

## RESEARCH

# Vaping-related news events and their relationship with sentiment in the online vaping environment: A computational interrupted time series analysis with large-scale public data

Navin Kumar<sup>1\*</sup> and Kamila Janmohamed<sup>2</sup>

\*Correspondence:

navin.kumar@yale.edu

<sup>3</sup>210 Prospect St, New Haven, CT 06511

Full list of author information is available at the end of the article

## Abstract

**Background:** Vaping-related news coverage may have furthered misconceptions around the relative harms of vapes. Also, some positive opinions around vaping may be derived from misinformation, perhaps creating inimical health outcomes. Thus, we need to study how vaping-related news events (e.g. 2019 vaping illness epidemic, COVID-19) are associated with sentiment in the online vaping environment, to better understand how to promote vaping as a potential harm reduction technique for those who smoke and are unable to quit, and to minimize vape-centric misinformation that could lead to reduced health outcomes.

**Methods:** We obtained vaping-related online data through web-scraping several online environments from August 1 2019 - April 21 2020. Sentiment analysis was performed to understand changes in sentiment in the online vaping environment in relation to vaping-related events, such as the Trump administration's planned ban on flavored vaping products, and when COVID-19 was first reported to the WHO.

**Results:** For all online environments, we observed a statistically significant negative association of 15% (Estimate: -0.16; 95% CI: -0.29, -0.03; P: 0.01) between sentiment score and the Trump administration's move towards a ban on flavored vaping products, and a statistically significant positive association of 7% between sentiment score (Estimate: 0.07; 95% CI: 0.01, 0.14; P: 0.02) and when COVID-19 was first reported to the WHO (December 31 2019).

**Conclusions:** News events may be related to sentiment in the online vaping environment, depending on the event. Depending on the nature of the event, we suggest that public health messaging may improve health outcomes.

**Keywords:** News events; Vaping; E-cigarette; COVID-19; Harm reduction; Pandemic

## Background

E-cigarette use (vaping) is likely less injurious to health compared to combustible cigarettes, due to reduced production of toxic chemicals and carcinogens [1, 2]. Despite this evidence, many people who smoke in the US perceive e-cigarettes (vapes) to be at least as dangerous to health as combustible cigarettes [1, 3]. Such misconceptions may dissuade people who smoke and are unable to quit from switching to vaping [4]. If quitting smoking is not a viable option, switching to vaping may improve overall health outcomes [5, 6]. While youth vape use has declined since 2019,

its prevalence remains high. As of 2020, 4.5% of US adults and 3.6 million middle and high school students used e-cigarettes [7]. Sales from 2010-2016 show strong early growth followed by considerable slowing over time [8]. In the US, the current consensus is that vaping is not a smoking cessation method, as no vape has been approved by the Food and Drug Administration as a safe and effective cessation product. The US scientific consensus is that vape aerosol contains fewer numbers and lower levels of toxicants than smoke from combustible tobacco cigarettes [9]. However, use of vapes results in dependence on the devices, but with apparently less risk and severity than that of combustible tobacco cigarettes [9]. Among youths, vape use is associated with increased risk for cigarette initiation [10]. Among adults, a Cochrane review found that nicotine vapes probably do help people to stop smoking for at least six months, working better than nicotine replacement therapy and nicotine-free e-cigarettes [11].

News coverage around vaping-related events may have furthered misconceptions around the relative harms of vapes [4]. For example, regarding the recent outbreak of vaping-related lung injury (EVALI), most cases were related to consumption of vitamin E acetate, an additive included in some tetrahydrocannabinol (THC) devices [12]. However, news reports did not always differentiate between THC devices and standard nicotine-based vapes [4], perhaps disproportionately characterizing vaping harms. In addition, in response to EVALI, the Trump administration proposed to ban some nicotine-based vaping products [13], despite most EVALI cases being related to vapes that contained THC and vitamin E acetate [14]. This event was heavily featured in the news, with rising news coverage peaking in September 2019 [15].

More recently, there have been similar news articles and research involving COVID-19. Several studies have indicated that those who vape are more vulnerable to COVID-19 infections or more likely to develop serious complications once contracting severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as compared to people who do not vape [16, 17, 18]. Such information may dissuade people who smoke from switching to vaping, perhaps potentially increasing the overall tobacco mortality burden [19]. However, improved opinions around vaping may not always be beneficial to tobacco control. For example, during COVID-19, some individuals had positive sentiment about vaping, usually around vaping cannabidiol (CBD) as a possible COVID-19 cure, derived from incorrect beliefs that vaping may be a COVID-19 treatment [17]. Vaping CBD as COVID-19 treatment is still largely unsubstantiated, and likely a form of misinformation [20, 21]. Thus, we need to study how such news events are related to sentiment around vaping, to better understand how to promote vaping as a potential harm reduction technique for those who smoke and are unable to quit [22] and also ensure that positive sentiment around vaping is not borne from misinformation, possibly affecting health outcomes of those who vape.

A previous study detailed how exposure to vaping-centric news shaped individuals' normative understandings around real-life vape use [23]. This study used an experiment to show that news articles could influence individuals' perceptions of vaping prevalence [23]. Another study used data on consumer risk perceptions from two surveys conducted before and during EVALI to examine differences in risk

perceptions between these periods [24]. This study reported that the first EVALI news event around the Centers for Disease Control and Prevention (CDC) warning consumers to avoid all vaping products [25] increased the proportion of individuals who believed vapes to be harmful. As news around EVALI died down, the perceived risk of vaping decreased [24]. Past work has demonstrated that news sources can influence risk perceptions and normative perceptions around vaping, especially around perceived risks of these products. Risk perceptions are associated with health-related behaviors and thus perceptions around vaping may influence use [26].

Previous research has not explored how news events are associated with sentiment in vaping-related online data from a range of sources, nor explored how COVID-19 news related to vaping is associated with sentiment in online vaping environments. Detailing a large scope of sources is necessary to document the broad range of vaping conversations online and how they are related to news events. Exploring how COVID-19 news events are associated with sentiment around vaping is key to ensuring that positive sentiment around vaping is not driven by inaccurate information, inimical to health outcomes.

We mapped temporal trends (August 1 2019 - April 21 2020) in the online vaping arena using sentiment analysis to show how various news events were related to sentiment around vaping, in various online environments (e.g. social media, forums, news media). Sentiment analysis can identify if expressions in text detail positive or negative opinions. We aim to provide insight around improving health outcomes of people who smoke, amid increased risk perception around vaping [5, 6], and improved sentiment around vaping drawn from possibly inaccurate information.

## Methods

### Ethics statement

Approval and informed consent were not needed as all data was publicly available, based on practices in similar past research [27, 28].

### Data acquisition and processing

Data was obtained using a textual query (web-scraping) which scanned a data pool of approximately 200000 different US-based domains such as public forum posts, blogs, news articles, message boards, healthcare provider forums and social media (see Supplement for full list). The data collection process was conducted following guidelines established by Kim et. al (2016) [29]. We attempted to collect data from all publicly available sources that addressed vaping, both generalist and specialized sources. We thus obtained data from generalist sources such as Facebook posts, and specialized vaping forums.

Textual queries automatically search indicated sources for text fragments related to keywords, in this case, keywords such as vape, vaping, and e-cigarette (see Appendix for full list). The textual query extracted text fragments (e.g. sentences or paragraphs surrounding each keyword) instead of the full articles or posts. