

Original Paper

# Intersection of the Web-Based Vaping Narrative With COVID-19: Topic Modeling Study

Kamila Janmohamed<sup>1</sup>; Abdul-Nasah Soale<sup>2</sup>, MSc; Laura Forastiere<sup>1,3</sup>, PhD; Weiming Tang<sup>4,5,6</sup>, PhD; Yongjie Sha<sup>4</sup>, BA; Jakob Demant<sup>7</sup>, PhD; Edoardo Airoidi<sup>2</sup>, PhD; Navin Kumar<sup>1</sup>, MPhil

<sup>1</sup>Human Nature Lab, Department of Sociology, Yale University, New Haven, CT, United States

<sup>2</sup>Department of Statistical Science, Fox School of Business, Temple University, Philadelphia, PA, United States

<sup>3</sup>Department of Biostatistics, Yale School of Public Health, New Haven, CT, United States

<sup>4</sup>UNC Project-China, University of North Carolina at Chapel Hill, Guangzhou, China

<sup>5</sup>Southern Medical University Dermatology Hospital, Guangzhou, China

<sup>6</sup>School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

<sup>7</sup>University of Copenhagen, Copenhagen, Denmark

**Corresponding Author:**

Navin Kumar, MPhil

Human Nature Lab

Department of Sociology

Yale University

17 Hillhouse Ave

New Haven, CT, 06511

United States

Phone: 1 4754148375

Email: [navin.kumar@yale.edu](mailto:navin.kumar@yale.edu)

## Abstract

**Background:** The COVID-19 outbreak was designated a global pandemic on March 11, 2020. The relationship between vaping and contracting COVID-19 is unclear, and information on the internet is conflicting. There is some scientific evidence that vaping cannabidiol (CBD), an active ingredient in cannabis that is obtained from the hemp plant, or other substances is associated with more severe manifestations of COVID-19. However, there is also inaccurate information that vaping can aid COVID-19 treatment, as well as expert opinion that CBD, possibly administered through vaping, can mitigate COVID-19 symptoms. Thus, it is necessary to study the spread of inaccurate information to better understand how to promote scientific knowledge and curb inaccurate information, which is critical to the health of vapers. Inaccurate information about vaping and COVID-19 may affect COVID-19 treatment outcomes.

**Objective:** Using structural topic modeling, we aimed to map temporal trends in the web-based vaping narrative (a large data set comprising web-based vaping chatter from several sources) to indicate how the narrative changed from before to during the COVID-19 pandemic.

**Methods:** We obtained data using a textual query that scanned a data pool of approximately 200,000 different domains (4,027,172 documents and 361,100,284 words) such as public internet forums, blogs, and social media, from August 1, 2019, to April 21, 2020. We then used structural topic modeling to understand changes in word prevalence and semantic structures within topics around vaping before and after December 31, 2019, when COVID-19 was reported to the World Health Organization.

**Results:** Broadly, the web-based vaping narrative can be organized into the following groups or archetypes: harms from vaping; Vaping Regulation; Vaping as Harm Reduction or Treatment; and Vaping Lifestyle. Three archetypes were observed prior to the emergence of COVID-19; however, four archetypes were identified post-COVID-19 (Vaping as Harm Reduction or Treatment was the additional archetype). A topic related to CBD product preference emerged after COVID-19 was first reported, which may be related to the use of CBD by vapers as a COVID-19 treatment.

**Conclusions:** Our main finding is the emergence of a vape-administered CBD treatment narrative around COVID-19 when comparing the web-based vaping narratives before and during the COVID-19 pandemic. These results are key to understanding how vapers respond to inaccurate information about COVID-19, optimizing treatment of vapers who contract COVID-19, and possibly minimizing instances of inaccurate information. The findings have implications for the management of COVID-19

among vapers and the monitoring of web-based content pertinent to tobacco to develop targeted interventions to manage COVID-19 among vapers.

(*J Med Internet Res* 2020;22(10):e21743) doi: [10.2196/21743](https://doi.org/10.2196/21743)

## KEYWORDS

vaping; COVID-19; topic modeling; web-based narrative; modeling; trend; narrative; social media; internet; web-based health information

## Introduction

COVID-19 is spreading rapidly and has been declared a global pandemic [1]. COVID-19 was first reported in Wuhan, China, in December 2019 [2], and it was declared a pandemic by the World Health Organization (WHO) on March 11, 2020 [3]. With the pandemic currently in progress, research on the determinants of disease progression and communities that may be more vulnerable to COVID-19 is of key importance [4].

There is some scientific knowledge that vaping cannabidiol (CBD) or other substances may be associated with more severe manifestations of COVID-19 [5,6]. Use of electronic cigarettes (e-cigarettes) has been associated with a reduction in the ability of the lungs to respond to infection [5,6]; thus, people who use e-cigarettes may be at increased risk of contracting COVID-19 [4,7]. Several studies have indicated that smokers, including vapers, are more vulnerable to COVID-19 infections or more likely to develop serious complications after contracting SARS-CoV-2, the virus that causes COVID-19 [8-10]. Vaping devices may also be possible sites of COVID-19 transmission [9]. However, inaccurate information that vaping can aid COVID-19 treatment is also circulating [9], and experts have expressed the opinion that CBD, which can be administered through vaping, can mitigate COVID-19 symptoms [11,12].

For example, some Twitter posts around vaping indicated that e-cigarette devices may increase lung humidity and prevent COVID-19, and other posts stated that these devices can be used to administer COVID-19 medication to the lungs and possibly destroy the virus [9]. It has also been suggested that CBD products, often delivered through vaping, can be used to treat COVID-19, perhaps by augmenting the immune system [11,12]. Much of the information around vaping and other tobacco products is disseminated through the internet [13,14] and can affect health outcomes [15,16]. For example, vapers who develop COVID-19 may mistakenly believe that vaping CBD or other substances can alleviate COVID-19 symptoms; however, it may instead create additional disease complications. Thus, it is necessary to study the spread of inaccurate information to better understand how to promote scientific knowledge and curb misinformation, which may be critical to vapers' health [17].

A previous study analyzed Twitter content around COVID-19 and vaping [9]. The indicated study surveilled vaping tweets and detailed conversations around COVID-19 and vaping. The web-based conversations centered on the possibly heightened risk of COVID-19 for vapers and how vaping could potentially protect against COVID-19 [9]. However, past work did not use web-based vaping-related data from a range of sources or

detailed data from before the COVID-19 pandemic and through its progression. Detailing a large scope of sources is necessary to document the broad range of web-based vaping conversations, and collecting data from both before and during the COVID-19 pandemic is key to understanding the narrative prior to the emergence of COVID-19 and how it is changing as the pandemic progresses. Our study builds on previous work by using a large data set to represent the web-based vaping narrative (August 1, 2019, to April 21, 2020) that combines analysis of a multitude of sources, such as blogs, forums, and social media posts; also, we used novel computational techniques to examine how the vaping narrative has changed from before to during the COVID-19 pandemic.

