2022. URL: https://hints.cancer.gov/view-questions/question-detail.aspx?PK_Cycle=14&qid=1849 [accessed 2023-09-20]
2. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. J Med Internet Res 2021;23(1):e17187 [FREE Full text] [doi: 10.2196/17187] [Medline: 33470931]
3. Johnson SB, Park HS, Gross CP, Yu JB. Use of alternative medicine for cancer and its impact on survival. J Natl Cancer Inst 2017;110:121-124.
4. Johnson SB, Park HS, Gross CP, Yu JB. Complementary medicine, refusal of conventional cancer therapy, and survival among patients with curable Cancers. JAMA Oncol 2018;4(10):1375-1381 [FREE Full text] [doi: 10.1001/jamaoncol.2018.2487] [Medline: 30027204]
5. Weil CR, Hutten RJ, Barney BM, Fagerlin A, Gaffney DK, Gill DM, et al. Shifting perceptions of alternative therapies in cancer patients during the COVID-19 pandemic: results from the complementary and alternative medicine exposure in oncology (CAMEO) study. J. Clin. Oncol 2022;40(16_suppl):e24130-e24130. [doi: 10.1200/jco.2022.40.16_suppl.e24130]
6. Johnson SB, Bylund CL. Identifying cancer treatment misinformation and strategies to mitigate its effects with improved radiation oncologist-patient communication. Pract Radiat Oncol 2023;13(4):282-285. [doi: 10.1016/j.prro.2023.01.007] [Medline: 36736620]
7. Johnson SB, Parsons M, Dorff T, Moran MS, Ward JH, Cohen SA, et al. Cancer misinformation and harmful information on facebook and other social media: a brief report. J Natl Cancer Inst 2022;114(7):1036-1039 [FREE Full text] [doi: 10.1093/jnci/djab141] [Medline: 34291289]
8. Alsyouf M, Stokes P, Hur D, Amasyali A, Ruckle H, Hu B. 'Fake News' in urology: evaluating the accuracy of articles shared on social media in genitourinary malignancies. BJU Int 2019;124(4):701-706. [doi: 10.1111/bju.14787] [Medline: 31044493]
9. Yussof I, Ab Muin NF, Mohd M, Hatah E, Mohd Tahir NA, Mohamed Shah N. Breast cancer prevention and treatment misinformation on Twitter: an analysis of two languages. Digit Health 2023;9:20552076231205742 [FREE Full text] [doi: 10.1177/20552076231205742] [Medline: 37808244]
10. Chen L, Wang X, Peng T. Nature and diffusion of gynecologic cancer–related misinformation on social media: analysis of tweets. J Med Internet Res 2018;20(10):e11515. [doi: 10.2196/11515] [Medline: 30327289]
11. Lazard AJ, Nicolla S, Vereen RN, Pendleton S, Charlot M, Tan H, et al. Exposure and reactions to cancer treatment misinformation and advice: survey study. JMIR Cancer 2023;9:e43749 [FREE Full text] [doi: 10.2196/43749] [Medline: 37505790]
12. Health misinformation. The US Surgeon General's Advisory. 2021. URL: https://www.hhs.gov/surgeongeneral/priorities/health-misinformation/index.html [accessed 2023-09-05]

13. Finding cancer information on the internet. American Cancer Society. 2020. URL: https://www.cancer.org/cancer/understanding-cancer/cancer-information-on-the-internet.html [accessed 2023-07-20]
14. How to find cancer resources you can trust. National Cancer Institute Internet. 2015. URL: https://www.cancer.gov/about-cancer/managing-care/using-trusted-resources [accessed 2022-06-28]
15. Bondielli A, Marcelloni F. A survey on fake news and rumour detection techniques. Inf Sci 2019;497:38-55. [doi: 10.1016/j.ins.2019.05.035]
16. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J. Lang. Soc. Psychol 2009;29(1):24-54. [doi: 10.1177/0261927x09351676]
17. Kapantai E, Christopoulou A, Berberidis C, Peristeras V. A systematic literature review on disinformation: toward a unified taxonomical framework. New Media Soc 2020;23(5):1301-1326. [doi: 10.1177/1461444820959296]
18. Zannettou S, Sirivianos M, Blackburn J, Kourtellis N. The web of false information. ACM J Data Inf Qual 2019;11(3):1-37. [doi: 10.1145/3309699]
19. Sendra A, Farré J, Vaagan RW. Seeking, sharing and co-creating: a systematic review of the relation between social support theory, social media use and chronic diseases. Soc Theory Health 2019;18(4):317-339. [doi: 10.1057/s41285-019-00106-z]
20. Wang Y, McKee M, Torbica A, Stuckler D. Systematic literature review on the spread of health-related misinformation on social media. Soc Sci Med 2019;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
21. Molina MD, Sundar SS, Le T, Lee D. "Fake News" is not simply false information: a concept explication and taxonomy of online content. Am Behav Sci 2019;65(2):180-212. [doi: 10.1177/0002764219878224]
22. Babineau J. Product review: Covidence (systematic review software). J Can Health Libr Assoc 2014;35:68-71.
23. Ghenai A, Mejova Y. Fake cures. Proc ACM Hum Comput Interact 2018;2:1-20. [doi: 10.1145/3274327]
24. The Free Encyclopedia. List of unproven and disproven cancer treatments. Wikipedia [Internet]. Wikimedia Foundation, Inc. 2013. URL: https://en.wikipedia.org/wiki/List_of_unproven_and_disproven_cancer_treatments [accessed 2023-07-21]
25. Illegally sold cancer treatments. US Food and Drug Administration. 2017. URL: https://www.fda.gov/consumers/health-fraud-scams/illegally-sold-cancer-treatments [accessed 2023-09-01]
26. White RW, Hassan A. Content bias in online health search. ACM Trans Web 2014;8(4):1-33. [doi: 10.1145/2663355]
27. Marzi G, Balzano M, Marchiori D. K-Alpha Calculator-Krippendorff's Alpha Calculator: a user-friendly tool for computing Krippendorff's alpha inter-rater reliability coefficient. MethodsX 2024;12:102545 [FREE Full text] [doi: 10.1016/j.mex.2023.102545] [Medline: 39669968]
28. Oortwijn Y, Ossenkoppele T, Betti A. Interrater disagreement resolution: a systematic procedure to reach consensus in annotation tasks. 2021 Presented at: Proceedings of the Workshop on Human Evaluation of NLP Systems (HumEval); 2021 April 01; Dublin, Ireland p. 131-141.
29. Minaee S, Kalchbrenner N, Cambria E, Nikzad N, Chenaghlu M, Gao J. Deep learning--based text classification. ACM Comput Surv 2021;54(3):1-40. [doi: 10.1145/3439726]
30. Qasim R, Bangyal WH, Alqarni MA, Ali Almazroi A. A fine-tuned BERT-based transfer learning approach for text classification. J Healthc Eng 2022;2022:3498123 [FREE Full text] [doi: 10.1155/2022/3498123] [Medline: 35013691]
31. Roitero K, Bozzato C, Della MV, Mizzaro S, Serra G. Twitter goes to the doctor: detecting medical tweets using machine learning and BERT. 2020. URL: http://hdl.handle.net/11390/1189151 [accessed 2023-07-12]
32. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv Preprint posted online on October 11, 2018. [doi: 10.48550/ARXIV.1810.04805]
33. Elroy O, Yosipof A. Semi-supervised learning classifier for misinformation related to earthquakes prediction on social media. 2023 Presented at: Artificial Neural Networks and Machine Learning – ICANN 2023: 32nd International Conference on Artificial Neural Networks; September 26-29, 2023; Heraklion, Greece. [doi: 10.1007/978-3-031-44207-0_22]
34. Swire-Thompson B, Lazer D. Public health and online misinformation: challenges and recommendations. Annu Rev Public Health 2020;41:433-451 [FREE Full text] [doi: 10.1146/annurev-publhealth-040119-094127] [Medline: 31874069]
35. Chapman C, Stolee K. Exploring regular expression usage and context in Python. 2016 Presented at: Proceedings of the 25th International Symposium on Software Testing and Analysis; July 18, 2016; New York, NY. [doi: 10.1145/2931037.2931073]
36. Boyd RL, Ashokkumar A, Seraj S, Pennebaker JW. The development and psychometric properties of LIWC-22. 2022. URL: https://www.liwc.app [accessed 2023-07-19]
37. Tausczik YR, Pennebaker JW. The psychological meaning of words: LIWC and computerized text analysis methods. J Lang Soc Psychol 2009;29(1):24-54. [doi: 10.1177/0261927x09351676]
38. Goyal A, Gupta V, Kumar M. Recent named entity recognition and classification techniques: a systematic review. Comput. Sci. Rev 2018;29:21-43. [doi: 10.1016/j.cosrev.2018.06.001]
39. Ranganathan J, Tsahai T. Sentiment analysis of tweets using deep learning. In: Advanced Data Mining and Applications. Cham, Switzerland: Springer International Publishing AG; 2022:106-117.
40. Kuhn M, Wickham H. Tidymodels: a collection of packages for modeling and machine learning using tidyverse principles. 2020. URL: https://www.tidymodels.org [accessed 2023-07-24]
41. Bach F. Consistency of the group Lasso and multiple kernel learning. J Mach Learn Res 2008;9:1179-1225.

42. Bach FR. Bolasso: model consistent Lasso estimation through the bootstrap. 2008 Presented at: Proceedings of the 25th International Conference on Machine Learning; July 5, 2008; Helsinki, Finland p. 33-40. [doi: 10.1145/1390156.1390161]
43. Agarwal S, Samavedhi A. Profiling fake news: learning the semantics and characterisation of misinformation. In: Advanced Data Mining and Applications. ADMA 2022. Lecture Notes in Computer Science, vol 13087. Cham, Switzerland: Springer International Publishing; 2022:203-216.
44. Agarwal R, Gupta S, Chatterjee N. Profiling fake news spreaders on Twitter: a clickbait and linguistic feature based scheme. In: Natural Language Processing and Information Systems. New York City, NY: Springer International Publishing; 2022:345-357.
45. Alkhodair SA, Fung BCM, Ding SHH, Cheung WK, Huang S. Detecting high-engaging breaking news rumors in social media. ACM Trans Manage Inf Syst 2020;12(1):1-16. [doi: 10.1145/3416703]
46. Garg S, Kumar Sharma D. Linguistic features based framework for automatic fake news detection. Comput Ind Eng 2022;172(Part A):108432. [doi: 10.1016/j.cie.2022.108432]
47. Agrawal C, Pandey A, Goyal S. Fake news detection system based on modified bi-directional long short term memory. Multimed Tools Appl 2022;81(17):24199-24223. [doi: 10.1007/s11042-022-12772-9]
48. Ahmad I, Yousaf M, Yousaf S, Ahmad MO. Fake news detection using machine learning ensemble methods. Complexity 2020;2020:1-11 [FREE Full text] [doi: 10.1155/2020/8885861]
49. Alsmadi I, O'Brien MJ. Rating news claims: feature selection and evaluation. Math Biosci Eng 2019;17(3):1922-1939 [FREE Full text] [doi: 10.3934/mbe.2020101] [Medline: 32233515]
50. Aneja S, Aneja N, Kumaraguru P. Predictive linguistic cues for fake news: a societal artificial intelligence problem. arXiv Preprint posted online on November 26, 2022 [FREE Full text]
51. Ebrahimi Fard A, Mohammadi M, Chen Y, van de Walle B. Computational rumor detection without non-rumor: a one-class classification approach. IEEE Trans Comput Soc Syst 2019;6(5):830-846. [doi: 10.1109/tcss.2019.2931186]
52. Sharma AK, Gangeya RA, Kumar H, Chaurasia S, Srivastava DK. Ensemble rumor text classification model applied to different tweet features. Int J Fuzzy Log Intell Syst 2022;22(3):325-338. [doi: 10.5391/ijfis.2022.22.3.325]
53. Sharma D, Shrivastava P, Garg S. Utilizing word embedding and linguistic features for fake news detection. 2022 Presented at: 9th International Conference on Computing for Sustainable Global Development (INDIACom); March 23-25, 2022; New Delhi, India. [doi: 10.23919/indiacom54597.2022.9763294]
54. Shelke S, Attar V. Rumor detection in social network based on user, content and lexical features. Multimed Tools Appl 2022;81(12):17347-17368 [FREE Full text] [doi: 10.1007/s11042-022-12761-y] [Medline: 35282405]
55. Shelke S, Attar V. Role of various features in identification of rumors in the social network. 2021 Presented at: 12th International Conference on Computing Communication and Networking Technologies (ICCCNT); July 06-08, 2021; Kharagpur, India. [doi: 10.1109/icccnt51525.2021.9579856]
56. Sicilia R, Giudice S, Pei Y, Pechenizkiy M, Soda P. Health-related rumour detection on Twitter. 2017 Presented at: IEEE International Conference on Bioinformatics and Biomedicine (BIBM); November 13-16, 2017; Kansas City, MO. [doi: 10.1109/BIBM.2017.8217899]
57. Sun QJ. A machine learning analysis of the features in deceptive and credible news. arXiv Preprint posted online on October 5, 2019. [doi: 10.48550/arXiv.1910.02223]
58. Tuarob S, Tucker CS, Salathe M, Ram N. An ensemble heterogeneous classification methodology for discovering health-related knowledge in social media messages. J Biomed Inform 2014;49:255-268 [FREE Full text] [doi: 10.1016/j.jbi.2014.03.005] [Medline: 24642081]
59. van de Guchte L, Raaijmakers S, Meeuwissen E, Spenader J. Near real-time detection of misinformation on online social networks. In: Disinformation in Open Online Media. Cham, Switzerland: Springer International Publishing; 2020:240-260.
60. Vijeev A, Mahapatra A, Shyamkrishna A, Murthy S. A hybrid approach to rumour detection in microblogging platforms. 2018 Presented at: International Conference on Advances in Computing, Communications and Informatics (ICACCI); September 19-22, 2018; Bangalore, India. [doi: 10.1109/ICACCI.2018.8554371]
61. Volkova S, Jang JY. Misleading or falsification: inferring deceptive strategies and types in online news and social media. 2018 Presented at: Companion Proceedings of The Web Conference 2018; April 23, 2018; Lyon, France p. 575-583. [doi: 10.1145/3184558.3188728]
62. Xuan K, Xia R. Rumor stance classification via machine learning with text, user and propagation features. 2019 Presented at: International Conference on Data Mining Workshops (ICDMW); November 8-11, 2019; Beijing, China. [doi: 10.1109/icdmw.2019.00085]
63. Zhou X, Jain A, Phoha VV, Zafarani R. Fake news early detection: an interdisciplinary study. arXiv Preprint posted online on April 26, 2019. [Medline: 10.48550/ARXIV.1904.11679]
64. Wang L, Wang Y, de Melo G, Weikum G. Five shades of untruth: finer-grained classification of fake news. 2018 Presented at: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM); August 28-31, 2018; Barcelona, Spain. [doi: 10.1109/asonam.2018.8508256]
65. Gautam A, Jerripothula KR. SGG: spinbot, grammarly and glove based fake news detection. 2020 Presented at: IEEE Sixth International Conference on Multimedia Big Data (BigMM); September 24-26, 2020; New Delhi, India. [doi: 10.1109/bigmm50055.2020.00033]

66. Mahbub S, Pardede E, Kayes ASM. COVID-19 rumor detection using psycho-linguistic features. IEEE Access 2022;10:117530-117543. [doi: 10.1109/access.2022.3220369]
67. Niven T, Kao HY, Wang HY. Profiling spreaders of disinformation on Twitter: IKMLab and softbank submission. CLEF (Working Notes) 2020;2696:1-8 [FREE Full text]
68. Mondal T, Pramanik P, Bhattacharya I, Boral N, Ghosh S. Analysis and early detection of rumors in a post disaster scenario. Inf Syst Front 2018;20(5):961-979. [doi: 10.1007/s10796-018-9837-8]
69. Basharat S, Ahmad M. Inferring trust from message features using linear regression and support vector machines. In: Communications in Computer and Information Science. Singapore: Springer Singapore; 2018:577-598.
70. Butt S, Sharma S, Sharma R, Sidorov G, Gelbukh A. What goes on inside rumour and non-rumour tweets and their reactions: a psycholinguistic analyses. Comput Human Behav 2022;135:107345. [doi: 10.1016/j.chb.2022.107345]
71. Casillo M, Colace F, Conte D, de Santo M, Lombardi M, Mottola S. A multi-feature bayesian approach for fake news detection. In: Computational Data and Social Networks. Cham, Switzerland: Springer International Publishing; 2020:333-344.
72. Castillo JM, Fadera KDF, Ladao AAA, Go JG, Tamayo MB, Octaviano MV. Fake news detection on English news article?s title. 2021 Presented at: 1st International Conference in Information and Computing Research (iCORE); December 11-12, 2021; Manila, Philippines. [doi: 10.1109/icore54267.2021.00046]
73. Cerbin L, DeJesus J, Warnken J, Gokhale S. Unmasking the mask debate on social media. 2021 Presented at: IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC); July 12-16, 2021; Madrid, Spain. [doi: 10.1109/compsac51774.2021.00098]
74. Choudhary A, Arora A. Linguistic feature based learning model for fake news detection and classification. Expert Syst Appl 2021;169:114171. [doi: 10.1016/j.eswa.2020.114171]
75. Choudhry A, Khatri I, Jain M, Vishwakarma DK. An emotion-aware multitask approach to fake news and rumor detection using transfer learning. IEEE Trans Comput Soc Syst 2024;11(1):588-599. [doi: 10.1109/tcss.2022.3228312]
76. Felber T. Constraint 2021: machine learning models for COVID-19 fake news detection shared task. arXiv Preprint posted online on January 11, 2021 . [doi: 10.48550/arXiv.2101.03717]
77. Ghanem B, Ponzetto SP, Rosso P. FacTweet: profiling fake news Twitter accounts. In: Statistical Language and Speech Processing. Cham, Switzerland: Springer International Publishing; 2020:35-45.
78. Jogalekar NS, Attar V, Palshikar GK. Rumor detection on social networks: a sociological approach. 2020 Presented at: IEEE International Conference on Big Data (Big Data); December 10-13, 2020; Atlanta, GA. [doi: 10.1109/bigdata50022.2020.9378149]
79. Kumar A, Sangwan SR, Nayyar A. Rumour veracity detection on Twitter using particle swarm optimized shallow classifiers. Multimed Tools Appl 2019;78(4):24083-24101. [doi: 10.1007/s11042-019-7398-6]
80. Kumari R, Ashok N, Ghosal T, Ekbal A. A multitask learning approach for fake news detection: novelty, emotion, and sentiment lend a helping hand. 2021 Presented at: International Joint Conference on Neural Networks (IJCNN); July 18-22, 2021; Shenzhen, China. [doi: 10.1109/ijcnn52387.2021.9534218]
81. Kwon S, Cha M, Jung K. Rumor detection over varying time windows. PLoS One 2017;12(1):e0168344 [FREE Full text] [doi: 10.1371/journal.pone.0168344] [Medline: 28081135]
82. Li C, Chun SA, Geller J. Stemming the tide of fake news about the COVID-19 pandemic. Int FLAIRS Conf Proc 2022;35 [FREE Full text]
83. Rajabi Z, Shehu A, Purohit H. User behavior modelling for fake information mitigation on social web. In: Social, Cultural, and Behavioral Modeling. Cham, Switzerland: Springer International Publishing; 2019.
84. Sandrilla R, Devi MS. FNU-BiCNN: fake news and fake URL detection using Bi-CNN. Int J Adv Comput Sci Appl 2022;13(2) [FREE Full text]
85. Seddari N, Derhab A, Belaoued M, Halboob W, Al-Muhtadi J, Bouras A. A hybrid linguistic and knowledge-based analysis approach for fake news detection on social media. IEEE Access 2022;10:62097-62109. [doi: 10.1109/access.2022.3181184]
86. Singh JP, Kumar A, Rana NP, Dwivedi YK. Attention-based LSTM network for rumor veracity estimation of tweets. Inf Syst Front 2020;24(2):459-474. [doi: 10.1007/s10796-020-10040-5]
87. Singh JP, Rana NP, Dwivedi YK. Rumour veracity estimation with deep learning for Twitter. In: FIP Advances in Information and Communication Technology. Cham, Switzerland: Springer International Publishing; 2019.
88. Slimi H, Bounhas I, Slimani Y. URL-based tweet credibility evaluation. 2019 Presented at: IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA); November 3-7, 2019; Abu Dhabi, United Arab Emirates. [doi: 10.1109/aiccsa47632.2019.9035255]
89. Sotirakou C, Karampela A, Mourlas C. Evaluating the role of news content and social media interactions for fake news detection. In: Bright J, Giachanou A, Spaiser V, Spezzano F, George A, Pavliuc A, editors. Disinformation in Open Online Media. MISDOOM 2021. Lecture Notes in Computer Science, vol 12887. Cham, Switzerland: Springer International Publishing; 2021.
90. Verma PK, Agrawal P, Amorim I, Prodan R. WELFake: word embedding over linguistic features for fake news detection. IEEE Trans Comput Soc Syst 2021;8(4):881-893. [doi: 10.1109/tcss.2021.3068519]

91. Kula S, Choraś M, Kozik R, Ksieniewicz P, Woźniak M. Sentiment analysis for fake news detection by means of neural networks. In: Krzhizhanovskaya VV, Závodszky G, Lees MH, Dongarra JK, Sloot PMA, Brissos S, et al, editors. Lecture Notes in Computer Science. Cham, Switzerland: Springer International Publishing; 2020.
92. Vadavalli A, Al E. Feed forward neural network based effective feature extraction technique for better classification accuracy. TURCOMAT 2021;12(11):4576-4588.
93. Saikh T, Anand A, Ekbal A, Bhattacharyya P. A novel approach towards fake news detection: Deep learning augmented with textual entailment features. In: Natural Language Processing and Information Systems. Cham, Switzerland: Springer International Publishing; 2019:345-358.
94. Gravanis G, Vakali A, Diamantaras K, Karadais P. Behind the cues: a benchmarking study for fake news detection. Expert Syst Appl 2019;128:201-213. [doi: 10.1016/j.eswa.2019.03.036]
95. Hassan FM, Lee M. Political fake statement detection via multistage feature-assisted neural modeling. 2020 Presented at: IEEE International Conference on Intelligence and Security Informatics (ISI); November 9-10, 2020; Arlington, VA. [doi: 10.1109/isi49825.2020.9280531]
96. Patro J, Rathore PS. A sociolinguistic route to the characterizationdetection of the credibility of events on Twitter. 2020 Presented at: Proceedings of the 31st ACM Conference on Hypertext and Social Media; July 15, 2020; Virtual event, United States. [doi: 10.1145/3372923.3404795]
97. Kumar JM, Gopalani D, Kumar MY, Kumar R. Machine learning based fake news detection using linguistic features and word vector features. 2020 Presented at: 2020 IEEE 7th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON); November 27-29, 2020; Prayagraj, India. [doi: 10.1109/upcon50219.2020.9376576]
98. Rastogi S, Bansal D. Disinformation detection on social media: an integrated approach. Multimed Tools Appl 2022;81(28):40675-40707 [FREE Full text] [doi: 10.1007/s11042-022-13129-y] [Medline: 35582207]
99. Reddy H, Raj N, Gala M, Basava A. Text-mining-based fake news detection using ensemble methods. Int J Autom Comput 2020;17(2):210-221. [doi: 10.1007/s11633-019-1216-5]
100. Volkova S, Shaffer K, Jang JY, Hodas N. Separating facts from fiction: linguistic models to classify suspicious trusted news posts on Twitter. 2017 Presented at: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers); December 24, 2024; Stroudsburg, PA. [doi: 10.18653/v1/p17-2102]
101. Lotfi S, Mirzarezaee M, Hosseinzadeh M, Seydi V. Rumor conversations detection in Twitter through extraction of structural features. Inf Technol Manag 2021;22(4):265-279. [doi: 10.1007/s10799-021-00335-7]
102. Khanam Z, Alwasel BN, Sirafi H, Rashid M. Fake news detection using machine learning approaches. IOP Conf Ser Mater Sci Eng 2021;1099(1):012040. [doi: 10.1088/1757-899x/1099/1/012040]
103. Braşoveanu AMP, Andonie R. Integrating machine learning techniques in semantic fake news detection. Neural Process Lett 2020;53(5):3055-3072. [doi: 10.1007/s11063-020-10365-x]
104. Hamidian S, Diab MT. Rumor detection and classification for Twitter data. arXiv Preprint posted online on November 25, 2019 [FREE Full text] [doi: 10.1090/mbk/121/79]
105. Iwendi C, Mohan S, Khan S, Ibeke E, Ahmadian A, Ciano T. Covid-19 fake news sentiment analysis. Comput Electr Eng 2022;101:107967 [FREE Full text] [doi: 10.1016/j.compeleceng.2022.107967] [Medline: 35474674]
106. Kasseropoulos D, Tjortjis C. An Approach Utilizing Linguistic Features for Fake News Detection. IFIP Advances in Information and Communication Technology. Cham: Springer International Publishing; 2021.
107. Rezayi S, Soleymani S, Arabnia HR, Li S. Socially aware multimodal deep neural networks for fake news classification. 2021 Presented at: IEEE 4th International Conference on Multimedia Information Processing and Retrieval (MIPR). IEEE; 2021; September 10, 2021; Tokyo, Japan. [doi: 10.1109/mipr51284.2021.00048]
108. Anggrainingsih R, Hassan GM, Datta A. BERT based classification system for detecting rumours on Twitter. arXiv Preprint posted online on September 7, 2021. [doi: 10.48550/ARXIV.2109.02975]
109. Tafannum F, Sharear Shopnil MN, Salsabil A, Ahmed N, Rabiul Alam MG, Tanzim Reza M. Demystifying black-box learning models of rumor detection from social media posts. 2021 Presented at: 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). IEEE; 2021; December 4, 2021; New York, NY. [doi: 10.1109/uemcon53757.2021.9666567]
110. Al-Rakhami MS, Al-Amri AM. Lies kill, facts save: detecting COVID-19 misinformation in Twitter. IEEE Access 2020;8:155961-155970. [doi: 10.1109/access.2020.3019600]
111. Kalyanam J, Velupillai S, Doan S, Conway M, Lanckriet G. Facts and fabrications about ebola: a Twitter based study. arXiv Preprint posted online on August 9, 2015. [doi: 10.48550/ARXIV.1508.02079]
112. Kwon S, Cha M, Jung K, Chen W, Wang Y. Prominent features of rumor propagation in online social media. 2013 Presented at: IEEE 13th International Conference on Data Mining; December 10, 2013; Dallas, TX. [doi: 10.1109/icdm.2013.61]
113. Mazzeo V, Rapisarda A, Giuffrida G. Detection of fake news on COVID-19 on web search engines. Front Phys 2021;9 [FREE Full text] [doi: 10.3389/fphy.2021.685730]
114. Puraivan E, Godoy E, Riquelme F, Salas R. Fake news detection on Twitter using a data mining framework based on explainable machine learning techniques. IET Conf Proc 2021;2021(1):157-162. [doi: 10.1049/icp.2021.1450]

115. Sahana VP, Pias AR, Shastri R, Mandloi S. Automatic detection of rumoured tweets and finding its origin. 2015 Presented at: International Conference on Computing and Network Communications (CoCoNet); December 16-19, 2015; Trivandrum, India. [doi: 10.1109/coconet.2015.7411251]
116. Sailunaz K, Kawash J, Alhajj R. Tweet and user validation with supervised feature ranking and rumor classification. Multimed Tools Appl 2022;81(22):31907-31927. [doi: 10.1007/s11042-022-12616-6]
117. Sharma D, Singhal S. Detection of fake news on social media using classification data mining techniques. Int J Eng Adv Technol 2019;9(1):3132-3138.
118. Afroz S, Brennan M, Greenstadt R. Detecting hoaxes, frauds, and deception in writing style online. 2012 Presented at: IEEE Symposium on Security and Privacy. IEEE; May 23, 2012; San Francisco, CA. [doi: 10.1109/sp.2012.34]
119. Khan S, Hakak S, Deepa N, Prabadevi B, Dev K, Trelova S. Detecting COVID-19-related fake news using feature extraction. Front Public Health 2021;9:788074 [FREE Full text] [doi: 10.3389/fpubh.2021.788074] [Medline: 35059379]
120. Levi O, Hosseini P, Diab M, Broniatowski D. Identifying nuances in fake news vs. satire: using semantic and linguistic cues. 2019 Presented at: Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda; November 01, 2019; Hong Kong, China. [doi: 10.18653/v1/d19-5004]
121. Argyris YA, Zhang N, Bashyal B, Tan PN. Using deep Learning to identify linguistic features that facilitate or inhibit the propagation of anti- and pro-vaccine content on social media. 2022 Presented at: 2022 IEEE International Conference on Digital Health (ICDH). IEEE; July 10-16, 2022; Barcelona, Spain p. 107-116 URL: https://europepmc.org/abstract/MED/37975063 [doi: 10.1109/icdh55609.2022.00025]
122. Maan M, Jain M, Trivedi S, Sharma R. Machine learning based rumor detection on Twitter data. In: Communications in Computer and Information Science. New York, NY: Springer International Publishing; 2022.
123. Raj C, Meel P. Is dynamic rumor detection on social media viable? An unsupervised perspective. arXiv Preprint posted online on November 23, 2021. [doi: 10.48550/arXiv.2111.11982]
124. Office of the Commissioner. 6 Tip-offs to rip-offs: don't fall for health fraud scams. US Food and Drug Administration. 2021. URL: https://www.fda.gov/consumers/consumer-updates/6-tip-offs-rip-offs-dont-fall-health-fraud-scams [accessed 2023-07-20]
125. Apter AJ, Paasche-Orlow MK, Remillard JT, Bennett IM, Ben-Joseph EP, Batista RM, et al. Numeracy and communication with patients: they are counting on us. J Gen Intern Med 2008;23(12):2117-2124 [FREE Full text] [doi: 10.1007/s11606-008-0803-x] [Medline: 18830764]
126. Büchter RB, Fechtelpeter D, Knelangen M, Ehrlich M, Waltering A. Words or numbers? Communicating risk of adverse effects in written consumer health information: a systematic review and meta-analysis. BMC Med Inform Decis Mak 2014;14:76 [FREE Full text] [doi: 10.1186/1472-6947-14-76] [Medline: 25155972]
127. Omura M, Maguire J, Levett-Jones T, Stone TE. The effectiveness of assertiveness communication training programs for healthcare professionals and students: a systematic review. Int J Nurs Stud 2017;76:120-128. [doi: 10.1016/j.ijnurstu.2017.09.001] [Medline: 28964979]
128. Richard C, Lussier M, Millette B, Tanoubi I. Healthcare providers and patients: an essay on the importance of professional assertiveness in healthcare today. Med Educ Online 2023;28(1):2200586 [FREE Full text] [doi: 10.1080/10872981.2023.2200586] [Medline: 37040516]
129. ChatGPT 3.5. OpenAI. URL: https://chatgpt.com/g/g-F00faAwkE-open-a-i-gpt-3-5 [accessed 2025-01-16]

Abbreviations

API: application programming interface
AUC: area under the curve
BERT: Bidirectional Encoder Representations from Transformers
LIWC: Linguistic Inquiry and Word Count
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by T Mackey; submitted 31.05.24; peer-reviewed by D Boatman, A King, W Ahmed; comments to author 05.07.24; revised version received 22.08.24; accepted 23.11.24; published 12.02.25.

Please cite as:
Fridman I, Boyles D, Chheda R, Baldwin-SoRelle C, Smith AB, Elston Lafata J
Identifying Misinformation About Unproven Cancer Treatments on Social Media Using User-Friendly Linguistic Characteristics: Content Analysis
JMIR Infodemiology 2025;5:e62703
URL: https://infodemiology.jmir.org/2025/1/e62703
doi:10.2196/62703
PMID:

Original Paper

A Model of Trust in Online COVID-19 Information and Advice: Cross-Sectional Questionnaire Study

Elizabeth Sillence[1], PhD; Dawn Branley-Bell[1*], PhD; Mark Moss[1*], PhD; Pam Briggs[1*], PhD

Department of Psychology, Northumbria University, Newcastle upon Tyne, United Kingdom

*these authors contributed equally

Corresponding Author:
Elizabeth Sillence, PhD
Department of Psychology
Northumbria University
Ellison Terrace
Newcastle upon Tyne, NE1 8ST
United Kingdom
Phone: 44 1912437251
Email: elizabeth.sillence@northumbria.ac.uk

Abstract

Background: During the COVID-19 pandemic, many people sought information from websites and social media. Understanding the extent to which these sources were trusted is important in relation to health communication.

Objective: This study aims to identify the key factors influencing UK citizens' trust and intention to act on advice about COVID-19 found via digital resources and to test whether an existing model of trust in eHealth provided a good fit for COVID-19–related information seeking online. We also wished to identify any differences between the evaluation of general information and information relating specifically to COVID-19 vaccines.

Methods: In total, 525 people completed an online survey in January 2022 encompassing a general web trust questionnaire, measures of information corroboration, coping perceptions, and intention to act. Data were analyzed using principal component analysis and structural equation modeling. The evaluation responses of general information and COVID-19 vaccine information were also compared.

Results: The principal component analysis revealed 5 trust factors: (1) credibility and impartiality, (2) familiarity, (3) privacy, (4) usability, and (5) personal experiences. In the final structural equation modeling model, trust had a significant direct effect on intention to act (β=.65; *P*<.001). Of the trust factors, credibility and impartiality had a significant positive direct effect on trust (β=.82; *P*<.001). People searching for vaccination information felt less at risk, less anxious, and more optimistic after reading the information. We noted that most people sought information from "official" sources. Finally, in the context of COVID-19, "credibility and impartiality" remain a key predictor of trust in eHealth resources, but in comparison with previous models of trust in online health information, checking and corroborating information did not form a significant part of trust evaluations.

Conclusions: In times of uncertainty, when faced with a global emergent health concern, people place their trust in familiar websites and rely on the perceived credibility and impartiality of those digital sources above other trust factors.

(JMIR Infodemiology 2025;5:e59317) doi:10.2196/59317

KEYWORDS

eHealth; electronic health; digital intervention; trust; online information seeking; scientific credibility; digital resources; COVID-19; SARS-CoV-2; respiratory; infectious; pulmonary; pandemic; public health; health information; global health; surveys; social media

Introduction

Background

The COVID-19 pandemic understandably led to an increase in "official" sources of information and advice from politicians, public health officials, clinicians, and scientists. This public-facing information was communicated via the mainstream press, through live-streamed press briefings, and online. However, "unofficial" sources of information were also circulated, primarily via social media. For individuals, access to good quality information during the pandemic was critical, not least because official messaging was constantly being updated in relation to recommended or mandated behaviors such as social distancing, mask-wearing, and self-isolation.

During this time, many people sought their information online [1] through websites, social media, and mobile apps. People looked for information on the signs and symptoms of the virus, measures to avoid catching and spreading the virus, self-care once infected, and vaccination information. In addition to health advice, people also sought related information on rules and guidance regarding self-isolating, masks, and social distancing.

Accurate and appropriate health communication is an important tool in tackling any pandemic and it can directly influence individuals' affective and behavioral responses to a crisis [2]. In relation to the COVID-19 pandemic, studies have shown that access to a larger and more diverse set of information sources led to increased worry [1,3] and greater confusion, in part because of the infodemic of misinformation and rumors that were promoted about the pandemic [4]. The UK Government's approach to tackling COVID-19 relied upon broad public trust, but issues with inconsistent and unclear messaging, as well as general political mistrust, were apparent [5]. In short, it sometimes became difficult for people to know who to trust in relation to taking appropriate actions to reduce the spread of COVID-19 and minimize personal risk.

Against this backdrop, the aim of this study was to understand more about the digital resources people in the United Kingdom used for COVID-19–related information and the extent to which they trusted these resources. Although we know that online health formed a key source of information for many people during the pandemic, we do not know how people evaluated these digital sources and what factors were important in trusting the information, the source, and ultimately deciding whether or not to act on the advice given. We also wished to test whether an existing model of trust in eHealth provided a good fit for COVID-19–related information seeking online. We begin by briefly reviewing the literature on trust and eHealth before introducing the COVID-19 context and outlining the study objectives.

Trust in Online Health Information

Over the last 20 years, research has consistently pointed to the importance of both the design and the content of websites in terms of establishing trustworthiness [6,7]. Commonly reported indicators of trust and credibility include site owners or sponsors; consensus among multiple sources; characteristics of writing and language; advertisements; content authorship; and interface design [8]. Related studies have looked at the quality of web-based health information and have highlighted navigability, aesthetics, and ease of understanding as important factors [9]. As digital resources for health have developed and diversified, we have seen a move away from government and medically driven sources towards more charity and patient-led sites [10] and the use of social media [11,12] meaning that shared patient experience has also become a critical factor in determining trust and appropriateness of online advice [13].

Despite concerns about the quality and reliability of some digital sources [14], they are often well-used and well-liked. Interestingly, they are not necessarily trusted and the advice they contain is not always acted upon. In part, this may relate to a dislike in the United Kingdom for commercial funding models underpinning health websites [10]. A recent model of trust in eHealth [15] found that credibility and impartiality are the key predictors of trust in eHealth websites and noted that websites containing patients' experiences can have a positive impact on trust but only if those sources have been checked against other sources first. The authors also noted that the need to corroborate digital information sources may be reduced in cases where there is strong familiarity with a well-used website.

COVID-19 Context

The COVID-19 pandemic led to a global surge in information seeking online in relation to the spread of the virus, best means of protection, access to health care, local rules and guidance, and, subsequently, information about COVID-19 vaccines, tracing apps and COVID-19 passports [16]. While official sources moved quickly to try and fill these information gaps, social media platforms provided a space for information and misinformation to circulate widely [17]. Conspiracy theories and rumors in relation to the virus and the vaccine were prevalent online as was poor-quality information [18-20]. The unique situation increased attention on governments as a source of information however historically government and official health sources have been subject to mistrust and their health messages resisted especially concerning vaccinations for example in the case of the Measles Mumps Rubella vaccination and the H1N1 (swine flu) vaccination program [21,22]. In these cases, trust in nonofficial information sources and the media is often higher.

United Kingdom Context

In response to the global pandemic, the UK prime minister announced a national lockdown on March 23rd, 2023 [23]. Daily press briefings followed, led by politicians and National Health Service (NHS) leaders providing coordinated information on COVID-19 legislation and guidance, health advice, and subsequently the vaccine rollout.

Survey data indicates there was a slight increase in political trust in the United Kingdom as the lockdown commenced [24] and most people supported the government enforcement of behavior in the early months [5] with positive views on government decision-making related to response transparency. Although people looked to government and health leaders for information and guidance these officials were not immune from criticism. Politicians and advisors often found themselves at the center of news stories that challenged perceptions of trust [24], and of privacy and security, for example in relation to the rollout of contact tracing apps [25] and COVID-19 passports. Low trust in scientists and medics was also associated with COVID-19 vaccine hesitancy [26].

The sudden onset of COVID-19 and its impact not just on UK citizens but worldwide highlighted the public's need for information. Understanding how individuals sought information from digital sources and whether they trusted this information is the focus of this study. Note that this distinct aim is different from many of the studies of information-seeking behavior during the pandemic that were more focused on the motives that drive online interrogation. Typically, these searches adopted the Risk Information Seeking and Processing model [27] which sees risk information seeking as driven by factors such as information insufficiency, subjective norms, and relevant channel beliefs. Although the Risk Information Seeking and Processing model has been used effectively to model information-seeking behaviors in relation to COVID-19 [28,29] it says relatively little about the extent to which people decide whether to trust the information they are exposed to.

Other studies have examined overall levels of trust in traditional information sources concerning COVID-19 by comparing television, radio, and newspapers with websites [30] but to our knowledge, this is the first study that examines trust and the antecedents of trust in different digital resources in relation to COVID-19. Focusing on the antecedents of trust at this time alongside individuals' behavioral and attitudinal responses to the information they found is key for our future understanding of trusted health communication during health emergencies.

Rationale for This Study

The revised model of trust in eHealth [15] indicates a number of antecedents for trust in online health information and advice and for intention to act on that advice. This study builds upon that work by asking whether existing trust models are a good fit for COVID-19 information-seeking online. The uncertainty provided by the COVID-19 pandemic provides a unique opportunity to examine how people search for, evaluate, and make trust decisions about health information and advice.

The COVID-19 pandemic provides an opportunity to examine in more depth the type of health information seeking that has been taking place. As described previously, people's information needs vary including information on symptoms and symptom management, self-isolation, and vaccination. Vaccination in particular presents a unique opportunity to explore health information seeking within the context of heightened uncertainty and self-reported behavioral outcomes.

It may be that the global nature of the pandemic and people's desire for information exchange fueled social media sources of health information and increased visibility of patient experiences. On the other hand, information corroboration is effortful, and in times of heightened stress and uncertainty, it may not be appropriate or lead to better coping outcomes. Relying on a single source of information may be more straightforward but trust in government or health professionals may impact trust perceptions around such information sources.

Therefore, the study has three aims: (1) to examine whether an existing trust model is a good fit for COVID-19–related information seeking online, (2) to examine differences in affective responses to digital resources about COVID-19 vaccination versus general information about COVID-19, and (3) to examine whether searching had a self-reported impact on vaccination decisions or attitude toward COVID-19 passports.

Methods

Design

A cross-sectional survey was conducted in January 2022. At this time in the United Kingdom, the Omicron variant wave had just peaked, mask use was still advised but no longer compulsory in indoor settings, and self-isolation after a positive test result was still a legal requirement. We collected quantitative data from eHealth users regarding their use of health websites in relation to COVID-19. We used Prolific to recruit a representative UK sample.

Participants

A total of 600 people completed the survey. In total, 525 participants indicated they had looked for COVID-19 information online. Of these 85.3% (448/525) had looked for more general information and advice about COVID-19 while 14.7% (77/525) had looked for information specifically on the vaccine. Full details of participant demographics can be found in Table 1.

Participants were asked whether they had gone online to look for health advice and information about COVID-19. Those answering "yes" were asked to indicate whether they had been searching for general health advice about COVID-19 or whether they had been searching for health advice about COVID-19 vaccinations. Participants then completed a series of questions relating to the last time they searched for health advice about COVID-19 online. Specifically, they were asked to "think about any one digital source that you visited during that search" and to answer the remaining questions with respect to that source. They answered questions relating to the impact of health advice on their coping perceptions and intention to act on the advice, the degree to which they trusted the information and the digital source, their attitude toward COVID passports, for example, the NHS app that shows proof of vaccination and demographic information.

Table 1. Participant demographics (of those who reported looking for COVID-19 information, N=525). All participants were from the United Kingdom.

Characteristics	Values, n (%)
Age group (years)	
18-25	54 (10.3)
26-34	85 (16.2)
35-54	197 (37.5)
55-64	123 (23.4)
65 years or older	66 (12.6)
Sex	
Male	249 (47.4)
Female	273 (52)
Transgender	2 (0.4)
Other	1 (0.2)
Ethnicity	
Caucasian	430 (81.9)
Latino or Hispanic	3 (0.6)
Middle Eastern	5 (1)
African	11 (2.1)
Caribbean	10 (1.9)
South Asian	31 (5.9)
East Asian	11 (2.1)
Mixed	12 (2.3)
Other	7 (1.3)
Prefer not to say	5 (1)
Education level	
Less than secondary school	2 (0.4)
Secondary school	68 (13)
Further education (eg, college, A-level)	177 (33.7)
Bachelor’s degree	194 (37)
Postgraduate degree (eg, MSc, PhD, MD)	82 (15.6)
Prefer not to say	2 (0.4)
Employment	
Full time	254 (48.4)
Part time	87 (16.4)
Retired	85 (16.2)
Unemployed	60 (11.4)
Student	29 (5.5)
Prefer not to say	10 (1.9)
Relationship status	
Single	143 (27.2)
Married or civil partnership or cohabiting	333 (63.4)
Divorced	30 (5.6)
Widowed	10 (1.9)
Prefer not to say	9 (1.7)

Measures

Unless stated otherwise, participants answered the following measures on a 5-point Likert scale (1=strongly disagree to 5=strongly agree).

General Web Trust Questionnaire

The general web trust questionnaire contained 36 items from the study by Sillence et al [15] alongside measures of coping, information corroboration, and affective responses also taken from Sillence et al [15]. Specifically, coping was measured by asking participants to respond to the following stem and variables "After I read the information about COVID-19 I felt…" (1) in control and (2) optimistic using a 5-point scale with the labels: 1=less, 2=slightly less, 3=no different, 4=slightly more, and 5=more (Cronbach α=.84.). Additional affective responses, worried, reassured, at risk, confused and anxious were measured using the same format.

Information corroboration with other sources of information was measured with the following 4 items: (1) "I checked other websites," (2) "I checked other sources," (3) I found the advice consistent across other websites or apps, and (4) I found the advice consistent across other sources (Cronbach α=.87).

Impact on vaccination decision was measured using a single item developed for this study: "To what extent did the information and advice you read online impact your decision regarding COVID vaccinations?" Responses were given on a 5-point scale from "1=It did not influence at all" to "5=It influenced to a very large degree."

Attitude toward COVID-19 passports was measured using a single item developed for this study, that is, "I think COVID passports are a good idea" (1=strongly disagree to 5=strongly agree).

Outcome Measures

Trust was measured following Sillence et al [15], using the mean response to the following 2 items: (1) "I trusted the site" and (2) "I felt I could trust the information on the site" (Cronbach α=.95). Intention to act was an outcome measure, assessed with 1 item "I intended to act upon the advice." This item was taken from Sillence et al [15].

Ethical Considerations

The study received full ethical approval from Northumbria University ethics committee (REF:33639). The survey was hosted on Qualtrics and all data was anonymized. The first page provided participants with information detailing the aim, length, data storage, contact details, and withdrawal process of the study. They were then asked to provide informed consent. Participants received £1.25 (€1.49; US $1.66) for taking part in the study and the average completion time was around 7 minutes.

Results

Overview

We first explored the general web trust questionnaire by performing principal component analysis (PCA). We then explored the relationship between the factor structure and outcomes by testing its fit to the sampled data using structural equation modeling (SEM).

Properties of the General Web Trust Questionnaire

The 36 items of the scale were entered into the PCA. All items loaded onto the extracted components but any items with factor loadings lower than 0.30 were suppressed (Table 2). The analysis indicated that 5 components possessed eigenvalues greater than 1 and together explained 68.7% of the variance in keeping with accepted conventions for successful PCA [31]. The Familiarity factor is the weakest of those extracted although it does meet the minimum threshold of comprising three items [32].

Table 2. Factor loadings for each item (factor loadings lower than .30 are suppressed).

Item	Rotation factor loadings				
	Personal experience (PEx)	Credibility and impartiality	Usability	Privacy	Familiarity
The language made it easy to understand	—[a]	—	.69	—	—
It helped me understand the issue better	—	—	.70	—	—
It was easy to use	—	—	.77	—	—
It told me most of what I needed to know	—	—	.59	—	—
The layout was consistent with other digital sources	—	—	.61	—	—
The advice appeared to be prepared by an expert	—	.69	—	—	—
The advice seemed to be offered in my best interests	—	.73	—	—	—
The advice came from a knowledgeable source	—	.73	—	—	—
The advice seemed credible	—	.80	—	—	—
It was owned by a well-known organization	—	—	—	—	.73
It featured familiar logos	—	—	—	—	.78
It had a professional design	—	—	—	—	.64
It had an attractive design	—	—	.47	—	—
It provided reassurances about my privacy	—	—	—	.66	—
It gave the option to post anonymously	—	—	—	.45	—
It gave reassurances about how they used your information	—	—	—	.78	—
It had a privacy policy	—	—	—	.82	—
It explained their use of cookies	—	—	—	.75	—
It contained accounts of other people's experiences	.87	—	—	—	—
There was a chance to share my experiences	.90	—	—	—	—
There were opportunities to interact with other people on the digital source	.87	—	—	—	—
I saw a wide range of experiences rather different to mine	.88	—	—	—	—
It offered powerful accounts of health experiences	.85	—	—	—	—
It felt like the advice was tailored to me personally	.62	—	—	—	—
I was offered the chance to see experiences from people just like me	.91	—	—	—	—
It contained contributions from likeminded people	.92	—	—	—	—
I was able to contribute to content on the digital source	.88	—	—	—	—
The personal accounts on the digital source were written by people similar to me	.91	—	—	—	—
I found personal accounts that reflected my own experience	.92	—	—	—	—
I found personal accounts that were relevant to my condition	.93	—	—	—	—
There were opportunities to gather information from the personal accounts on the digital source	.91	—	—	—	—
The personal accounts contained advice for readers	.91	—	—	—	—
The personal accounts provided social or emotional support	.89	—	—	—	—
The advice appeared to be impartial and independent	—	.78	—	—	—

Item	Rotation factor loadings				
	Personal experience (PEx)	Credibility and impartiality	Usability	Privacy	Familiarity
The advice seemed objective (ie, no hidden agenda)	—	.81	—	—	—
It was free from advertisements		.54	—	—	—
Eigenvalues	11.8	4.7	3.2	3.0	2.1
Variance explained (%)	32.7	13.1	8.9	8.2	5.8

[a]Not available.

Exploring the Relationship Between the Trust Questionnaire and Self-Reported Behavioral Outcomes

The data were further analyzed using SEM performed in analysis of moment structures using the maximum likelihood estimation method on the item covariance matrix. The specified model was based on Sillence et al [15] and modified to incorporate the new 5-factor structure. The goodness of fit indices supports the specified model. The chi-square value indicated poor fit (χ^2_{773}=1265.5; P<.001). However, this test is considered too sensitive for samples over 200 and here the sample size is 448. The Cmin/*df* value of 1.64 indicates a good fit. The goodness of fit and adjusted goodness of fit values of .89 and .86 respectively indicate adequate fit [33]. The comparative fit index value of .97 indicates good fit [34], as does the root mean square of approximation value of .04, 90% CI .034-.041 [35].

The model accounted for 64.7% of the variance in trust, 8.7% in coping, 9.7% in information corroboration, and 40.3% in intention to act. All beta path coefficients including those in Figure 1 and those that were not significant were inspected in evaluating the predictive power of the model and are presented for completeness in Table 3.

Figure 1. The trust model with significant standardized path coefficients.

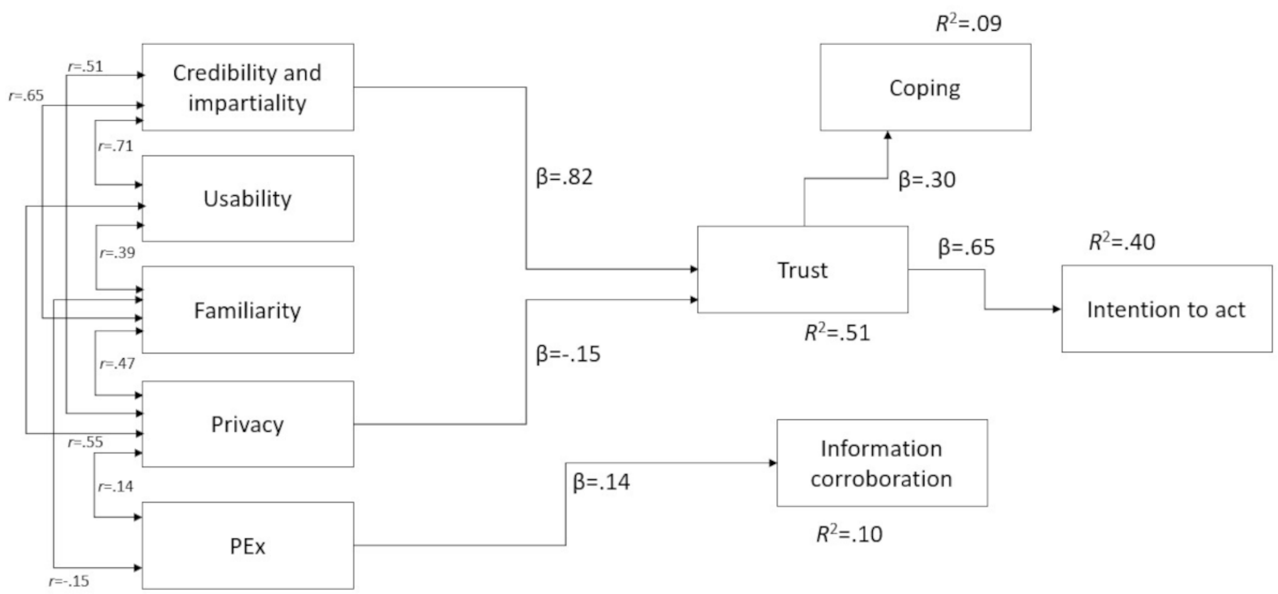


Table 3. The unstandardized path weights and critical ratio (ie, *z* score) values for the main effects of the hypothesized full model.

Parameter	Unstandardized path coefficient	Critical ratio	*P* value
Credibility and impartiality			
Trust	.93	9.79	<.001
Information corroboration	.17	1.07	.29
Usability			
Trust	–.05	–.36	.72
Information corroboration	.39	1.56	.12
Familiarity			
Trust	–.04	–.64	.52
Information corroboration	.12	.98	.33
Privacy			
Trust	–.19	–2.43	.02
Information corroboration	.06	.41	.68
Personal experience			
Trust	–.001	–.03	.98
Information corroboration	.09	2.78	.01
Trust			
Coping	.27	4.89	<.001
Intention to act	.80	15.23	<.001
Coping–intention to act	–.04	–.67	.50
Information corroboration			
Trust	.001	.03	.98
Intention to act	–.02	–.61	.54

Only Credibility and Impartiality were found to possess a significant positive path to Trust. Privacy had a weaker yet significant negative path, meaning privacy assurances were associated with lower trust. Familiarity, usability, and personal experience (PEx) were not significantly predictive of Trust. Only Trust was found to significantly predict the intention to act on the advice. In addition, Trust significantly predicted Coping, suggesting that trustworthy websites heighten individuals' coping perceptions, making them feel more in control and optimistic. PEx significantly predicts Information Corroboration, suggesting that people are exploring a little further than the original digital source; however, this corroboration process does not appear to be affecting their level of trust or intention to act.

Comparison of Two Populations

Although the relatively small sample size for the vaccine information group meant that a comparable SEM model could not be reliably tested a series of independent samples *t* tests were used to compare the two groups on the key variables of interest (Tables 4 and 5).

Table 4. Mean (SD) values for key outcome variables.

Group	Trust	Intention to act	Corroboration	Impact on the decision regarding vaccination	Attitude toward COVID-19 passports
Searching for information on vaccinations (N=77)	4.22 (.91)	4.10 (1.05)	3.49 (1.24)	2.90 (1.21)	3.38 (1.51)
Searching for information on COVID-19 (N=448)	4.33 (.74)	4.13 (.89)	3.49 (1.06)	2.74 (1.39)	3.51 (1.36)

Table 5. Mean (SD) values for "after I read the information" variables.

Group	Worried	Reassured	At risk	Confused	Anxious	Optimistic	In control
Searching for information on vaccinations (N=77)	2.27 (1.11)	3.84 (.95)	2.40 (.98)	2.14 (1.13)	2.42 (1.20)	3.66 (1.11)	3.57 (1.13)
Searching for information on COVID-19 (N=448)	2.48 (.88)	3.68 (.77)	2.84 (.88)	2.15 (.98)	2.76 (.97)	3.27 (.81)	3.42 (.85)

Independent Sample *t* tests

There was no significant difference between groups for trust (t_{523}=–1.169; *P*=.24; Cohen *d*=–.14, 95% CI –.386 to .098), intention to act (t_{523}=–.187; *P*=.85; Cohen *d*=–.02, 95% CI –.265 to .219), corroboration (t_{523}=–.038; *P*=.97; Cohen *d*=–.01, 95% CI –.247 to .237), impact on decision regarding vaccination (t_{523}=.934; *P*=.35; Cohen *d*=.115, 95% CI –.127 to .357), or COVID-19 passports (t_{523}=–.773; *P*=.44; Cohen *d*=–.095, 95% CI –.337 to .146).

Those searching for information on vaccinations (mean 2.40) felt significantly less at risk than those searching for general information on COVID-19 (mean 2.84; t_{523}=3.988; *P*<.001; Cohen *d*=–49, 95% CI –.735 to –.2348) and felt significantly less anxious (mean 2.42) than those searching for general information on COVID-19 (mean 2.76; t_{523}=–2.758; *P*=.003; Cohen *d*=–.34, 95% CI –.583 to –.097). Those searching for information on vaccinations (mean=3.66) felt significantly more optimistic than those searching for general information on COVID-19 (mean=3.27; t_{523}=3.760; *P*<.001; Cohen *d*=.464, 95% CI .220-.707).

There was no significant difference for the variable "In Control" (t_{523}=1.335; *P*=.18; Cohen *d*=–.165, 95% CI –.077 to .407) or for "Confused" (t_{523}=–.054; *P*=.96; Cohen *d*=–.007, 95% CI –.248 to .235). Finally, the variables "Worried" and "Reassured" approached but did not reach statistical significance (t_{523}=–1.813; *P*=.07; Cohen *d*=–.224, 95% CI –.466 to .019 and t_{523}=1.712; *P*=.09; Cohen *d*=.211, 95% CI –.031 to .453, respectively).

Digital Sources of Information

Table 6 shows the digital sources used. The majority of participants used either the NHS health care sources or the governmental sources for both general information and vaccine-specific information.

Digital sources were categorized as: (1) Governmental sources: official UK government website (Gov.uk), World Health Organization, Office of National Statistics, and Centre for Disease Control. (2) NHS health care sources: any page hosted on the NHS website (nhs.uk). (3) Other health care sources: any non-NHS health care website. This included The Mayo Clinic, WedMD, patient.co.uk, and the Health Check podcast. (4) News websites: any of the mainstream news providers, the majority of those reported were the BBC. (5) Search engines: where participants did not go to one source but reported explicitly using search engines, such as Google, to intentionally search for COVID-19–related information, rather than, for example, visiting a particular source (perhaps a source perceived as authoritative or trusted), such as the NHS, government, or BBC websites, and browsing the content from there. (6) Scientific journal: any peer-reviewed journal publishing academic research. (7) Specific health condition websites: any website dedicated to a specified health condition rather than a general health website, including asthma.org and Crohn's & Colitis UK. (8) Social media and forums: any online forum or social networking platform defined as user-driven and facilitating sharing of content, dialogue creation, and communication by and between users (in keeping with Kapoor et al, 2018 [36]). (9) Other: all instances where resources were not explicitly specified or where participants reported visiting multiple sources. All other resources are named individually in Table 6.

Table 6. Digital sources used.

Source	General information (N=448), n (%)	Vaccine specific information (N=77), n (%)
National Health Service health care sources	262 (58.48)	39 (50.65)
Governmental sources	64 (14.30)	11 (14.29)
Multiple resources or unspecific	37 (8.30)	13 (16.88)
News websites	30 (6.70)	3 (3.90)
Other health care sources	6 (1.34)	1 (1.30)
Social media and forums	20 (4.46)	2 (2.60)
Search engines	19 (4.24)	7 (9.09)
Zoe COVID-19 study	6 (1.34)	0 (0)
Scientific journals	1 (0.22)	0 (0)
Specific health condition websites	2 (0.45)	0 (0)
Wikipedia	1 (0.22)	0 (0)
TripAdvisor	0 (0)	1 (1.30)

Discussion

Principal Results

Trust continues to significantly influence self-reported intention to act on health information. In terms of trust predictors, only credibility and impartiality have a significant, direct, and positive relationship with trust. Privacy has a significant negative relationship with trust. Other predictors (familiarity, usability, and PEx) may be indirect and mediated through other trust variables. The variable PEx had a significant direct effect on information corroboration and trust was found to significantly relate to coping perceptions. The findings suggest a number of important discussion points.

First, the Sillence et al [15] trust model provides a reasonable fit for COVID-19–related health information online. Trust continues to predict intention and the credibility and impartiality of the digital source remains the strongest predictor of trust in digital health sources. However, compared to the 2019 model, the picture here is of a simpler trust process in which the credibility and impartiality factor does the "heavy lifting" in relation to trust compared to the other variables. Another key difference is the lack of a relationship between corroboration and trust. In earlier models, health information seekers looked to verify the information they found online by cross-checking with other digital and nondigital sources. Here we see only a direct relationship between the credibility and impartiality of the website and trust. One reason for this, given the predominance of the NHS as the most popular site for information and advice, is that our health information seekers are simply taking the website at face value providing it appears sufficiently credible and impartial. However, it is interesting that in an American sample, information seekers relied heavily upon often unreliable social media sources for information and advice, yet still engaged in relatively low levels of fact-checking [37] and so we must consider the possibility that people are being bombarded with so much information in relation to the pandemic that they simply switch off.

The role of PEx within digital sources is interesting here. While PEx significantly predicts information corroboration there was no subsequent relationship with trust. In the 2019 model [15] it was suggested that patient experiences can positively influence trust but only if users first corroborate the information through other sources. In our study, we suggest that people are checking up on these patient stories and experiences simply out of interest rather than as a way of assessing the trustworthiness of the information. When faced with a high degree of uncertainty and with limited detailed information, assessments of risk may be emotion-based [38], and people may well seek out other people's personal accounts of their COVID-19 experiences. Personal accounts are often engaging and are seen as more relatable than statistical information when it comes to decision-making [39]. While PEx is now embedded within a diverse range of digital resources, those more closely associated with personal content, for example, social media platforms or individual blogs, were generally underrepresented in the data we collected. Instead, we observed a reliance on official digital sources, in particular, the NHS website and government sources. In terms of pandemic or emergency, reliance on official sources may be more commonplace. Sillence et al [15] found that the majority of UK respondents cited the NHS website as their source of health information, and McNeill, Harris, and Briggs [40] noted that the main UK source to be retweeted during the H1N1 pandemic was NHS Choices. In this study, there was little reported use of social media, which is perhaps surprising and contrasts with other recent health pandemics in which social media use and misinformation have been prevalent [37,41,42] as well as in earlier studies examining the COVID-19 pandemic and the facilitation of conspiracy theories [43,44].

Despite generally high levels of mistrust in the government's overall handling of the pandemic [5], UK citizens still sought information from government sites. Moreover, we see a reliance on health professionals and public health information. In a time of limited information, there may be fewer options available to information seekers and individuals may be satisfied with seeking official sources of information even if they contain basic knowledge as opposed to more detailed, specific information. This contrasts with earlier work on trust in digital health information in which personalization or tailoring is seen as important to trust. People with long-term experience of a particular health condition often become experts by experience and may seek more specific, tailored digital resources to support their health conditions. This involves making more fine-grained assessments of the personal relevance of the information before deciding to trust or act upon the advice it contains [10,45] and is especially true where the condition is rare or less well known [46]. In the case of COVID-19, a worldwide pandemic affecting all age groups, it might be that generic information applicable to all sufficed in this case. There was little sense that people were checking COVID-19 information in relation to their other, pre-existing health conditions and specific health websites may not have had that information readily available. In light of research that shows how health information overload may lead to increased anxiety [47], our participants' reliance on relatively few, authoritative websites seems like a reasonable strategy. Too much, possibly conflicting, information about COVID-19 can leave an individual feeling overwhelmed and will ultimately lead to "information avoidance," which is clearly a poor outcome in the face of a global pandemic.

Unlike Sillence et al's [15] 2019 model, we note that privacy has a weak negative relationship with trust. The topic of privacy was raised repeatedly in relation to the discussion of contact tracing apps and COVID-19 passports and so while not directly related to the digital source being used it may be that being asked to think about the privacy features of sources stimulates a wider consideration of privacy and mistrust. Rather than privacy policies etc. being seen as an example of good practice, the very fact that these options were present on digital sources may have served as a reminder that data are being collected, processed, and often shared. Privacy nudges may well remind people of the need to be mindful of privacy but can also raise awareness of the data that is available for collection [48,49].

Second, trust significantly predicted coping suggesting that trustworthy websites heighten individuals' coping perceptions, making them feel able to cope. Interestingly, Wang et al [1] did not find an association between the use of the internet as an information source on COVID-19 and self-confidence in coping with COVID-19 but did not focus on trusted websites.

Looking at the affective variables in more detail for the two groups (general information seeking and vaccination information), we see that those searching for vaccination information felt more positive—specifically, they felt less at risk, less anxious, and more optimistic after reading the information. Wang et al [1] found that vaccination information sources have different effects on students' coping appraisal of COVID-19 with information from medical personnel leading to greater knowledge about the mechanism of vaccination and greater response efficacy of vaccination compared to information from coworkers or colleagues. In terms of coping, during the H1N1 pandemic, those people who adopted a more problem-focused coping strategy including seeking out information to help solve problems were more likely to indicate they would be vaccinated [22]. In our data, those individuals who have gone looking for information about vaccination feel better for having done so.

Zheng et al [50] noted that vaccine information seeking is related to vaccination intention and suggested that health information seeking can be viewed as a coping behavior when people do not have sufficient knowledge of a particular health topic. Although seeking vaccine-related information online was also positively related to perceived vaccine information overload [50], it may be that sticking with a single trusted source is preferable for improved coping. Finally, there were no differences in terms of trust, intention to act on information, or attitude toward COVID-19 passports between participants who were searching for general COVID-19 health information versus those who had searched for vaccination information. This is unsurprising given the similarity of digital sources used.

In summary, people searching for general COVID-19 information as well as those searching for COVID-19 vaccine-specific information sought out official sources of information online. In terms of uncertainty when faced with a global emergent health concern people place their trust in familiar websites and rely on the perceived credibility and impartiality of those digital sources.

Limitations

It is important to note that data was purposely not collected during a period of national lockdown in the United Kingdom. The vaccination program was already well underway and COVID-19 passports were very much still on the agenda. People may have sought information from alternative digital sources had data collection taken place during a period of lockdown. Focusing on the United Kingdom made sense given the local regulations and practices in place, but it would be interesting to make comparisons with other countries going forward. The reliance on the NHS website in the United Kingdom would be interesting to compare with countries where different funding models exist for example where health insurance schemes mean there is no single free at the point of service system. Vaccine hesitancy is relatively low in the U and has declined since the start of the vaccination rollout program from 10% to 3% in September 2021 [51]. Other countries, for example, France, have much higher levels of vaccine hesitancy [52], and comparisons here in relation to trust around digital health resources would warrant further investigation. Finally, it is interesting to note that although we have used a one-shot cross-sectional methodology, we mirror findings from Zhang et al [53], who examined trust over several waves earlier in the pandemic and noted a decrease in the use of social media over time and an increase in trust in government information.

Conclusion

In conclusion, in the context of COVID-19, "credibility and impartiality" remain a key predictor of trust in eHealth resources but in comparison with previous models of trust in online health information, checking and corroborating information did not form a significant part of trust evaluations. In times of uncertainty when faced with a global emergent health concern, people placed their trust in familiar websites and relied on the perceived credibility and impartiality of those digital sources.

Acknowledgments

This project contributes towards the EPSRC-funded Centre for Digital Citizens (grant number EP/T022582/1).

Data Availability

The datasets generated and analyzed during this study are available in the Open Science Framework repository [54].

Conflicts of Interest

None declared.