Using the novel computational technique of structural topic modeling (STM), we mapped temporal trends in the web-based vaping narrative (a large data set comprising web-based vaping chatter from several sources) to show how discourse differed before versus during the COVID-19 pandemic. Topic modeling is a computer-aided content analysis technique where texts are organized into themes known as topics. These topics are not provided to the machine prior to modeling but emerge inductively as the algorithm learns patterns within the texts. The model assumes a relational theory of meaning by identifying structures of co-occurrence of words in individual texts and across all the texts. The model thus provides content analysis of text data sets that are too large to code by hand. Topic models use machine learning to uncover patterns and relationships that may be omitted by hand coding or traditional content analysis. Unsupervised machine learning methods have performed similarly to human coders on identical documents [18]. Unsupervised machine learning is a variant of machine learning that looks for new patterns in a data set without pre-existing labels and with limited human supervision. We used an approach to topic modeling known as STM [19,20]. STM enables discovery of topics and their prevalence based on document metadata, such as dates, or other important attributes, such as the number of new COVID-19 cases worldwide per day. Adding this metadata is useful, as the data are obtained over several months (August 2019 to April 2020), and the web-based vaping narrative may be susceptible to thematic change based on the progression of the COVID-19 pandemic. STM has been used to address several social scientific research questions around areas such as climate change [21,22] and web-based drug marketplaces [23]. As vapers may be at greater risk for contracting SARS-CoV-2 and COVID-19 disease progression [8-10], and CBD, which is often administered through vaping, may have interactions with COVID-19 treatment outcomes, we hope to provide insight on how vapers are responding to the pandemic. This may help improve the treatment outcomes of vapers who develop COVID-19.

## Methods

### Ethics Statement

Approval and informed consent were not needed because the data were collected using publicly available textual query techniques. All data are publicly available and can be accessed by anyone. The data were provided to the research team with all identifiers removed.

### Data Acquisition and Processing

Data were obtained using a textual query that scanned a data pool of approximately 200,000 different domains, such as public forum posts, blogs, news articles, message boards, health care provider forums, and social media (see [Multimedia Appendix 1](#) for the full list of sources). Textual queries were used to automatically search the indicated sources for text fragments related to keywords such as *vape*, *vaping*, and *e-cigarette*. The data that comprised vaping-related text fragments were collected from August 1, 2019, to April 21, 2020. As the data set represents a multitude of sources for web-based chatter related to vaping, our data set is likely representative of the web-based vaping narrative during the indicated period. The start date for the COVID-19 pandemic was denoted as December 31, 2019, when the Chinese government disclosed the existence of COVID-19 to the World Health Organization (WHO) [3]. Although the date of the first COVID-19 case is prior to December 31, 2019 [2], COVID-19 is unlikely to have influenced vaping-related discourse in the United States prior to December 31, 2019, due to low global awareness. The time period of August 1, 2019, to April 21, 2020, was chosen to provide sufficient data to detail the vaping narrative prior to the COVID-19 pandemic. Given that the date demarcating pre- and post-COVID-19 is December 31, 2019, our time period allowed for approximately four months before and approximately the same period after the first report of COVID-19 to the WHO.

### Word Prevalence and Topic Modeling

To prepare the data for word prevalence and topic modeling analysis, English stop words such as “the,” “a,” and “an” were removed, along with abbreviations, and terms were stemmed using Porter’s stemming algorithm [24]. Stemming converts words with the same “stem” or root (eg, “innovative” and “innovator”) to a single word type (eg, “innovate”). As our study centered on the intersection of vaping and COVID-19, it was expected that words such as “cigarette,” “vape,” and “coronavirus” would dominate our findings. However, these terms may crowd out other words, perhaps causing us to miss key topics occurring in the text. For example, if we were interested in understanding different cooking techniques such as roasting and frying, and we sourced data from web-based forums frequented by amateur chefs, the most frequent words in the data might be “cook” and “recipe.” However, these words might obscure information around the cooking techniques we were interested in. Thus, in some cases, such as our study, it may be necessary to remove frequently occurring words to detail underlying themes in the data. All data were first processed to remove mentions of COVID-19, tobacco, and vaping. These data were used to generate word clouds by word prevalence. As we will later detail, the word clouds generated by word

prevalence contained significant mentions of CBD after the emergence of COVID-19. When conducting topic modeling, mentions of CBD may crowd out other words and reduce our ability to identify salient topics. As such, we further processed the data set for topic modeling by removing mentions of cannabis, inclusive of CBD.

We first generated word clouds based on the top 200 terms ranked by prevalence before and after COVID-19 was reported to the WHO. In a word cloud, a larger font size indicates a greater prevalence of a single word. Word clouds thus provide a relative yardstick of the importance of a word in a particular time period. This visualization enabled us to qualitatively assess words by importance. Documents were processed (words removed) for mentions of cannabis, COVID-19, tobacco, and vaping. References to cannabis were determined using these search terms: [*bud OR cannabis OR cannabidiol OR cbd OR ganja OR hash OR hashish OR hemp OR indica OR joint OR marijuana OR mary jane OR ruderalis OR pot OR sativa OR weed OR THC*]. References to COVID-19 were determined using these search terms: [*COVID-19 OR covid 19 OR novel coronavirus OR coronavirus OR sars cov-2 OR sars cov 2 OR sars-cov-2 OR n-cov OR cov OR covid*]. References to tobacco were determined using these search terms: [*baccy OR bidi OR cig OR cigar OR cigarillo OR cigarette OR ciggy OR fag OR hookah OR pipe OR shag OR sheesha OR shisha OR snuff OR snus OR tobacco*]. References to vaping were determined using these search terms: [*e-cig OR electronic cigarette OR vape OR vaper OR vaping OR vapelife OR vapist OR vapin OR vaplyfe*].

We then used topic modeling to understand changes in word prevalence within topics around vaping and COVID-19. Topic modeling is a computer-aided content analysis technique in which texts are organized into themes known as topics [25,26]. In topic modeling, a topic is a distribution over a vocabulary of words that represent semantically interpretable themes [19]. For example, in a topic denoted “vape,” the terms “smoke” and “device” are more likely to occur than the words “peanut” and “tomato.” “Smoke” may appear in both “vape” and “cooking” topics with different contextual meanings. Given that the topic is a distribution, “smoke” may appear with other high-probability terms such as “roast” and “fry” in the “cooking” topic but may appear with terms such as “nicotine” and “device” in the “vape” topic. Thus, topics can be understood by considering that a person who was talking about the topic of “cooking” would tend to use some words more frequently than others compared to if they were talking about the topic of “vape.” Topic models are suitable for analyzing large quantities of textual data via an automated technique for providing context.

We used an approach to topic modeling known as STM [18,20]. STM [18,20] enables the generation of topics regarding document metadata such as date and source as well as other covariates relevant to the research question, such as new COVID-19 cases. This is vital to understanding how the narrative and topic proportions change over time. This enabled a robust quantitative analysis of how the COVID-19 pandemic has shaped the web-based narrative on vaping [19]. The key innovation of STM is that it can incorporate metadata or information about each document. This allows metadata covariates, such as new COVID-19 cases per day, to influence



topic discovery. Metadata can affect both the prevalence and content of a topic. Metadata covariates for topical prevalence allow the metadata to affect topic frequency. Similarly, covariates in topical content allow the metadata to affect the word rate within a topic or how a topic is discussed [20]. The STM process will output documents and vocabulary for analysis [20]. This output can be investigated in a range of ways, such as detailing words associated with topics or the relationships between metadata and topics. Model output can be used to conduct hypothesis testing around these relationships. STM [18,20] was applied to the whole data set (August 1, 2019, to April 21, 2020); the data prior to the reporting of COVID-19 to the WHO only (August 1 to December 31, 2019); and the data after COVID-19 was reported to the WHO only (January 1 to April 21, 2020). We subsetted the data to see if the vaping narratives were different before and after COVID-19 was reported to the WHO. We used the following metadata covariates for the STM models. For the full data set, the covariates were the binary variable for before and after COVID-19 was reported to the WHO, COVID-19 content covariate (variable to control for COVID-19-related content), date (the first day was denoted as 1, and the days were numbered sequentially after), source (0=social media, 1=news), new COVID-19 cases per day worldwide, and new COVID-19 deaths per day worldwide. For the pre-COVID-19 data set, the covariates were the date and source. For the post-COVID-19 data set, the covariates were the date, source, new COVID-19 cases per day worldwide, and new COVID-19 deaths per day worldwide.

Because STM is an unsupervised approach, the number of topics to estimate ( $k$ ) is key to the analysis. We first estimated several models ranging from 5 to 200 topics. These models were then evaluated qualitatively by their ability to produce coherent topics and capture topics regarding vaping and COVID-19 [27]. The number of topics was based on our understanding of the data set and how other researchers interpreted STM results [27,28]. The choice of the number of topics was also influenced by postestimation validation outcomes and past work [27]. As per standard content analysis [29], topic model validation also requires qualitative review, where researchers assess the interpretability and relative efficacy of models based on their subject matter expertise and data context. Our final models ( $k=15$  for the whole data set;  $k=20$  for the pre-COVID-19 report data set;  $k=20$  for the post-COVID-19 report data set) provided the greatest external validity and the most semantically coherent output of distinctive topics. When the number of topics was greater than indicated above, there were diminishing returns for solutions, as the substantive meaning and coherence of categories started to break down [21]. When the number of topics was lower, variation decreased and specific topics were placed into more generic categories. Validating a topic model is not the same as evaluating a statistical model regarding a population sample [30]. The goal is to identify the framework that best describes the data, not to estimate population parameters [30].

We conducted qualitative analysis to determine the number of topics based on past social science studies in which topic modeling was used to extract meaning from large text samples.

These studies [21,31] determined topics by qualitative coding based on word prevalence and researchers' topic expertise. We applied similar techniques in this study. Methods such as interrater reliability ratings may guard against subjective bias based on subject matter expertise and data context [32]. Adding interrater reliability to the qualitative component of topic modeling may improve data quality. However, we sought to use topic modeling to answer a specific research question, not to improve on methodological techniques. Thus, we used best practices implemented in previous studies regarding topic interpretation but did not advance these methods.

Topic interpretation was influenced by prior knowledge about why texts were written and what they sought to accomplish. Most of the text was produced and consumed by people who engaged in vaping and other forms of tobacco use, and this lens was used to interpret the presence or absence of topics and words. Most of the topic labels were straightforward and did not require much interpretation. To characterize topics in the vaping narrative, we qualitatively coded each topic by investigating word clouds based on each topic and reviewing exemplar documents which detailed high proportions of each topic [19]. The topic we classified as "tobacco company merger called off" had the following most frequently occurring words: "sue," "analyst," "halt," "1st," "imperial," "judge," "backlash," "advisor," "merger," "stake," "acquire," "outbreak," "carolina," "confirm," and "mint." Exemplar documents that exhibited high proportions of this topic indicated a preoccupation with these words. This detailed a topical preoccupation with a tobacco firm merger being called off. Thus, the interpretation of the topic was clear, given the genre of the narrative and the reliance on research regarding prominent topics around vaping. Two authors assigned the topics, and a third author resolved disagreements when they arose.

Topic validation is key to assessing whether the substantive meaning of the topic and its words are parallel with the qualitative meaning of the text, and we used methodological guidance from past research for this purpose [19,26]. Past work advocated the use of sample documents to validate the substantive meaning of each topic. Determining the number of sample documents to use is based on the amount of resolution needed by a social scientist to answer the research question using topic modeling methods [33]. Thus, determining the number of sample documents is a largely qualitative process that is dependent on the research question at hand. To determine the appropriate number of documents to sample, we searched the social science literature for studies that used topic modeling based on whether they asked similar research questions to our study, addressed similar topic areas to our study, and drew their study data from similar sources to those used in our study. We searched databases such as Web of Science Core Collection, Embase, PsycINFO, MEDLINE, and Sociological Abstracts. We used keywords such as *contrarian*, *polarized*, and *topic modeling*. The 2016 paper by Farrell [21] was determined to be the most similar to our study based on the assessed characteristics. Farrell explored ideological polarization in climate change and used a broad range of sources, such as press releases, published papers, and website articles. Based on the nature of the research question and large range of sources,

Farrell determined that a sample of 50 documents was sufficient to validate the substantive meaning of the topic output. Given the similarities between Farrell's 2016 study and ours in a range of characteristics, we similarly determined that a sample of 50 documents was adequate to validate the topics. We used `findThoughts` and `plotQuote` within the STM package to examine the top 50 associated documents for each topic to validate a topic's substantive meaning. Determination of the top 50 documents was based on ranking topics by the maximum a posteriori estimate of the topic's theta value, which represents the modal estimate of the proportion of word tokens assigned to the topic with the model. These top 50 documents were read by two of the authors to determine validity. A third author resolved disagreements where necessary. As indicated above, interrater reliability was not determined; however, we believe our methods were sufficiently robust.