References

1. Wang PW, Lu WH, Ko NY, Chen YL, Li DJ, Chang YP, et al. COVID-19-related information sources and the relationship with confidence in people coping with COVID-19: facebook survey study in Taiwan. J Med Internet Res 2020;22(6):e20021 [FREE Full text] [doi: 10.2196/20021] [Medline: 32490839]
2. Finset A, Bosworth H, Butow P, Gulbrandsen P, Hulsman RL, Pieterse AH, et al. Effective health communication - a key factor in fighting the COVID-19 pandemic. Patient Educ Couns 2020;103(5):873-876 [FREE Full text] [doi: 10.1016/j.pec.2020.03.027] [Medline: 32336348]

3. Ho HY, Chen YL, Yen CF. Different impacts of COVID-19-related information sources on public worry: an online survey through social media. Internet Interv 2020;22:100350 [FREE Full text] [doi: 10.1016/j.invent.2020.100350] [Medline: 33072524]
4. Mheidly N, Fares J. Leveraging media and health communication strategies to overcome the COVID-19 infodemic. J Public Health Policy 2020;41(4):410-420 [FREE Full text] [doi: 10.1057/s41271-020-00247-w] [Medline: 32826935]
5. Enria L, Waterlow N, Rogers NT, Brindle H, Lal S, Eggo RM, et al. Trust and transparency in times of crisis: results from an online survey during the first wave (April 2020) of the COVID-19 epidemic in the UK. PLoS One 2021;16(2):e0239247 [FREE Full text] [doi: 10.1371/journal.pone.0239247] [Medline: 33591985]
6. Kim Y. Trust in health information websites: a systematic literature review on the antecedents of trust. Health Informatics J 2016;22(2):355-369 [FREE Full text] [doi: 10.1177/1460458214559432] [Medline: 25518944]
7. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. J Med Internet Res 2017;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
8. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators. J Med Internet Res 2019;21(5):e12522 [FREE Full text] [doi: 10.2196/12522] [Medline: 31045507]
9. Zhang Y, Kim Y. Consumers' evaluation of web-based health information quality: meta-analysis. J Med Internet Res 2022;24(4):e36463. [doi: 10.2196/36463] [Medline: 35482390]
10. Sillence E, Hardy C, Briggs P. Why don't we trust health websites that help us help each other? An analysis of online peer-to-peer healthcare. 2013 Presented at: WebSci '13: Proceedings of the 5th Annual ACM Web Science Conference; 2013, May 2-4; Paris, France p. 396-404. [doi: 10.1145/2464464.2464488]
11. O'Leary K, Coulson N, Perez-Vallejos E, McAuley D. Towards understanding how individuals with inflammatory bowel disease use contemporary social media platforms for health-related discourse. Computers in Human Behavior 2020;112:106463. [doi: 10.1016/j.chb.2020.106463]
12. Sasse M, Ohrndorf S, Palmowski A, Wagner AD, Burmester GR, Pankow A, et al. Digital health information on autoinflammatory diseases: a YouTube quality analysis. Rheumatol Int 2023;43(1):163-171 [FREE Full text] [doi: 10.1007/s00296-022-05243-9] [Medline: 36374326]
13. Sillence E. Having faith in the online voice. Exploring contemporary issues of trust, language and advice in the context of e-health. LO 2017;87(8). [doi: 10.13092/lo.87.4175]
14. Osman W, Mohamed F, Elhassan M, Shoufan A. Is YouTube a reliable source of health-related information? A systematic review. BMC Med Educ 2022;22(1):382 [FREE Full text] [doi: 10.1186/s12909-022-03446-z] [Medline: 35590410]
15. Sillence E, Blythe JM, Briggs P, Moss M. A revised model of trust in internet-based health information and advice: cross-sectional questionnaire study. J Med Internet Res 2019;21(11):e11125 [FREE Full text] [doi: 10.2196/11125] [Medline: 31710297]
16. Zhao X, Fan J, Basnyat I, Hu B. Online health information seeking using '#COVID-19 Patient Seeking Help' on Weibo in Wuhan, China: descriptive study. J Med Internet Res 2020;22(10):e22910 [FREE Full text] [doi: 10.2196/22910] [Medline: 33001838]
17. Van den Broucke S. Why health promotion matters to the COVID-19 pandemic, and vice versa. Health Promot Int 2020;35(2):181-186 [FREE Full text] [doi: 10.1093/heapro/daaa042] [Medline: 32297931]
18. Zeng J, Schäfer MS. Conceptualizing 'Dark Platforms'. Covid-19-related conspiracy theories on 8kun and Gab. Digital Journalism 2021;9(1):1-23. [doi: 10.1080/21670811.2021.1938165]
19. Islam MS, Kamal AM, Kabir A, Southern DL, Khan SH, Hasan SMM, et al. COVID-19 vaccine rumors and conspiracy theories: the need for cognitive inoculation against misinformation to improve vaccine adherence. PLoS One 2021;16(5):e0251605 [FREE Full text] [doi: 10.1371/journal.pone.0251605] [Medline: 33979412]
20. Chan C, Sounderajah V, Daniels E, Acharya A, Clarke J, Yalamanchili S, et al. The reliability and quality of YouTube videos as a source of public health information regarding COVID-19 vaccination: cross-sectional study. JMIR Public Health Surveill 2021;7(7):e29942 [FREE Full text] [doi: 10.2196/29942] [Medline: 34081599]
21. Brown KF, Kroll JS, Hudson MJ, Ramsay M, Green J, Long SJ, et al. Factors underlying parental decisions about combination childhood vaccinations including MMR: a systematic review. Vaccine 2010;28(26):4235-4248. [doi: 10.1016/j.vaccine.2010.04.052] [Medline: 20438879]
22. Taha SA, Matheson K, Anisman H. The 2009 H1N1 influenza pandemic: the role of threat, coping, and media trust on vaccination intentions in Canada. J Health Commun 2013;18(3):278-290. [doi: 10.1080/10810730.2012.727960] [Medline: 23301849]
23. Prime Minister's Statement on Coronavirus (COVID-19): 23 March 2020. GOV.UK. URL: https://www.gov.uk/government/speeches/pm-address-to-the-nation-on-coronavirus-23-march-2020 [accessed 2025-10-15]
24. Davies B, Lalot F, Peitz L, Heering M, Cifterler HO, Babaian J, et al. Changes in political trust in Britain during the COVID-19 pandemic in 2020: integrated public opinion evidence and implications. Humanit Soc Sci Commun 2021;8(166):1-9. [doi: 10.1057/s41599-021-00850-6]

25. Samuel G, Roberts SL, Fiske A, Lucivero F, McLennan S, Phillips A, et al. COVID-19 contact tracing apps: UK public perceptions. Crit Public Health 2022;32(1):31-43 [FREE Full text] [doi: 10.1080/09581596.2021.1909707] [Medline: 35221546]
26. Allington D, McAndrew S, Moxham-Hall V, Duffy B. Coronavirus conspiracy suspicions, general vaccine attitudes, trust and coronavirus information source as predictors of vaccine hesitancy among UK residents during the COVID-19 pandemic. Psychol Med 2023;53(1):236-247 [FREE Full text] [doi: 10.1017/S0033291721001434] [Medline: 33843509]
27. Griffin RJ, Dunwoody S, Neuwirth K. Proposed model of the relationship of risk information seeking and processing to the development of preventive behaviors. Environ Res 1999;80(2 Pt 2):S230-S245. [doi: 10.1006/enrs.1998.3940] [Medline: 10092438]
28. Zhou X, Roberto AJ. An application of the risk information seeking and processing model in understanding college students' COVID-19 vaccination information seeking and behavior. Sci Commun 2022;44(4):446-474 [FREE Full text] [doi: 10.1177/10755470221120415] [Medline: 38603181]
29. Ahn J, Kim HK, Kahlor LA, Atkinson L, Noh GY. The impact of emotion and government trust on individuals' risk information seeking and avoidance during the COVID-19 pandemic: a cross-country comparison. J Health Commun 2021;26(10):728-741. [doi: 10.1080/10810730.2021.1999348] [Medline: 34779340]
30. Ali SH, Foreman J, Tozan Y, Capasso A, Jones AM, DiClemente RJ. Trends and predictors of COVID-19 information sources and their relationship with knowledge and beliefs related to the pandemic: nationwide cross-sectional study. JMIR Public Health Surveill 2020;6(4):e21071 [FREE Full text] [doi: 10.2196/21071] [Medline: 32936775]
31. Hinkin TR. A brief tutorial on the development of measures for use in survey questionnaires. Organ Res Methods 1998;1(1):104-121. [doi: 10.1177/109442819800100106]
32. Yong AG, Pearce S. A beginner’s guide to factor analysis: focusing on exploratory factor analysis. TQMP 2013;9(2):79-94. [doi: 10.20982/tqmp.09.2.p079]
33. Marsh H, Grayson D. Latent variable models of multitrait-multimethod data. In: Structural Equation Modeling: Concepts, Issues, and Applications. Thousand Oaks: Sage; 1995:177-198.
34. Garson GD. Statnotes: Topics in Multivariate Analysis. 2009. URL: https://faculty.chass.ncsu.edu/garson/PA765/statnote.htm [accessed 2025-01-08]
35. Browne MW, Cudeck R. Alternative ways of assessing model fit. Sociological Methods & Research 1992;21(2):230-258. [doi: 10.1177/0049124192021002005]
36. Kapoor KK, Tamilmani K, Rana NP, Patil P, Dwivedi YK, Nerur S. Advances in social media research: past, present and future. Inf Syst Front 2017;20(3):531-558. [doi: 10.1007/s10796-017-9810-y]
37. Neely S, Eldredge C, Sanders R. Health information seeking behaviors on social media during the COVID-19 pandemic among American social networking site users: survey study. J Med Internet Res 2021;23(6):e29802 [FREE Full text] [doi: 10.2196/29802] [Medline: 34043526]
38. Leppin A, Aro AR. Risk perceptions related to SARS and avian influenza: theoretical foundations of current empirical research. Int J Behav Med 2009;16(1):7-29 [FREE Full text] [doi: 10.1007/s12529-008-9002-8] [Medline: 19214752]
39. Winterbottom A, Bekker HL, Conner M, Mooney A. Does narrative information bias individual's decision making? A systematic review. Soc Sci Med 2008;67(12):2079-2088. [doi: 10.1016/j.socscimed.2008.09.037] [Medline: 18951673]
40. McNeill A, Harris PR, Briggs P. Twitter influence on UK vaccination and antiviral uptake during the 2009 H1N1 pandemic. Front Public Health 2016;4:26 [FREE Full text] [doi: 10.3389/fpubh.2016.00026] [Medline: 26942174]
41. Dredze M, Broniatowski DA, Hilyard KM. Zika vaccine misconceptions: a social media analysis. Vaccine 2016;34(30):3441-3442 [FREE Full text] [doi: 10.1016/j.vaccine.2016.05.008] [Medline: 27216759]
42. Sharma M, Yadav K, Yadav N, Ferdinand KC. Zika virus pandemic-analysis of Facebook as a social media health information platform. Am J Infect Control 2017;45(3):301-302. [doi: 10.1016/j.ajic.2016.08.022] [Medline: 27776823]
43. Jennings W, Stoker G, Bunting H, Valgarðsson VO, Gaskell J, Devine D, et al. Lack of trust, conspiracy beliefs, and social media use predict COVID-19 vaccine hesitancy. Vaccines (Basel) 2021;9(6):593 [FREE Full text] [doi: 10.3390/vaccines9060593] [Medline: 34204971]
44. Theocharis Y, Cardenal A, Jin S, Aalberg T, Hopmann DN, Strömbäck J, et al. Does the platform matter? Social media and COVID-19 conspiracy theory beliefs in 17 countries. New Media & Society 2021;25(12):3412-3437. [doi: 10.1177/14614448211045666]
45. Entwistle VA, France EF, Wyke S, Jepson R, Hunt K, Ziebland S, et al. How information about other people's personal experiences can help with healthcare decision-making: a qualitative study. Patient Educ Couns 2011;85(3):e291-e298. [doi: 10.1016/j.pec.2011.05.014] [Medline: 21652162]
46. Coulson NS, Buchanan H, Aubeeluck A. Social support in cyberspace: a content analysis of communication within a Huntington's disease online support group. Patient Educ Couns 2007;68(2):173-178 [FREE Full text] [doi: 10.1016/j.pec.2007.06.002] [Medline: 17629440]
47. Sultana T, Dhillon G, Oliveira T. The effect of fear and situational motivation on online information avoidance: the case of COVID-19. Int J Inf Manage 2023;69:102596 [FREE Full text] [doi: 10.1016/j.ijinfomgt.2022.102596] [Medline: 36415624]

48. Zhang B, Xu H. Privacy nudges for mobile applications: Effects on the creepiness emotion and privacy attitudes. : ACM; 2016 Presented at: CSCW '16: Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing; 2016 27 February - 2 March; San Francisco California USA p. 1676-1690. [doi: 10.1145/2818048.2820073]
49. Wang Y, Leon P, Acquisti A, Cranor LF, Forget A, Sadeh N. A field trial of privacy nudges for Facebook. 2014 Presented at: CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems; 2014 April 26; Toronto Ontario Canada p. 2367-2376. [doi: 10.1145/2556288.2557413]
50. Zheng H, Jiang S, Rosenthal S. Linking online vaccine information seeking to vaccination intention in the context of the COVID-19 pandemic. Science Communication 2022;44(3):320-346. [doi: 10.1177/10755470221101067]
51. Coronavirus and changing attitudes towards vaccination, England: 7 to 16 September 2021. Office for National Statistics. URL: https://tinyurl.com/bp8aczwh [accessed 2025-01-09]
52. Robinson E, Jones A, Lesser I, Daly M. International estimates of intended uptake and refusal of COVID-19 vaccines: a rapid systematic review and meta-analysis of large nationally representative samples. Vaccine 2021;39(15):2024-2034 [FREE Full text] [doi: 10.1016/j.vaccine.2021.02.005] [Medline: 33722411]
53. Zhang Y, Suhaimi N, Yongsatianchot N, Gaggiano JD, Kim M, Patel SA, et al. Shifting trust: examining how trust and distrust emerge, transform, and collapse in COVID-19 information seeking. : ACM; 2022 Presented at: CHI '22: Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems; 2022 April 29; New Orleans LA USA p. 1-21. [doi: 10.1145/3491102.3501889]
54. Open software foundation. URL: https://osf.io/b4a9m/ [accessed 2025-01-24]

Abbreviations

NHS: National Health Service
PCA: principal component analysis
PEx: personal experience
SEM: structural equation modeling

Edited by R Cuomo; submitted 09.04.24; peer-reviewed by A Frik, J Rowley, S Pesälä; comments to author 06.09.24; revised version received 27.10.24; accepted 21.11.24; published 13.02.25.

Please cite as:
Sillence E, Branley-Bell D, Moss M, Briggs P
A Model of Trust in Online COVID-19 Information and Advice: Cross-Sectional Questionnaire Study
JMIR Infodemiology 2025;5:e59317
URL: https://infodemiology.jmir.org/2025/1/e59317
doi:10.2196/59317
PMID:

Original Paper

Transformer-Based Tool for Automated Fact-Checking of Online Health Information: Development Study

Azadeh Bayani[1,2], MSc; Alexandre Ayotte[1,2*], MSc; Jean Noel Nikiema[1,2,3*], PhD

[1]Laboratoire Transformation Numérique en Santé, LabTNS, Montreal, QC, Canada

[2]Centre de recherche en santé publique, Université de Montréal et CIUSSS du Centre-Sud-de-l'Île-de-Montréal, Montreal, QC, Canada

[3]Department of Management, Evaluation and Health Policy, School of Public Health, Université de Montréal, Montreal, QC, Canada

[*]these authors contributed equally

Corresponding Author:
Azadeh Bayani, MSc
Laboratoire Transformation Numérique en Santé, LabTNS
7101 Av. du Parc, Montréal,
Montreal, QC, H3N 1X9
Canada
Phone: 1 4389980241
Email: azadeh.bayani@umontreal.ca

Abstract

Background: Many people seek health-related information online. The significance of reliable information became particularly evident due to the potential dangers of misinformation. Therefore, discerning true and reliable information from false information has become increasingly challenging.

Objective: This study aimed to present a pilot study in which we introduced a novel approach to automate the fact-checking process, leveraging PubMed resources as a source of truth using natural language processing transformer models to enhance the process.

Methods: A total of 538 health-related web pages, covering 7 different disease subjects, were manually selected by Factually Health Company. The process included the following steps: (1) using transformer models of bidirectional encoder representations from transformers (BERT), BioBERT, and SciBERT, and traditional models of random forests and support vector machines, to classify the contents of web pages into 3 thematic categories (semiology, epidemiology, and management), (2) for each category in the web pages, a PubMed query was automatically produced using a combination of the "WellcomeBertMesh" and "KeyBERT" models, (3) top 20 related literatures were automatically extracted from PubMed, and finally, (4) the similarity checking techniques of cosine similarity and Jaccard distance were applied to compare the content of extracted literature and web pages.

Results: The BERT model for the categorization of web page contents had good performance, with F_1-scores and recall of 93% and 94% for semiology and epidemiology, respectively, and 96% for both the recall and F_1-score for management. For each of the 3 categories in a web page, 1 PubMed query was generated and with each query, the 20 most related, open access articles within the category of systematic reviews and meta-analyses were extracted. Less than 10% of the extracted literature was irrelevant; those were deleted. For each web page, an average of 23% of the sentences were found to be very similar to the literature. Moreover, during the evaluation, it was found that cosine similarity outperformed the Jaccard distance measure when comparing the similarity between sentences from web pages and academic papers vectorized by BERT. However, there was a significant issue with false positives in the retrieved sentences when compared with accurate similarities, as some sentences had a similarity score exceeding 80%, but they could not be considered similar sentences.