Finally, word clouds were again generated, this time to visualize topics with the top 100 words ranked by STM-generated weights per topic, for the models representing the full data set and the data sets from before and after COVID-19 was reported to the WHO. In these word clouds, a larger font size indicated a greater weight, with word clouds indicating the importance of a word within a topic. We then grouped topics and their associated word clouds into larger categories or archetypes based on shared concepts across topics [34]. All analysis was conducted using R with the following packages: `dplyr`, `quanteda`, `textclean`, `tm`, and `stm` [20,24,35-37].

## Results

### Data

We collected 4,027,172 documents (361,100,284 words) comprised of text from blogs (86.01%, N=310,546,244), news articles (11.02%, N=39,721,031), forums (3.01%, N=10,833,008), comments (<1%), professional reviews and Facebook posts (both <1%).

### Word Prevalence

The most frequently observed words in the data set were "ban" (216,735/361,100,284 words, 0.06%, rank 1), "product" (135,607/361,100,284 words, 0.04%, rank 2), and "make" (115,413/361,100,284 words, 0.03%, rank 3).

Figure 1 shows a word cloud displaying the 200 most frequently featured words in the data sets divided by time period. The words are colored, sized, and positioned radially based on frequency of appearance, with larger, more central words appearing most frequently. In the period before COVID-19 was reported to the WHO (prior to December 31, 2019), commonly featured words included "ban," "lung," and "quit." Over the next four months, mentions of "CBD" and "oil" increased, along with positive words such as "good" and "best." There was a clear shift in word prevalence before COVID-19 was reported to the WHO compared to after the COVID-19 report. Word prevalence shifted from words related to the vaping ban to positive words and words associated with CBD. The vaping ban was a move by the US government on September 11, 2019, to remove all flavored vaping products from the market [38].





**Figure 2.** Major archetypes of vaping-related topics with example topics per archetype obtained by structural topic modeling for all observations (August 1, 2019, to April 21, 2020). The word clouds are generated from the weights of the top 100 terms within a topic. Terms with larger weights are depicted in larger font sizes. Terms with approximately the same weight are depicted in the same color.



- 1a. Toxic chemicals in vape products
- 2a. Vaping health crisis
- 3a. Vitamin E acetate responsible for vaping illnesses
- 4a. Vaping-related chest pains
- 5a. Minimum age for vaping raised
- 6a. Walmart ending e-cigarette sales
- 7a. Flavored vape ban reversed
- 8a. Disposable vape devices not covered by flavor ban
- 9a. Vaping lung damage
- 10a. Lung injury outbreak linked to vaping





**Figure 4.** Major archetypes of vaping-related topics with example topics per archetype obtained by structural topic modeling after COVID-19 was reported to the WHO (January 1 to April 21, 2020). The word clouds are generated from the weights of the top 100 terms within a topic. Terms with larger weights are depicted in larger font sizes. Terms with approximately the same weight are depicted in the same color.



Topic modeling captured significant events in the vaping environment, such as the vaping health crisis and Walmart stopping e-cigarette sales. The vaping health crisis referred to the 2019 emergence of vaping-associated pulmonary injury (VAPI) in the United States [39]. Walmart stopping e-cigarette sales denoted the September 2019 termination of vape sales at Walmart after the advent of VAPI [40]. These were likely significant events in the vaping narrative, as they emphasized the possible harms of vaping and were highly prominent in US media. Accordingly, these events were identified as individual topics. We organized the topics into the following groups or archetypes: Harms from Vaping; Vaping Regulation; Vaping as Harm Reduction or Treatment; and Vaping Lifestyle. We

generated word clouds from the weights of the top 100 terms within each topic, and Figures 2–4 show sample word clouds for each of the three archetypes. Broadly, across the whole time period (Figure 2), the web-based vaping narrative was centered around harms from vaping and vaping regulation. Archetypes pre-COVID-19 (Figure 3) versus during COVID-19 (Figure 4) were largely similar, except that three archetypes were identified pre-COVID-19 (Harms From Vaping; Vaping Regulation; Vaping Lifestyle) but four archetypes were identified post-COVID-19 (Harms From Vaping; Vaping Regulation; Vaping as Harm Reduction or Treatment; Vaping Lifestyle). This suggests that the emergence of COVID-19 is related to the appearance of topics around vaping as a form of harm reduction

or treatment. There was also variation in the topics within an archetype. For example, pre-COVID-19, the Harms From Vaping archetype included topics such as “illness from black market vape products” and “toxic chemicals in cannabis vape products.” After COVID-19 was reported to the WHO, example topics in the same archetype were “vaping and dental health” and “vaping and lung damage.” This suggests that while underlying themes in the vaping narrative are largely stagnant, specific topics in an archetype may vary over time. In line with the difference in archetypes before and during the COVID-19 pandemic, there were also different event-based topics before and during COVID-19. Before COVID-19, several topics represented events significant to vapers in that time period, such as “Walmart stops e-cigarette sales” and “Thanksgiving vape device sales.” These event-related topics were not present after COVID-19 was reported to the WHO. Instead, we noticed new event-based topics, such as “Rush Limbaugh supports vaping” and “disposable vape devices not covered by ban.” After COVID-19 was reported to the WHO, we noticed the emergence of topics specific to COVID-19 that were not present pre-COVID-19. These topics were “vape stores as essential service” (vape stores not being designated as essential services when COVID-19 lockdowns occurred in the United States) and “sharing vapes in COVID-19” (vape devices being possible sites of SARS-CoV-2 transmission). Similarly, a “CBD product preference” topic emerged after the COVID-19 report (Figure 4, word cloud 8c) under the Vaping as Harm Reduction or Treatment archetype that may be related to the advent of the pandemic. As indicated in the Methods section, we removed words around cannabis to provide for more nuanced analysis; accordingly, CBD and other related terms did not appear in the topic-based word clouds. This CBD topic was not present pre-COVID-19. The appearance of the “CBD product preference” topic may be related to vapers using CBD as a treatment for COVID-19. As a validity check, we examined the top 50 associated documents for the “CBD product preference” topic to validate the substantive meaning of the topic. Convenience sampling was not used to sample the top 50 topics; instead, we used the theta values of the topics, as detailed earlier. The number of top-ranked documents to be sampled was based on the methodology outlined earlier. There were 114,622 documents in total for the “CBD product preference” topic. These documents were read by two of the authors to determine validity. A third author resolved disagreements where necessary. As indicated above, interrater reliability was not used. We found that a majority of these top 50 documents (31, 62%) detailed CBD administered through vaping as a possible COVID-19 cure or protective agent. Example text fragments regarding how vaping CBD can prevent or treat COVID-19:

*Pot smoke is the best expectorant I've ever used and fresh cbd oil or weed brownies are verrrrrry healing, promotes good sleep and good healing. If you are a non-smoker, tobacco smoke should help you clear out your lungs if you have nothing else.* [March 10, 2020]

*COVID-19 deaths invariably involve a "cytokine storm," an excessive, un-checked immune system response. Cannabinoids from cannabis, cbd in*

*particular, can lower cytokine production naturally. research needed asap!* [March 25, 2020]

Example text fragments regarding websites marketing CBD for COVID-19 prevention or treatment:

*Why hemp cbd flowers and a vaporizer are the best COVID-19 coronavirus prepping tools.* [February 18, 2020]

*REDACTED COMPANY NAME applauds the use of cbd during the coronavirus outbreak* [March 30, 2020]

*Toronto-based cannabis seller testing cbd's effectiveness on reducing symptoms of coronavirus.* [April 1, 2020]

## Discussion

### Principal Findings

Our main finding was the emergence of discourse around vape-administered CBD treatment for COVID-19 when comparing web-based vaping narratives before and after the outbreak of COVID-19. Recent work suggested that CBD use may increase COVID-19 risks [41]. Other studies indicated that CBD may aid COVID-19 treatment outcomes [12,42]. Vaping CBD products as a treatment for COVID-19 is still largely unsubstantiated. Beliefs around CBD as a COVID-19 treatment, bolstered by marketing campaigns and early-stage research [11,12,42], may be responsible for the emergence of discussion around CBD.

There is limited empirical research on the intersection of the vaping narrative and COVID-19, especially around the emergence of CBD-related COVID-19 treatments and comparing the narratives before and during the COVID-19 pandemic. The role of web-based narratives in tobacco control, especially within social media, is a growing field of study [43]. We expand on past work that used novel computational approaches to examine trends in digital media to understand how these web-based behaviors may influence health behaviors [44,45]; our results indicate that the web-based environment is key to comprehending vaping and related health outcomes, especially in response to public health events. Previous research suggested the need to monitor social media content around tobacco to protect youth and mitigate tobacco use [43]. We expand on these studies, bolstering the need to surveil web-based tobacco content given our findings around increased discussion of inaccurate COVID-19 vape-administered treatments that are not evidence-based and may worsen health outcomes. A recent review detailed the role of misinformation in public health outcomes [17], and we expand on past work by providing evidence of how large-scale events may create misinformation in the health sphere. The strength of this study is our use of innovative computational methods to explore the content of the vaping narrative and how it is affected by COVID-19, comparing narrative content before the COVID-19 outbreak and after the pandemic took shape. This outcome measurement is key to understanding how vapers respond to COVID-19, enabling optimized treatment of vapers who develop COVID-19, and possibly minimizing instances of inaccurate information.

Our findings around the use of CBD as a non-evidence-based and possibly injurious COVID-19 treatment, likely administered through vaping, are indicative of the earlier discussed point. Inaccurate information on the internet may create complications in COVID-19 treatment. Vapers who develop COVID-19 may use vape-administered CBD treatments; meanwhile, CBD is associated with reduced immune system functioning [41] and may heighten the disease progression of COVID-19. It is possible that upon contracting COVID-19, people may use vape devices to administer medication to themselves. Given the possibility of device malfunction [46,47], some individuals may further harm their health if they develop COVID-19 symptoms. As levels of misinformation around devices as a means to administer purported COVID-19 treatments increases, more people may share these modified devices; this creates possible sites of transmission [9,48] and may further increase SARS-CoV-2 transmission rates.

If health care professionals are aware that vapers with COVID-19 may use CBD as a treatment based on inaccurate information, these professionals may be better able to respond to vapers with COVID-19 who demonstrate CBD-related complications. Professionals can provide accurate information regarding COVID-19 to vapers who seek health care. Misinformation can also be combated with trusted information. Public health authorities can include COVID-19-specific information in targeted vaping-related messaging, perhaps mitigating consumption of inaccurate web-based information. Information-based campaigns can target inaccurate information in line with the topics identified by our results, such as vaping CBD as a COVID-19 treatment. There are several experimental interventions around reducing levels of inaccurate information [49,50], with some centering specifically on COVID-19 [51] and other health issues [17]. Interventions that harness similar techniques, such as asking respondents to judge information accuracy, may nudge individuals toward obtaining accurate information around COVID-19 and vaping. Interventions can also center on vaping-related health literacy in various media outlets, which may reduce misinformation about the topic [52]. Thus, our results may improve COVID-19 treatment for individuals who may have received inaccurate information around COVID-19 and vaping; they may also provide insight on reducing levels of misinformation among vapers during the pandemic.

---

## Acknowledgments

This study was funded by The Foundation for a Smoke-Free World. The funding body had no role in the design, analysis, or interpretation of the data in the study. This study was preregistered on the Open Science Framework (DOI 10.17605/OSF.IO/XBVF2).

---

## Authors' Contributions

KJ, NK, AN, and EA contributed to the study design, hypothesis generation, data collection, data analysis, data interpretation, and manuscript writing and review. LF, WT, YS, DM, and JD contributed to the manuscript writing and review.

---

## Conflicts of Interest

None declared.

---

Our findings have several implications. From a policy standpoint, we suggest that vaping forums be mandated to provide accurate data around the interactions between vaping and COVID-19. These efforts may reduce levels of inaccurate information around COVID-19 and vaping and may minimize any COVID-19 complications associated with vaping. Future research can identify changes in the vaping narrative as the pandemic progresses further, allowing public health authorities to adjust treatment provision for vapers at risk of contracting COVID-19. Future work can also address how inaccurate information on the internet can be mitigated, especially as the pandemic progresses.

## Limitations

Our results depended on the validity of the data collected through our textual query. We searched a wide range of web-based media and identified key themes that validated our results (eg, Walmart stops e-cigarette sales, Vitamin E acetate and vaping illness), and we are thus confident in the comprehensiveness of our data. We may have overlooked some slang terms for vaping and thus underestimated the web-based narrative. We did not obtain location data for individual text fragments; thus, we were not able to determine how COVID-19 cases in certain areas affected the narrative. Our data were drawn from August 1, 2019, to April 21, 2020, and we were not able to determine changes in the narrative after April. We were not able to collect all web-based vaping chatter and may have missed some themes. In future research, we will collect qualitative and survey data from vapers to enhance the current findings. We did not use interrater reliability for our qualitative analysis, and we will use such methods in future research.

## Conclusions

We demonstrated the advent of discourse around vape-administered CBD treatment for COVID-19 by comparing the web-based vaping narratives before and during the COVID-19 pandemic. The increase in CBD-related discussion within the vaping narrative may be due to the marketing of CBD products consumed through vaping as a COVID-19 treatment [11]. Our findings have implications for the management of COVID-19 among vapers and for monitoring of web-based content pertinent to tobacco.



## Multimedia Appendix 1

List of sources from which data were obtained.

[\[XLS File \(Microsoft Excel File\), 1228 KB-Multimedia Appendix 1\]](#)

### References

1. Remuzzi A, Remuzzi G. COVID-19 and Italy: what next? *Lancet* 2020 Apr;395(10231):1225-1228. [doi: [10.1016/s0140-6736\(20\)30627-9](https://doi.org/10.1016/s0140-6736(20)30627-9)]
2. Wu Z, McGoogan JM. Characteristics of and Important Lessons From the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72 314 Cases From the Chinese Center for Disease Control and Prevention. *JAMA* 2020 Apr 07;323(13):1239-1242. [doi: [10.1001/jama.2020.2648](https://doi.org/10.1001/jama.2020.2648)] [Medline: [32091533](https://pubmed.ncbi.nlm.nih.gov/32091533/)]
3. Timeline of WHO's response to COVID-19. World Health Organization. URL: <https://www.who.int/news-room/detail/29-06-2020-covid-timeline> [accessed 2020-05-29]
4. Volkow ND. Collision of the COVID-19 and Addiction Epidemics. *Ann Intern Med* 2020 Jul 07;173(1):61-62. [doi: [10.7326/m20-1212](https://doi.org/10.7326/m20-1212)]
5. Madison MC, Landers CT, Gu BH, Chang CY, Tung HY, You R, et al. Electronic cigarettes disrupt lung lipid homeostasis and innate immunity independent of nicotine. *J Clin Invest* 2019 Oct 01;129(10):4290-4304 [FREE Full text] [doi: [10.1172/JCI128531](https://doi.org/10.1172/JCI128531)] [Medline: [31483291](https://pubmed.ncbi.nlm.nih.gov/31483291/)]
6. Gilpin DF, McGown K, Gallagher K, Bengoechea J, Dumigan A, Einarsson G, et al. Electronic cigarette vapour increases virulence and inflammatory potential of respiratory pathogens. *Respir Res* 2019 Dec 18;20(1):267 [FREE Full text] [doi: [10.1186/s12931-019-1206-8](https://doi.org/10.1186/s12931-019-1206-8)] [Medline: [31847850](https://pubmed.ncbi.nlm.nih.gov/31847850/)]
7. McAlinden KD, Eapen MS, Lu W, Chia C, Haug G, Sohal SS. COVID-19 and vaping: risk for increased susceptibility to SARS-CoV-2 infection? *Eur Respir J* 2020 Jul 19;56(1):2001645. [doi: [10.1183/13993003.01645-2020](https://doi.org/10.1183/13993003.01645-2020)] [Medline: [32430427](https://pubmed.ncbi.nlm.nih.gov/32430427/)]
8. Glantz SA. Reduce your risk of serious lung disease caused by corona virus by quitting smoking and vaping. Center for Tobacco Control Research and Education. 2020. URL: <https://tobacco.ucsf.edu/reduce-your-risk-serious-lung-disease-caused-corona-virus-quitting-smoking-and-vaping> [accessed 2020-04-08]
9. Majmudar A, Allem J, Cruz TB, Unger JB. Public Health Concerns and Unsubstantiated Claims at the Intersection of Vaping and COVID-19. *Nicotine Tob Res* 2020 Aug 24;22(9):1667-1668 [FREE Full text] [doi: [10.1093/ntr/ntaa064](https://doi.org/10.1093/ntr/ntaa064)] [Medline: [32285129](https://pubmed.ncbi.nlm.nih.gov/32285129/)]
10. Lewis T. Smoking or Vaping May Increase the Risk of Severe Coronavirus Infection. *Scientific American*. 2020 Mar 17. URL: <https://www.scientificamerican.com/article/smoking-or-vaping-may-increase-the-risk-of-a-severe-coronavirus-infection1/> [accessed 2020-10-02]
11. Fraudulent Coronavirus Disease 2019 (COVID-19) Products. US Food and Drug Administration. URL: <https://www.fda.gov/consumers/health-fraud-scams/fraudulent-coronavirus-disease-2019-covid-19-products> [accessed 2020-05-29]
12. Esposito G, Pesce M, Seguela L, Sanseverino W, Lu J, Corpetti C, et al. The potential of cannabidiol in the COVID-19 pandemic. *Br J Pharmacol* 2020 Jun 10;2020 [FREE Full text] [doi: [10.1111/bph.15157](https://doi.org/10.1111/bph.15157)] [Medline: [32519753](https://pubmed.ncbi.nlm.nih.gov/32519753/)]
13. Ayers JW, Ribisl KM, Brownstein JS. Tracking the rise in popularity of electronic nicotine delivery systems (electronic cigarettes) using search query surveillance. *Am J Prev Med* 2011 Apr;40(4):448-453. [doi: [10.1016/j.amepre.2010.12.007](https://doi.org/10.1016/j.amepre.2010.12.007)] [Medline: [21406279](https://pubmed.ncbi.nlm.nih.gov/21406279/)]
14. Yamin CK, Bitton A, Bates DW. E-cigarettes: a rapidly growing Internet phenomenon. *Ann Intern Med* 2010 Nov 02;153(9):607-609. [doi: [10.7326/0003-4819-153-9-201011020-00011](https://doi.org/10.7326/0003-4819-153-9-201011020-00011)] [Medline: [21041581](https://pubmed.ncbi.nlm.nih.gov/21041581/)]
15. Freeman B. New media and tobacco control. *Tob Control* 2012 Mar 16;21(2):139-144. [doi: [10.1136/tobaccocontrol-2011-050193](https://doi.org/10.1136/tobaccocontrol-2011-050193)] [Medline: [22345236](https://pubmed.ncbi.nlm.nih.gov/22345236/)]
16. Forsyth SR, Malone RE. "I'll be your cigarette--light me up and get on with it": examining smoking imagery on YouTube. *Nicotine Tob Res* 2010 Aug 15;12(8):810-816 [FREE Full text] [doi: [10.1093/ntr/ntq101](https://doi.org/10.1093/ntr/ntq101)] [Medline: [20634267](https://pubmed.ncbi.nlm.nih.gov/20634267/)]
17. Swire-Thompson B, Lazer D. Public Health and Online Misinformation: Challenges and Recommendations. *Annu Rev Public Health* 2020 Apr 02;41(1):433-451. [doi: [10.1146/annurev-publhealth-040119-094127](https://doi.org/10.1146/annurev-publhealth-040119-094127)] [Medline: [31874069](https://pubmed.ncbi.nlm.nih.gov/31874069/)]
18. Roberts ME, Stewart BM, Airoidi EM. A Model of Text for Experimentation in the Social Sciences. *Journal of the American Statistical Association* 2016 Oct 18;111(515):988-1003. [doi: [10.1080/01621459.2016.1141684](https://doi.org/10.1080/01621459.2016.1141684)]
19. Roberts ME, Stewart BM, Tingley D, Lucas C, Leder-Luis J, Gadarian SK, et al. Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science* 2014 Mar 06;58(4):1064-1082. [doi: [10.1111/ajps.12103](https://doi.org/10.1111/ajps.12103)]
20. Roberts ME, Stewart BM, Tingley D. stm: An R Package for Structural Topic Models. *J Stat Soft* 2019;91(2):1-40. [doi: [10.18637/jss.v091.i02](https://doi.org/10.18637/jss.v091.i02)]
21. Farrell J. Corporate funding and ideological polarization about climate change. *Proc Natl Acad Sci USA* 2016 Jan 05;113(1):92-97. [doi: [10.1073/pnas.1509433112](https://doi.org/10.1073/pnas.1509433112)] [Medline: [26598653](https://pubmed.ncbi.nlm.nih.gov/26598653/)]
22. Tvinnereim E, Fløttum K. Explaining topic prevalence in answers to open-ended survey questions about climate change. *Nature Clim Change* 2015 Jun 1;5(8):744-747. [doi: [10.1038/nclimate2663](https://doi.org/10.1038/nclimate2663)]