Conclusions: In this pilot study, we have proposed an approach to automate the fact-checking of health-related online information. Incorporating content from PubMed or other scientific article databases as trustworthy resources can automate the discovery of similarly credible information in the health domain.

(JMIR Infodemiology 2025;5:e56831) doi:10.2196/56831

KEYWORDS

fact-checking automation; transformers; infodemic; credible health information; machine learning; automated; online health information; misinformation; natural language processing; epidemiology; health domain

Introduction

With rapid progressions in the digital age, and the vast dissemination of textual information available online, the likelihood of coming across misinformation has surged [1,2]. Misinformation refers to information that is untrue, incorrect, or deceptive in nature [3]. It is prevalent across various domains, with social media being a particularly prominent source [4]. Indeed, many people seek health-related topics on modern platforms and websites available online [5]. Inaccurate health-related information, however, poses an even greater risk, as it can directly impact lives [6,7]. Health misinformation is considered “a health-related claim or information which is not correct due to a lack of scientific evidence or knowledge” [4,8]. The importance of trustworthy online health information became particularly clear during the COVID-19 pandemic, which triggered a new crisis known as the COVID-19 infodemic. An infodemic refers to the excessive spread of false or misleading information across both digital and physical spaces [9] causing confusion and detrimental outcomes, as it underscores the potential risks posed by inaccurate or deceptive information to individuals [3,10]. The infodemic often manifests across 4 key areas: scientific research, policy and health care practice, news outlets, and social media platforms [11]. As a result, distinguishing between true and reliable information and falsehoods has become increasingly challenging. The labor-intensive process of manually verifying information specifically in health-related fields demands expert oversight and consumes significant time [4,9,12]. Therefore, it is crucial to establish an automated fact-checking process to help users identify the accuracy of health-related information available online.

The fact-checking process involves evaluating the truthfulness of information and consists of 3 key tasks: claim detection, evidence retrieval, and claim verification [12]. The first 2 tasks can be considered as factual verification, while the third focuses on assessing the accuracy of claims, which involves distinguishing reliable information from falsehoods to establish their factual validity [13].

Several studies have explored automating the fact-checking process, primarily focusing on misinformation in the form of fake news on websites [4,14,15] or social media [2,7,16-18]. These studies have generated synthetic datasets as the gold standard to facilitate the automation of evidence-based fact-checking. Thus, they compiled datasets comprising information or claims along with their corresponding evidence from trusted sources. Models were then trained using these datasets to automate the fact-checking process [7,10,15,17-20]. To create a database of verified claims, they used methods such as modifying phrases from Wikipedia [20], manual selection of quotation sentences and handpicking of claims from health news sites [14,15,21], and automatic selection of verified claims that were manually done by experts of journalists from fact-checking websites [10]. For example, the FEVER dataset, generated by modifying sentences taken from Wikipedia, consisted of 185,400 claims [22]. PUBHEALTH is another dataset containing false, true, unproven, and a mixture of health-related claims. The dataset also had a column containing journalist-crafted, gold-standard explanations designed to substantiate the fact-check labels assigned to each claim [6,18]. While synthetic datasets provide valuable contributions to advancing automatic fact-checking efforts, they cannot fully address real-world challenges, particularly the need for real-time, dynamic information [23]. Therefore, there is a need that claims and their associated evidence to be automatically extracted [24]. A study [25] developed a Large Language Model called TrumorGPT, which addresses limitations in fact-checking by incorporating retrieval-augmented generation and using continually updated knowledge graphs. This approach uses few-shot learning, knowledge graph construction, and semantic reasoning, which enhances the model’s ability to handle fact-checking tasks effectively. Another recent survey [12] explored automated techniques for predicting the veracity of claims, relying on natural language processing, knowledge representation, and databases. This study identified common challenges in fact-checking research and emphasized the importance of information retrieval and knowledge representation, particularly due to the rapid emergence of new claims.

Therefore, a key element of fact-checking involves identifying credible sources, and for health information, leveraging up-to-date scientific literature is essential as it is widely regarded as 1 of the most trustworthy references [26]. Indeed, numerous platforms and databases provide access to health-related and scientific literature, including Google Scholar, PubMed, ScienceDirect, and Web of Science, among others. These databases can be used as a reliable source for the automation of all the processes.

Numerous organizations have established guidelines to aid users in identifying trustworthy claims [27,28] where time-consuming manual recognition plays an important role in the process. In this pilot study, we proposed a novel automated evidence-based fact-checking approach that aims to identify and confirm accurate, truthful information using scientific literature and research databases as sources of truth. This exploratory evaluation highlights how using this approach may help users measure the extent of confidence in a web page and make informed decisions about accepting the health-related information of a website. Thus, the objective was to assess the truthfulness of health-related information through an evidence-based approach, without creating a synthetic database of claims-evidence but leveraging PubMed as a reliable source of fine-grained and up-to-date health-related information.

Methods

Approximately 1000 web pages were provided by Factually Health company on January 31, 2023. This company specializes in identifying reliable health-content websites [29]. The web pages were selected through random sampling within various disease categories to ensure a balanced dataset while minimizing the risk of overrepresentation of any single category. This approach accounted for variations in the number of available websites across disease categories. The web pages then underwent manual cleaning. Redundant pages were removed, and those unsuitable for research were excluded based on the following criteria: pages primarily featuring video content, pages related to clinical studies, pages resembling anecdotes rather than factual health information, or pages that restricted data extraction by Python (Python Software Foundation) libraries.

After this process, a dataset comprising 538 web pages was finalized. These web pages represented a diverse range of diseases, including arthritis (81 pages), chronic obstructive pulmonary disease (79 pages), COVID-19 (66 pages), hypertension (66 pages), lung cancer (70 pages), prostate cancer (66 pages), and diabetes (110 pages).

The selection of diverse disease categories was intended to minimize potential bias in the analysis. However, our previous study demonstrated that the selected diseases did not significantly impact classification results [29]. Using the URLs of each web page, the content was extracted as text files using the "justext" library in Python, to remove additional links and extraneous content from websites, such as navigation links, headers, and footers.

The process included the following three steps: (1) Classification of web page content into 3 thematic categories, semiology, epidemiology, and management by evaluating various transformer models, including bidirectional encoder representations from transformers (BERT), SciBERT, and BioBERT, as well as traditional models such as random forest (RF) and support vector machine (SVM), (2) automating the creation of PubMed queries combining "WellcomeBertMesh" and "KeyBERT" models, (3) automatic extraction of top 20 related literatures from PubMed, and (4) applying similarity checking techniques of cosine similarity and Jaccard distance to compare the content of extracted literature and web pages vectorized using BERT tokenizer. As a reliable source of truth, PubMed was a suitable choice to find evidence for health-related claims. PubMed, an open-source platform dedicated to facilitating searches and retrieval of health-related literature, encompasses over 36 million papers [30].

Classification of Web Page Contents

One of the necessary stages before determining the veracity of a claim or information is to detect the sentences that need to be verified [31]. These claims are crucial to the content's main point but require verification through an annotation schema and developing a benchmark for automated claim detection [14,31]. To detect sentences that need to be verified, two major steps were taken: (1) the identification of 3 thematic categories of content and (2) the classification of web page content according to these categories.

The Content Categories

To compare web page content with materials from the scientific literature database, it was essential to categorize the content, ensuring that comparisons were made within the relevant subject. Three distinct thematic categories have been identified for analysis: epidemiology, semiology, and management. In the epidemiology category, we included all sentences related to the statistics of a disease, the population, the frequencies, the causes, the risk assessment of the disease, and all public health-related information about the disease (eg, as of 2014, the global prevalence rate of rheumatoid arthritis was about 0.24%). In the semiology category, we considered all sentences related to signs (eg, high blood pressure is another sign of the disease) and symptoms (eg, this disease has symptoms such as pain, discomfort, weakness, fatigue). Finally, for the management category, we considered all the sentences linked to therapeutic approach (eg, drug treatment and surgical intervention, prevention, and the element of paraclinical diagnosis of diseases (eg, a complete medical examination carried out by a doctor can better determine if a person has chronic obstructive pulmonary disease and the degree of severity of the disease).

Manual Annotation and Model Development

Two authors (AB and AA) independently annotated 200 web pages on a sentence-by-sentence basis considering the 3 categories of epidemiology, semiology, management, and neutral until reaching a roughly balanced amount of data across all classes [32]. We used the Cohen κ score to assess the agreement between the 2 reviewers AB and AA). Any discrepancies were resolved by the third author (JNN).

Neutral sentences were those that did not correspond to any of the defined thematic categories. Table 1 shows the distribution of sentences for each category. The portable serverless text annotation tool of MedTator-1.3-11 [33] was used for the annotation process. A total of 3 transformer models of BERT, SciBERT, and BioBERT were used to classify the sentences into the 4 mentioned categories. The BERT model has demonstrated superior performance in several text classification tasks [29,34,35]. SciBERT is an extension of BERT and is trained on a vast corpus of scientific literature spanning multiple domains [36] and BioBERT is pretrained using an extensive corpus comprising PubMed abstracts (PubMed) and full-text articles from PubMed Central [37]. We have also conducted a performance comparison between the transformer models and 2 traditional machine learning models: RF and SVM.

Table 1. The distribution of classes.

Category	Number of sentences
Neutral	3162
Semiology	851
Epidemiology	1171
Management	1066

The “BertTokenizer” library has been used to tokenize the incoming sentences, with the following parameters: We applied a maximum sequence length of 128 to standardize the size of each input sentence. To optimize the model's hyperparameters, we applied the Bayesian optimization approach using the ‘BayesianOptimization’ library in Python. The hyperparameter tuning spaces are detailed in Table 2.

Table 2. Hyper-parameter tuning search space.

Hyper-parameters	Range	Best trial
Learning rate	10^{-7}, 10^{-2}	3×10^{-5}
Weight decay	10^{-5}, 10^{-1}	10^{-3}
Number of epochs	(1:5)	3
Batch size	(8,16,32,64)	32

Automating PubMed Query Generation

Overview

Literature extraction involved identifying scientific articles within PubMed to support the process. To achieve this, the approach requires the formulation of a query by combining keywords and Medical Subject Headings (MeSH) terms, which can be extracted from web page content. This process included three steps: (1) Automating PubMed subquery creation from MeSH terms and creating a subquery using the “WellcomeBertMesh” model, (2) Automating PubMed subquery creation from keywords using KeyBert model and creating a subquery, and (3) Construction of the final query by combining the different subqueries.

Automating PubMed Subquery Creation Using MeSH Terms Extracted by Transformers

All the MeSH terms were extracted from the text using a pretrained model of “WellcomeBertMesh,” which takes its inspiration from “BertMesh,” which undergoes the pretraining using the entire text of biomedical publications and is built upon the foundation of the BioBert pretrained model [38]. Given that our evidence for the websites primarily comprised health-related articles from PubMed, we selected this model. Its architecture is rooted in the latest advancements in the biomedical field, prominently featuring Microsoft’s cutting-edge “PubMedBert” as its core framework [38].

To enhance the accuracy of the subquery, the identified MeSH terms were initially organized according to their MeSH categories to construct subsubqueries. The MeSH has a tree structure that is organized hierarchically, visually presenting descriptors in broader and narrower relationships. The top tier of the MeSH tree structure encompasses 19 comprehensive categories. While these terms are not included in MeSH data maintenance and distribution, they can be used to search PubMed by using the search term “category” [39]. Therefore, we have considered the MeSH terms under each head category together using the “OR” operator in this subsubquery. Then, we constructed the subquery using the “AND” operator between extracted MeSH terms in different categories. The pseudo-code for this step is presented in Figure 1.

Figure 1. MeSH (medical subject heading) subquery builder.

```
Input: A list of sentences belonging to a web page S = [s1, s2, …] for a specific category
Input: category to consider ∈ {Epidemiology, Semiology, Management}
Output: A PubMed query extracted from the web page

model ←Load the "WellcomeBertMesh" pre-trained model
/* iterating through sentences to compute their vector representation then extracting the MeSH terms corresponding
to each sentence: */
for i ← 1, n do:
        vi ← model_vector (si)
        meshi ← model (vi)
end for
/* identifying the head categories for each MeSH term extracted*/
for j ← 1, length(mesh) do:
        categoryj ← extract_mesh_head_category(meshj)
end for
/* creating subqueries based on the MeSH terms belonging to the same or different categories */
for i ← 1, n do:
        for k ← 1, K do:
            sub-subquery1, sub-subquery2 ← null
        /* put OR for mesh terms in the same category, put AND for different categories*/
                if meshi belong to same categoryk then
                    sub-subquery1 ← (meshi OR sub-subquery1)
                Else
                  sub-subquery2 ← (meshi AND sub-subquery2)
                end if
            MeSH-sub_query ← (sub-subquery1 AND sub-subquery2)
        end for
end for
```

Automating PubMed Subquery Creation Using Key Phrases Extracted by Transformers

The key phrases from web page contents have been extracted using the transformer model "KeyBERT" library, which is described in previous literature as having the best performance in extracting the key phrases [40], especially for long texts [41], which aligns with our need of extracting the key phrases of the scientific papers. The extracted keywords were combined with the "AND" operator to create a subquery.

Figure 2 shows the proposed pseudo-code to extract the keywords for the creation of the subquery.

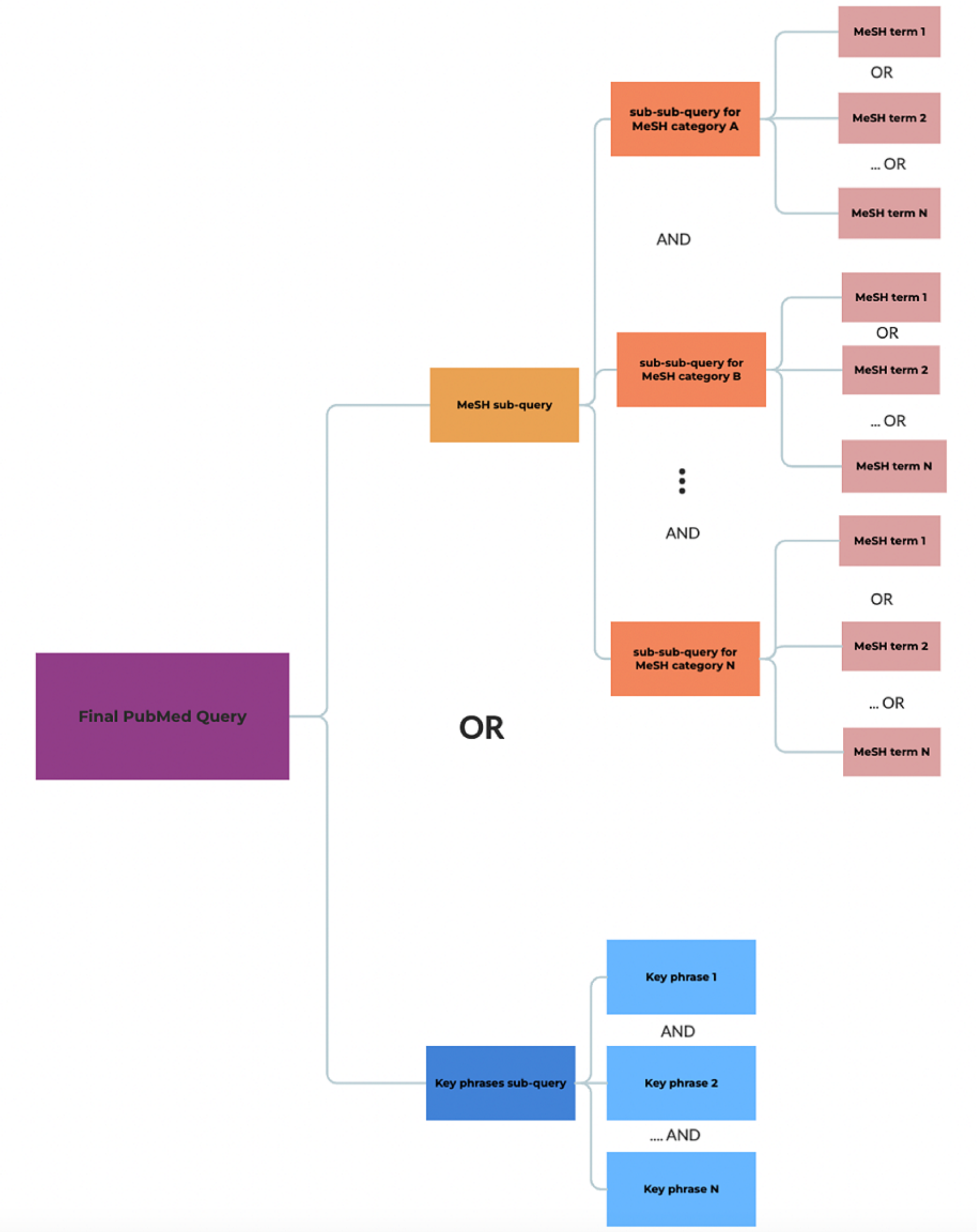
Figure 2. Key phrase extractor and subquery builder.

```
Input: A list of sentences belonging to a web page S = [s1, s2, …] for a specific category
Input: category to consider ∈ {Epidemiology, Semiology, Management}
Output: A PubMed query extracted from the web page
model ←Load the "KeyBERT" pre-trained model
/* computing a vector representation and extracting the key phrases corresponding to each sentence */
for i ← 1, n do
        vi ← model_vector (si)
        keyphrasei ← model (vi)
end for
/* creating key phrase subquery tailored to the specified categories */
keyphrase_query ← null
for i ← 1, length(keyphrases) do
        keyphrase_query ← (keyphrase_query AND keyphrasei)
end for
```

Construction of the Final Query

The subqueries extracted from the preceding processes were combined using the "OR" operator to construct the final query. Figure 3 presents a comprehensive overview of the process used to construct the final PubMed query, summarizing the structure and strategy behind its creation.