23. Guarino S, Santoro M. Multi-word structural topic modelling of ToR drug marketplaces. In: 2018 IEEE 12th International Conference on Semantic Computing (ICSC). 2018 Presented at: IEEE 12th International Conference on Semantic Computing (ICSC); Jan 31-Feb 2, 2018; Laguna Hills, CA p. 269-273. [doi: [10.1109/icsc.2018.00048](https://doi.org/10.1109/icsc.2018.00048)]
24. Feinerer I, Hornik K, Meyer D. Text Mining Infrastructure in R. *J Stat Soft* 2008;25(5):54. [doi: [10.18637/jss.v025.i05](https://doi.org/10.18637/jss.v025.i05)]
25. Mohr JW, Bogdanov P. Introduction—Topic models: What they are and why they matter. *Poetics* 2013 Dec;41(6):545-569. [doi: [10.1016/j.poetic.2013.10.001](https://doi.org/10.1016/j.poetic.2013.10.001)]
26. Blei DM. Probabilistic topic models. *Commun ACM* 2012 Apr;55(4):77-84. [doi: [10.1145/2133806.2133826](https://doi.org/10.1145/2133806.2133826)]
27. 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](https://doi.org/10.1093/pan/mps028)]
28. Wallach HM, Murray I, Salakhutdinov S, Mimno D. Evaluation methods for topic models. In: ICML '09: Proceedings of the 26th Annual International Conference on Machine Learning. 2009 Jun Presented at: 26th Annual International Conference on Machine Learning; Jun 14-18, 2009; Montréal, QC p. 1105-1112. [doi: [10.1145/1553374.1553515](https://doi.org/10.1145/1553374.1553515)]
29. Krippendorff K. *Content Analysis: An Introduction to its Methodology*. 4th Edition. Los Angeles, CA: Sage Publications; 2018.
30. DiMaggio P, Nag M, Blei D. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics* 2013 Dec;41(6):570-606. [doi: [10.1016/j.poetic.2013.08.004](https://doi.org/10.1016/j.poetic.2013.08.004)]
31. Munksgaard R, Demant J. Mixing politics and crime - The prevalence and decline of political discourse on the cryptomarket. *Int J Drug Policy* 2016 Sep;35:77-83. [doi: [10.1016/j.drugpo.2016.04.021](https://doi.org/10.1016/j.drugpo.2016.04.021)] [Medline: [27424175](https://pubmed.ncbi.nlm.nih.gov/27424175/)]
32. Saal FE, Downey RG, Lahey MA. Rating the ratings: Assessing the psychometric quality of rating data. *Psychol Bull* 1980;88(2):413-428. [doi: [10.1037/0033-2909.88.2.413](https://doi.org/10.1037/0033-2909.88.2.413)]
33. Nikolenko SI, Koltcov S, Koltsova O. Topic modelling for qualitative studies. *J Inf Sci* 2016 Jul 10;43(1):88-102. [doi: [10.1177/0165551515617393](https://doi.org/10.1177/0165551515617393)]
34. Evans MS. A computational approach to qualitative analysis in large textual datasets. *PLoS One* 2014 Feb 3;9(2):e87908 [FREE Full text] [doi: [10.1371/journal.pone.0087908](https://doi.org/10.1371/journal.pone.0087908)] [Medline: [24498398](https://pubmed.ncbi.nlm.nih.gov/24498398/)]
35. Wickham H, François R, Henry L, Müller K. *dplyr: A Grammar of Data Manipulation*. URL: <https://dplyr.tidyverse.org/reference/dplyr-package.html> [accessed 2020-10-02]
36. Benoit K, Watanabe K, Wang H, Nulty P, Obeng A, Müller S, et al. *quanteda: An R package for the quantitative analysis of textual data*. *JOSS* 2018 Oct;3(30):774. [doi: [10.21105/joss.00774](https://doi.org/10.21105/joss.00774)]
37. Rinker TW. *textclean: Text Cleaning Tools*. GitHub. URL: <https://github.com/trinker/textclean> [accessed 2020-10-02]
38. Furlow B. US Government contemplates a nationwide ban on flavoured e-cigarette products. *Lancet Respir Med* 2019 Nov;7(11):932. [doi: [10.1016/s2213-2600\(19\)30329-7](https://doi.org/10.1016/s2213-2600(19)30329-7)]
39. Hooper RW, Garfield JL. An Emerging Crisis: Vaping-Associated Pulmonary Injury. *Ann Intern Med* 2019 Oct 08;172(1):57. [doi: [10.7326/m19-2908](https://doi.org/10.7326/m19-2908)]
40. Lawrence K. *e-Cigarette Industry in Crisis: How Can Juul Survive?* SAGE Publications: SAGE Business Cases Originals 2020 Jan 15:online. [doi: [10.4135/9781529728613](https://doi.org/10.4135/9781529728613)]
41. Brown JD. Cannabidiol as prophylaxis for SARS-CoV-2 and COVID-19? Unfounded claims versus potential risks of medications during the pandemic. *Res Social Adm Pharm* 2020 Mar 31:online ahead of print [FREE Full text] [doi: [10.1016/j.sapharm.2020.03.020](https://doi.org/10.1016/j.sapharm.2020.03.020)] [Medline: [32273254](https://pubmed.ncbi.nlm.nih.gov/32273254/)]
42. Wang B, Kovalchuk A, Li D, Ilnytsky Y, Kovalchuk I, Kovalchuk O. In Search of Preventative Strategies: Novel Anti-Inflammatory High-CBD Cannabis Sativa Extracts Modulate ACE2 Expression in COVID-19 Gateway Tissues. Preprints. Preprint posted online on April 19, 2020 .
43. Kostygina G, Tran H, Shi Y, Kim Y, Emery S. 'Sweeter Than a Swisher': amount and themes of little cigar and cigarillo content on Twitter. *Tob Control* 2016 Oct 03;25(Suppl 1):i75-i82. [doi: [10.1136/tobaccocontrol-2016-053094](https://doi.org/10.1136/tobaccocontrol-2016-053094)] [Medline: [27697951](https://pubmed.ncbi.nlm.nih.gov/27697951/)]
44. Huang J, Kornfield R, Emery SL. 100 Million Views of Electronic Cigarette YouTube Videos and Counting: Quantification, Content Evaluation, and Engagement Levels of Videos. *J Med Internet Res* 2016 Mar 18;18(3):e67 [FREE Full text] [doi: [10.2196/jmir.4265](https://doi.org/10.2196/jmir.4265)] [Medline: [26993213](https://pubmed.ncbi.nlm.nih.gov/26993213/)]
45. Huang J, Kornfield R, Szczycka G, Emery SL. A cross-sectional examination of marketing of electronic cigarettes on Twitter. *Tob Control* 2014 Jul;23 Suppl 3:iii26-iii30. [doi: [10.1136/tobaccocontrol-2014-051551](https://doi.org/10.1136/tobaccocontrol-2014-051551)] [Medline: [24935894](https://pubmed.ncbi.nlm.nih.gov/24935894/)]
46. Chadi N, Belanger RE. Teen vaping: There is no vapour without fire. *Paediatr Child Health* 2020 Oct;25(6):337-339. [doi: [10.1093/pch/pxz137](https://doi.org/10.1093/pch/pxz137)] [Medline: [32968465](https://pubmed.ncbi.nlm.nih.gov/32968465/)]
47. Reasoner JJ, Regier BA, Beckendorf R, McAllister RK. Update on the Risks of Electronic Cigarettes-Vaping. *Ochsner J* 2020 Apr 03;20(1):2-4 [FREE Full text] [doi: [10.31486/toj.20.0012](https://doi.org/10.31486/toj.20.0012)] [Medline: [32284674](https://pubmed.ncbi.nlm.nih.gov/32284674/)]
48. Winickoff J, Hartman L, Reynolds M. Smoking and E-cigarette Use Put Individuals at Increased Risk of Coronavirus (COVID-19). Massachusetts Health Officers Association. URL: <https://mhoha.com/wp-content/uploads/2020/07/Smoking-and-E-cigarette-Use-Put-Individuals-at-Increased-Risk-of-COVID-19-1.pdf> [accessed 2020-08-15]
49. Pennycook G, Rand DG. Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proc Natl Acad Sci USA* 2019 Feb 12;116(7):2521-2526. [doi: [10.1073/pnas.1806781116](https://doi.org/10.1073/pnas.1806781116)] [Medline: [30692252](https://pubmed.ncbi.nlm.nih.gov/30692252/)]