Figure 3. Detailed process diagram for the development of the comprehensive final PubMed query.



Automating Related Literature Extraction

The final query was used to retrieve a compilation of articles, from which the top open access 20 resulting papers were extracted. The “PMC_ids” of papers were extracted using the “Entrez” library of Python that provides integrated access to PubMed Medline [42]. To evaluate the quality of our query results, we conducted a comprehensive review of the obtained full-text papers. In our assessment of the extracted papers in PubMed, those subjected to filtering within the systematic reviews and meta-analysis category exhibited more related papers to the subject of the research, compared with papers that were not subject to such filtering. Consequently, we selected them to encompass a wider range of relevant articles.

Finally, the automatically extracted papers were manually checked to be pertinent considering the title of the papers, the irrelevant papers were removed and excluded from the final process.

Similarity Detection and Fact-Checking

For the process of computing the similarity measure between different sentences, for each disease, we randomly selected 5 web pages in our dataset. For each of the 3 predefined thematic categories in a web page, 1 PubMed query was generated and with each query, the 20 most related, open access articles within the type of systematic reviews and meta-analysis were extracted. The following steps were then carried out: (1) Categorizing the extracted related literature content based on the 3 thematic categories. This was necessary to analyze sentences (from websites and scientific articles) that are relevant to the same topics. (2) Comparing by thematic category, the content from scientific articles and web pages to identify similar sentences.

Finally, after conducting a manual evaluation of the identified similar sentences, we calculated the average number of categorized sentences for each randomly selected web page, as well as the average number of credible sentences detected. Credible sentences refer to those in the related literature that demonstrated similarity with the sentences from the web pages.

Categorizing the Extracted Literature

The more performant fine-tuned model on the web page contents was used to categorize literature contents into 3 thematic categories. This approach enabled us to facilitate a direct comparison between sentences sharing the same thematic context.

Comparing the Content From Literature and Web Pages to Identify Similar Sentences

For the sentence comparison, we used the BERT vectorizer to transform the texts into vectors. This allowed us to encode the semantic significance of sentences as numerical values, facilitating the application of different similarity detection algorithms [43].

Both scientific articles and web page sentences were transformed into vector representations, taking into account their respective thematic categories. Subsequently, each web page sentence was compared with scientific article sentences of the same category using the cosine similarity and Jaccard technique. A similarity threshold of 87% was chosen to determine sentence selection, ensuring that sentences with over 87% similarity were chosen.

Figure 4 shows the proposed pseudo-code for the similarity-checking part.

Figure 4. Paper similarity detection.

```
Input: A list of sentences belonging to a web page and papers S = [s1, s2, …], P = [p1, p2, …]
Input: category to check ∈ {Epidemiology, Semiology, Management}, similarity_threshold
Output: percentage of similarity between two contents
model ←Load the "bert_base_uncase" pre-trained model
/* computing vector representation of paper sentences */
for i ← 1, n do:
        vi , v'i← model_vector (si, pi)

end for
/* computing the percentage of similarity between the contents of the web page and the papers*/
for k ← 1, length (vi) do
/* if the similarity between web page and the paper sentences be more than threshold*/
        if Cosine_similarity (vi, v'j∈(1…n)) > similarity_threshold then
          print (corresponding sentences of (vi, v'j))
        end if
end for
compute similarity percentage
```

For each disease, we randomly selected 5 web pages and extracted both their related papers and similar sentences. It was due to the inherent variability and specificity of medical information related to each disease. Diseases often exhibit unique characteristics, nuances, and clinical considerations. By prioritizing diseases, we aimed to provide a more granular and clinically relevant assessment of the similarity between the sentences. The outcomes, comprising sentences from the web pages and their corresponding similar sentences, underwent a manual verification by the authors to ensure semantic similarity between them. Subsequently, the proportion of semantically similar sentences between a web page and its related reference papers was calculated.

Ethical Considerations

This research relied solely on publicly accessible data and did not involve any human or animal participants, making it exempt from the need for ethical approval. The study strictly adheres to established data privacy norms to prevent any compromise of confidentiality or privacy. In addition, the project does not include any direct involvement or interactions with individuals, thereby minimizing potential ethical issues. The University of

Montreal's Research Committee has carefully examined our methodology and affirmed that this study falls outside the scope of Medical Research Involving.

Results

This section elaborates on the results of each part of the proposed pseudo-codes.

Classification of Web Page Contents

The annotation process for web page contents achieved a Cohen κ score of 87% among the 2 annotators (AA and AB), indicating high agreement between the annotators and ensuring the reliability of the data used for model evaluation.

The performance of transformer-based models (BERT, BioBERT, and SciBERT) was compared to traditional machine learning models (RF and SVM) for categorizing web page content into four categories. BERT emerged as the most effective model, consistently achieving superior precision, recall, and F_1-scores across all categories. Traditional models, in contrast, demonstrated lower performance, particularly in terms of F_1-scores, indicating limitations in balancing precision and recall effectively.

Table 3 illustrates the performance of the classification models used to classify the content of web pages. The performance matrix includes metrics such as precision, recall, and F_1-score.

Table 3. Performance evaluation of the BERT (Bidirectional Encoder Representations from Transformers) and machine learning models for web page content classification across considered categories.

	BERT[a]			BioBERT			SciBERT			RF[b]			SVM[c]		
Classes	Precision	Recall	F_1-score	Precision	Recall	F_1-score	Precision	Recall	F_1-score	Precision	Recall	F_1-score	Precision	Recall	F_1-score
Neutral	0.96	0.93	0.95	0.88	0.83	0.85	0.85	0.81	0.83	0.51	0.92	0.66	0.72	0.81	0.77
Semiology	0.91	0.94	0.93	0.81	0.81	0.81	0.77	0.79	0.78	0.96	0.05	0.09	0.71	0.59	0.64
Epidemiology	0.92	0.94	0.93	0.80	0.76	0.76	0.75	0.74	0.75	0.8	0.1	0.1	0.69	0.62	0.65
Management	0.95	0.96	0.96	0.83	0.89	0.89	0.83	0.87	0.85	0.59	0.58	0.59	0.74	0.73	0.74

[a]BERT: Bidirectional Encoder Representations from Transformers.

[b]RF: random forests.

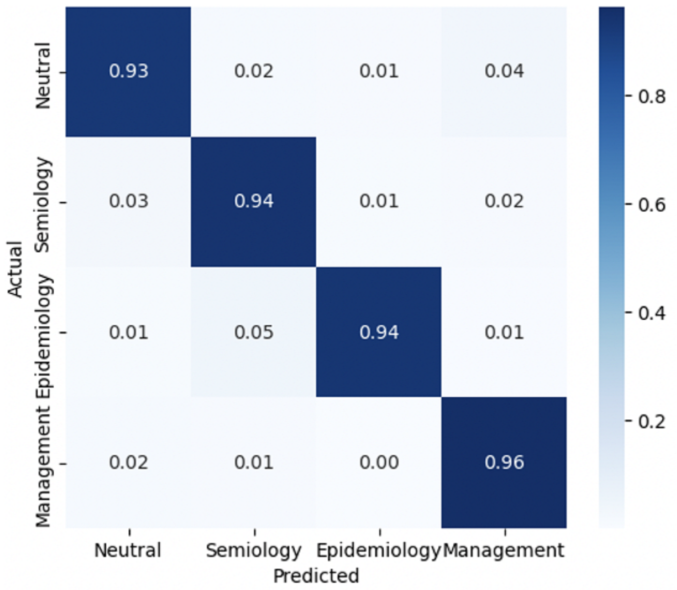
[c]SVM: support vector machines.

According to Table 3, among the transformer models, the BERT model had a promising performance with more than 93% recall for neutral sentences, 94% for semiology and epidemiology, and 96% for the management category. The model had an F_1-score of 95% for neutral sentences, 93% for semiology and epidemiology, and 96% for management. The model had 96% precision for neutral sentences, 91% for semiology, 92% for epidemiology, and 95% for management. Also, traditional models did not have high performance, the precision values for both RF and SVM were relatively low in some classes, indicating a high rate of false positives. Also, the F_1-scores for both RF and SVM were generally lower compared with the BERT model, indicating that they may not achieve a good balance between precision and recall. Therefore, the BERT model was selected for the classification of the web page contents.

The confusion matrix for the BERT model is shown in Figure 5, providing a detailed visualization of its classification performance across the different categories.

Figure 5 shows the confusion matrix for the BERT classifier, which correctly classified 0.93 of the neutral sentences, 0.94 for both the semiology and epidemiology sentences, and 0.96 for management sentences as true positives.

Figure 5. Bidirectional encoder representations from transformers model performance: confusion matrix for the classification of web page sentences into 3 thematic categories.



Automating PubMed Query Generation

To extract relevant literature for the web pages categorized thematically, a PubMed query was generated for each of the 7 diseases. Each query retrieved the 20 most related papers. The titles of the retrieved papers were manually evaluated, and less than 10% were deemed irrelevant, demonstrating the effectiveness of the generated queries. These irrelevant articles were excluded from further analysis.

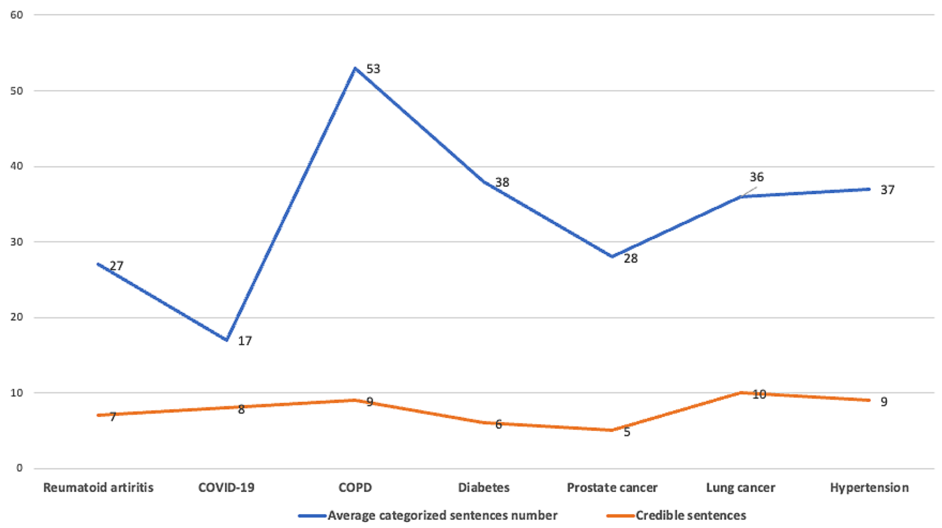
This result highlights the utility of using MeSH terms and key phrases in constructing PubMed queries, which efficiently yielded pertinent literature. The generated weblinks for accessing the papers followed the format: “https://pubmed.ncbi.nlm.nih.gov/PMID/,” with PMIDs obtained directly from the PubMed queries.

Similarity Detection and Fact-Checking

Figure 6 illustrates the average percentage of credible information found in the 5 randomly selected web pages categorized by related diseases. Credible information is defined as sentences in the web pages that were successfully matched with corresponding sentences in PubMed articles.

On average, 23% of the sentences on each web page were identified as similar to statements in the scientific literature. While this demonstrates the potential of the system to detect credible content, a significant challenge arose with false positives. Some sentences achieved a similarity score exceeding 80% but were semantically dissimilar upon closer inspection.

Figure 6. The average number of credible sentences on web pages (red line) versus the average number of all sentences on each web page (blue line). COPD: chronic obstructive pulmonary disease.



For instance, the following sentences from an extracted paper and a web page had a similarity score of 88% yet conveyed different meanings:

1. “Previous studies have documented residual symptoms that continue 12 weeks after the onset of acute COVID-19, known as post-acute or long COVID-19.”

2. “The acute phase of COVID itself can last for up to 14 days.”

This highlights the need for more sophisticated approaches to accurately distinguish between syntactic similarity and genuine semantic alignment.

As an illustrative example, for the rheumatoid arthritis category, we randomly selected 5 web pages, each containing an average of 27 sentences distributed across 3 thematic categories: epidemiology, semiology, and management (represented by the blue line). Among these, an average of 7 sentences per web page were deemed credible and successfully matched to corresponding statements in the scientific literature (depicted by the red line).

Discussion

Principal Findings

In the present pilot study, our objective was to automate aspects of the fact-checking process for online health information. While previous research [21,26] has explored automation in various stages of fact-checking, such as evidence retrieval or claim identification, this pilot serves as an initial step toward achieving full automation in the fact-checking process. Our approach includes the automation of identifying verifiable sentences through a classification process. Notably, our study used a fine-tuned BERT model, which exhibited notable efficacy in categorizing health-related sentences. Although BioBERT and SciBERT models have been reported to outperform BERT in various downstream tasks [36,37], in our investigation, the BERT model demonstrated superior performance. This discrepancy could be attributed to BERT training on general-purpose texts, such as Wikipedia or Book Corpus [35], which align more closely with the content of websites targeted at general populations. In contrast, BioBERT and SciBERT are trained on more specialized texts, such as scientific publications [36,37].

Previous research [14,31,44] has shown that the identification of claim-worthy sentences or the recognition of key information needing verification from reliable sources is a fundamental first step in automating the fact-checking process akin to our approach. This process is commonly structured as a text classification task. The previous studies used human annotators [44] or crowdsourcing [31] to tag claim-worthy sentences and trained machine learning models to classify them. A previous study [14] focused on detecting claims within news and public information, assigning each sentence a likelihood score for containing significant factual claims. Also, automating the fact-checking process is far from straightforward, as it necessitates the utilization of artificial intelligence tools to struggle with the complexity of text and context [10]. Studies often considered the problem as a binary classification to split the contents into credible or non-credible information, however, the decision is more complex since there may be several ambiguities in the sentences. In addition, several parts of the process depend on human judgment, which needs further research in the area. Building on this groundwork, our study applied a BERT-based classification approach to detect health information requiring verification and automatically proposing a sentence for this process. Previous studies relied on reviewer selections to develop claim and evidence datasets, lacking attempts to automate claim identification with real-world resources [17,18,45].

In addition, rather than constructing a manual reference dataset as the evidence for verifiable sentences, we leveraged the PubMed database as our source of truth. We automated the detection of evidence for claims made on web pages in an unsupervised approach, streamlining the verification process. This aligns with previous studies [21,26] that used PubMed publications as evidence, using transformer models to generate queries and retrieve documents from PubMed. We demonstrated the effectiveness of using transformer models to extract MeSH terms and key phrases from web page content, enabling the efficient generation of PubMed queries. This approach facilitated the retrieval of related articles from scientific references without requiring supervision. According to a previous study [14], to verify the veracity of the claims, it is crucial to translate them into queries against the reference databases. However, other studies [6,20,22] created a knowledge database as the references to compare with the claims. Notably, Sarrouti et al [6] introduced a dataset comprising evidence-claim pairs, manually annotated as SUPPORT, REFUTE, and NEUTRAL. They used BERT-based models to create a realistic testing ground for evidence-based fact-checking systems.

To assess the alignment between claim sentences and extracted references, we measured their similarity, a practice supported by [46]. This study underscores the necessity for a model in claim verification to measure the semantic similarity between claims and verified factual knowledge or references. To compare the semantic similarity, we used a transformer-based representation that converted the textual content into vectorial representation, allowing us to capture the contextual nuances of each sentence consistent with previous approaches [19,43,47]. This approach is more efficient and produces semantically richer sentence representations than simply averaging the vectors of words that appear in each sentence, and facilitates the similarity detection for the algorithms [48]. We successfully identified factual evidence for 23% of the health-related information extracted from web pages, indicating the complexity inherent in health information. Further research is required to enhance contextual comparison between claims and verified references. Also, the cosine similarity outperformed the Jaccard distance measure for comparing the claims and evidence in this study, which is different from the previous study [4], as they reported that the Jaccard distance was better at the similarity selection measure. The reason may be due to differences in the nature of the datasets in the 2 studies.

Limitations

This study had several limitations. First, we faced a challenge in identifying sentences within the papers that closely matched the content of the web pages. Numerous methods have been devised to tackle this issue [19,43,46]; however, a comprehensive consideration of the complete meaning of sentences requires further investigation. In addition, 77% of the sentences did not have matching counterparts in the academic literature that we retrieved. Regarding this proportion, 2 possible assumptions can be made: either the sentences themselves were not valid or the algorithm was unable to locate their related counterparts. Another potential reason could be that the sentences, though addressing a common subject such as the same medical condition, exhibited variations in meaning or contextual interpretation. Consequently, it would be premature to assert that these unmatched sentences are inherently not credible, given the vast volume of published papers that renders comprehensive verification computationally infeasible. Expanding the number of selected papers for comparison could therefore increase the likelihood of identifying additional relevant sentences in the literature. Nonetheless, quantifying the proportion of credible sentences offers valuable insights to aid users in their trust assessment.

It is worth acknowledging that authors in the realm of health-related data often simplify and rephrase content to cater to their target audience, making it more challenging to identify credible references for their statements. Therefore, the researchers propose exploring other models such as text generation models as potential solutions to address this particular challenge including WordNet or sequence-to-sequence (Seq2Seq) models.

A second limitation was the sample size of the academic papers used in the comparison. Due to the extensive volume of health-related publications, the assessment was limited to a selection of 20 papers. Expanding this scope to include more papers per content type could enhance the discovery of factual evidence in PubMed publications. Thus, further investigation into paper retrieval approaches is recommended.

A third limitation was that, although the thematic categorization of web page content, such as epidemiology, semiology, and management, ensured that the generated PubMed queries were more precise and contextually relevant, the need for quality assessment of the extracted PubMed articles remains evident. While our method provides users with essential information to assess the accuracy of health information, the ultimate determination of its truthfulness may depend on individual judgment, expert evaluation, source credibility, scientific article quality (eg, journal quality, impact factor for the domain) and the contemporaneity of the information (eg, date of publication, retracted).

The retrieved articles may vary in quality, ranging from high-impact studies to potentially outdated or retracted articles that could influence the reliability of the fact-checking process and the conclusions drawn from matched content. Addressing these characteristics within an automated process remains a key challenge. In our previous research, the credibility of the sources was automatically assessed [29]. In this study, while we evaluate comparability with scientific articles, developing a credibility scoring strategy for these articles is also necessary. Combining an algorithm that evaluates website credibility and assigns a credibility score to scientific articles with 1 that determines truthfulness could significantly enhance the effectiveness of fact-checking. These models can change the structure of sentences and may improve the possibility of finding more similar sentences. Finally, while the process could not be automated entirely since each step needed human supervision for the results, the suggested techniques have the potential to substantially alleviate the human effort required to locate valid information.

Conclusions

Our approach aimed to empower users in the decision-making process regarding the truthfulness of information by providing relevant evidence and enabling informed judgments. As a pilot, this research serves as an initial step toward exploring the feasibility of automating fact-checking processes in health information. Specifically, the methods presented here could be applied to create tailored fact-checking workflows for specific disease areas, such as diabetes, arthritis, or cancer, which were among the categories included in this study. For instance, thematic categorization (eg, management and epidemiology) could improve the precision and relevance of fact-checking tools in health care contexts. Using state-of-the-art models such as transformers may improve the performance of the model since the BERT embedding captures the meaning of the sentences [49]. The investigation also revealed that incorporating PubMed publications as a trustworthy resource can enhance the discovery of similar credible information as evidence. Finally, while the process could not be entirely automated and required human supervision, the suggested techniques demonstrate significant potential for integration into fact-checking tools. This integration could reduce the effort required to validate health information, ultimately increasing accessibility and reliability for end-users. Future work should focus on expanding the dataset and testing the approach in real-world scenarios to further refine its applicability across various health domains.

Acknowledgments

The authors express their sincere gratitude to Factually Health Company for generously providing a factual dataset covering 7 diseases and to IVADO Labs for their support.

Data Availability

The datasets generated and analyzed during this study are available from the corresponding author on reasonable request.

Conflicts of Interest

This research was funded by the Mitacs acceleration program in partnership with Factually Health Company, and the IVADO Funding for Collaborative Research in Data Science to Serve Sustainable Development.

References

1. Cuan-Baltazar JY, Muñoz-Perez MJ, Robledo-Vega C, Pérez-Zepeda MF, Soto-Vega E. Misinformation of COVID-19 on the internet: infodemiology study. JMIR Public Health Surveill 2020 Apr 09;6(2):e18444 [FREE Full text] [doi: 10.2196/18444] [Medline: 32250960]
2. Himelein-Wachowiak M, Giorgi S, Devoto A, Rahman M, Ungar L, Schwartz HA, et al. Bots and misinformation spread on social media: implications for COVID-19. J Med Internet Res 2021 May 20;23(5):e26933 [FREE Full text] [doi: 10.2196/26933] [Medline: 33882014]
3. Lee DKL, Ramazan O. Fact-checking of health information: the effect of media literacy, metacognition and health information exposure. J Health Commun 2021 Jul 03;26(7):491-500. [doi: 10.1080/10810730.2021.1955312] [Medline: 34292862]
4. Barve Y, Saini JR. Healthcare misinformation detection and fact-checking: a novel approach. IJACSA 2021;12(10). [doi: 10.14569/IJACSA.2021.0121032]
5. Ecker UKH, Lewandowsky S, Cook J, Schmid P, Fazio LK, Brashier N, et al. The psychological drivers of misinformation belief and its resistance to correction. Nat Rev Psychol 2022 Jan 12;1(1):13-29. [doi: 10.1038/s44159-021-00006-y]
6. Sarrouti M, Abacha AB, Mrabet Y, Demner-Fushman D. Evidence-based fact-checking of health-related claims. : Association for Computational Linguistics; 2021 Presented at: Findings of the Association for Computational Linguistics: EMNLP 2021; 2025 January 22; Punta Cana, Dominican Republic p. 3499-3512. [doi: 10.18653/v1/2021.findings-emnlp.297]
7. Sumayyia MD, Al-Madaney MM, Almousawi FH. Health information on social media. Perceptions, attitudes, and practices of patients and their companions. Saudi Med J 2019 Dec;40(12):1294-1298 [FREE Full text] [doi: 10.15537/smj.2019.12.24682] [Medline: 31828284]
8. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of health misinformation on social media: systematic review. J Med Internet Res 2021 Jan 20;23(1):e17187 [FREE Full text] [doi: 10.2196/17187] [Medline: 33470931]
9. Hang CN, Yu PD, Chen S, Tan CW, Chen G. MEGA: Machine Learning-Enhanced Graph Analytics for infodemic risk management. IEEE J Biomed Health Inform 2023 Dec;27(12):6100-6111. [doi: 10.1109/JBHI.2023.3314632] [Medline: 37713230]
10. Altuncu E, Nurse J, Bagriacik M, Kaleba S, Yuan H, Bonheme L, et al. aedFaCT: scientific fact-checking made easier via semi-automatic discovery of relevant expert opinions. arXiv:2305.07796 2023. [doi: 10.36190/2023.27]
11. Eysenbach G. How to fight an infodemic: the four pillars of infodemic management. J Med Internet Res 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: 10.2196/21820] [Medline: 32589589]
12. Guo Z, Schlichtkrull M, Vlachos A. A survey on automated fact-checking. Transactions of the Association for Computational Linguistics 2022;10:178-206. [doi: 10.1162/tacl_a_00454]
13. Walter N, Cohen J, Holbert RL, Morag Y. Fact-checking: a meta-analysis of what works and for whom. Political Communication 2019 Oct 24;37(3):350-375. [doi: 10.1080/10584609.2019.1668894]
14. Hassan N, Arslan F, Li C, Tremayne M. Toward automated fact-checking: detecting check-worthy factual claims by ClaimBuster. : ACM; 2017 Presented at: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining; 2017 August 13-17; Halifax NS Canada p. 1803-1812. [doi: 10.1145/3097983.3098131]
15. Samadi M, Mousavian M, Momtazi S. Deep contextualized text representation and learning for fake news detection. Information Processing & Management 2021 Nov;58(6):102723. [doi: 10.1016/j.ipm.2021.102723]
16. Zhang J, Featherstone JD, Calabrese C, Wojcieszak M. Effects of fact-checking social media vaccine misinformation on attitudes toward vaccines. Prev Med 2021 Apr;145:106408. [doi: 10.1016/j.ypmed.2020.106408] [Medline: 33388335]
17. Zhong B. Going beyond fact-checking to fight health misinformation: A multi-level analysis of the Twitter response to health news stories. International Journal of Information Management 2023 Jun;70:102626. [doi: 10.1016/j.ijinfomgt.2023.102626]
18. Akhtar M, Cocarascu O, Simperl E. PubHealthTab: a public health table-based dataset for evidence-based fact checking. : Association for Computational Linguistics; 2022 Presented at: Findings of the Association for Computational Linguistics: NAACL 2022; 2025 January 22; Seattle, United States p. 1-16. [doi: 10.18653/v1/2022.findings-naacl.1]
19. Martín A, Huertas-Tato J, Huertas-García Á, Villar-Rodríguez G, Camacho D. FacTeR-Check: semi-automated fact-checking through semantic similarity and natural language inference. Knowledge-Based Systems 2022. [doi: 10.1016/j.knosys.2022.109265]
20. Sathe A, Ather S, Le TM, Perry N, Park J. Automated fact-checking of claims from Wikipedia. : European Language Resources Association; 2020 Presented at: Proceedings of the Twelfth Language Resources and Evaluation Conference; 2025 January 22; Marseille, France p. 6874-6882.
21. Deka P, Jurek-Loughrey A, P D. Improved methods to aid unsupervised evidence-based fact checking for online heath news. JDI 2022;3(4):474-504. [doi: 10.26421/JDI3.4-5]

22. Thorne J, Vlachos A, Christodoulopoulos C, Mittal A. FEVER: a large-scale dataset for fact extraction and VERification. ACL Anthology 2018:1-9. [doi: 10.18653/v1/w18-5501]
23. Das A, Liu H, Kovatchev V, Lease M. The state of human-centered NLP technology for fact-checking. Information Processing & Management 2023 Mar;60(2):103219. [doi: 10.1016/j.ipm.2022.103219]
24. Nakov P, Corney D, Hasanain M, Alam F, Elsayed T, Barrón-Cedeño A, et al. Automated fact-checking for assisting human fact-checkers.. arXiv Preprint posted online on March 13, 2021. [doi: 10.48550/arXiv.2103.07769]
25. Hang CN, Yu PD, Tan CW. TrumorGPT: query optimization and semantic reasoning over networks for automated fact-checking. : IEEE; 2024 Presented at: 2024 58th Annual Conference on Information Sciences and Systems (CISS); 2024 March 13-15; Princeton, NJ, USA p. 1-6. [doi: 10.1109/ciss59072.2024.10480162]
26. Deka P, Jurek-Loughrey A, Deepak. Unsupervised keyword combination query generation from online health related content for evidence-based fact checking. : ACM; 2021 Presented at: iiWAS2021: The 23rd International Conference on Information Integration and Web Intelligence; 2021 November 29; Linz Austria p. 267-277. [doi: 10.1145/3487664.3487701]
27. Boyer C, Baujard V, Nater T, Scherrer J, Appel R. HONcode: a standard to improve the quality of medical/health information on the internet and HON's 5th survey on the use of internet for medical and health purposes. In: J Med Internet Res. 1999 Presented at: 6th Internet World Congress for Biomedical Sciences (INABIS 2000); 2000 December 7-16; Canada. [doi: 10.2196/jmir.1]
28. Padhy S, Rathore SS. An attention ‐ based deep learning model for credibility assessment of online health information. Computational Intelligence 2023 Jul 24;39(5):832-859. [doi: 10.1111/coin.12596]
29. Bayani A, Ayotte A, Nikiema JN. Automated credibility assessment of web-based health information considering health on the net foundation code of conduct (HONcode): model development and validation study. JMIR Form Res 2023;7:e52995 [FREE Full text] [doi: 10.2196/52995] [Medline: 38133919]
30. Giustini D, Barsky E. A look at Google Scholar, PubMed, and Scirus: comparisons and recommendations. J Can Health Libr Assoc 2005 Sep 01;26(3):85-89. [doi: 10.5596/c05-030]
31. Konstantinovskiy L, Price O, Babakar M, Zubiaga A. Toward automated factchecking. Digital Threats 2021 Apr 15;2(2):1-16. [doi: 10.1145/3412869]
32. Fernando KRM, Tsokos CP. Dynamically weighted balanced loss: class imbalanced learning and confidence calibration of deep neural networks. IEEE Trans Neural Netw Learn Syst 2022 Jul;33(7):2940-2951. [doi: 10.1109/TNNLS.2020.3047335] [Medline: 33444149]
33. He H, Fu S, Wang L, Liu S, Wen A, Liu H. MedTator: a serverless annotation tool for corpus development. Bioinformatics 2022 Mar 04;38(6):1776-1778 [FREE Full text] [doi: 10.1093/bioinformatics/btab880] [Medline: 34983060]
34. González-Carvajal S, Garrido-Merchán EC. Comparing BERT against traditional machine learning text classification.. arXiv Preprint published online May 26, 2020. [doi: 10.48550/arXiv.2005.13012]
35. Koroteev MV. BERT: a review of applications in natural language processing and understanding.. arXiv Preprint published online March 22, 2021. [doi: 10.48550/arXiv.2103.11943]
36. Beltagy I, Lo K, Cohan A. SciBERT: a pretrained language model for scientific text. ACL Anthology 2019:3615-3620. [doi: 10.18653/v1/d19-1371]
37. Lee J, Yoon W, Kim S, Kim D, Kim S, So C, et al. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics 2020;36(4):1234-1240 [FREE Full text] [doi: 10.1093/bioinformatics/btz682] [Medline: 31501885]
38. You R, Liu Y, Mamitsuka H, Zhu S. BERTMeSH: deep contextual representation learning for large-scale high-performance MeSH indexing with full text. Bioinformatics 2021;37(5):684-692 [FREE Full text] [doi: 10.1093/bioinformatics/btaa837] [Medline: 32976559]
39. Dhammi IK, Kumar S. Medical subject headings (MeSH) terms. Indian J Orthop 2014 Sep;48(5):443-444 [FREE Full text] [doi: 10.4103/0019-5413.139827] [Medline: 25298548]
40. Priyanshu A, Vijay S. arXiv 2022. [doi: 10.48550/arXiv.2211.07499]
41. Giarelis N, Kanakaris N. A comparative assessment of state-of-the-art methods for multilingual unsupervised keyphrase extraction. In: Artificial Intelligence Applications and Innovations. Cham: Springer International Publishing; 2021:635-645.
42. Schuler GD, Epstein JA, Ohkawa H, Kans JA. Entrez: molecular biology database and retrieval system. Methods Enzymol 1996;266:141-162. [doi: 10.1016/s0076-6879(96)66012-1] [Medline: 8743683]
43. Qadrud-Din J, Rabiou A, Walker R, Soni R, Gajek M, Pack G, et al. Transformer based language models for similar text retrieval and ranking.. arXiv Preprint published online on May 10, 2020. [doi: 10.48550/arXiv.2005.04588]
44. Pathak A, Shaikh MA, Srihari R. Self-supervised claim identification for automated fact checking.. arXiv. :213-227 Preprint published online on February 3, 2021. [doi: 10.48550/arXiv.2102.02335]
45. Kotonya N, Toni F. Explainable automated fact-checking for public health claims. ACL Anthology 2020:7740-7754. [doi: 10.18653/v1/2020.emnlp-main.623]
46. Barve Y, Saini JR. Detecting and classifying online health misinformation with 'Content Similarity Measure (CSM)' algorithm: an automated fact-checking-based approach. J Supercomput 2023;79(8):9127-9156 [FREE Full text] [doi: 10.1007/s11227-022-05032-y] [Medline: 36644509]

47. Vijjali R, Potluri P, Kumar S, Teki S. Two stage transformer model for COVID-19 fake news detection and fact checking.. arXiv Preprint published online on November 26, 2020. [doi: 10.48550/arXiv.2011.13253]
48. Nabożny A, Balcerzak B, Wierzbicki A, Morzy M, Chlabicz M. Active annotation in evaluating the credibility of web-based medical information: guidelines for creating training data sets for machine learning. JMIR Med Inform 2021;9(11):e26065 [FREE Full text] [doi: 10.2196/26065] [Medline: 34842547]
49. Jawahar G, Sagot B, Seddah D. What does BERT learn about the structure of language? : HAL open scince; 2019 Presented at: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics; 2019 July 28 to August 2; Florence (Italy) p. 3651-3657. [doi: 10.18653/v1/p19-1356]

Abbreviations

BERT: bidirectional encoder representations from transformers
MeSH: medical subject heading
RF: random forest
SVM: support vector machines

Edited by T Mackey; submitted 27.01.24; peer-reviewed by P Deka, CN Hang, O Ismaila; comments to author 14.03.24; revised version received 08.05.24; accepted 24.12.24; published 21.02.25.

Please cite as:
Bayani A, Ayotte A, Nikiema JN
Transformer-Based Tool for Automated Fact-Checking of Online Health Information: Development Study
JMIR Infodemiology 2025;5:e56831
URL: https://infodemiology.jmir.org/2025/1/e56831
doi:10.2196/56831
PMID:

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org