50. Barnett PA, Hoskins CE, Alhoti JA, Carpenter LJ. Reducing Public Misinformation About Organ Donation: An Experimental Intervention. *Journal of Social Distress and the Homeless* 2013 Jul 19;18(1-2):57-73. [doi: [10.1179/105307809805365235](https://doi.org/10.1179/105307809805365235)]
51. Pennycook G, McPhetres J, Zhang Y, Lu JG, Rand DG. Fighting COVID-19 Misinformation on Social Media: Experimental Evidence for a Scalable Accuracy-Nudge Intervention. *Psychol Sci* 2020 Jul;31(7):770-780. [doi: [10.1177/0956797620939054](https://doi.org/10.1177/0956797620939054)] [Medline: [32603243](https://pubmed.ncbi.nlm.nih.gov/32603243/)]
52. Nordheim LV, Gundersen MW, Espehaug B, Guttersrud Ø, Flottorp S. Effects of School-Based Educational Interventions for Enhancing Adolescents Abilities in Critical Appraisal of Health Claims: A Systematic Review. *PLoS One* 2016 Aug 24;11(8):e0161485 [FREE Full text] [doi: [10.1371/journal.pone.0161485](https://doi.org/10.1371/journal.pone.0161485)] [Medline: [27557129](https://pubmed.ncbi.nlm.nih.gov/27557129/)]

## Abbreviations

**CBD:** cannabidiol

**e-cigarette:** electronic cigarette

**STM:** structural topic modeling

**VAPI:** vaping-assisted pulmonary injury

**WHO:** World Health Organization

*Edited by G Eysenbach; submitted 29.06.20; peer-reviewed by C Harshbarger, M El Tantawi; comments to author 15.07.20; revised version received 16.07.20; accepted 16.09.20; published 30.10.20*

*Please cite as:*

*Janmohamed K, Soale AN, Forastiere L, Tang W, Sha Y, Demant J, Airoidi E, Kumar N*

*Intersection of the Web-Based Vaping Narrative With COVID-19: Topic Modeling Study*

*J Med Internet Res* 2020;22(10):e21743

URL: <http://www.jmir.org/2020/10/e21743/>

doi: [10.2196/21743](https://doi.org/10.2196/21743)

PMID: [33001829](https://pubmed.ncbi.nlm.nih.gov/33001829/)

©Kamila Janmohamed, Abdul-Nasah Soale, Laura Forastiere, Weiming Tang, Yongjie Sha, Jakob Demant, Edoardo Airoidi, Navin Kumar. Originally published in the *Journal of Medical Internet Research* (<http://www.jmir.org>), 30.10.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the *Journal of Medical Internet Research*, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.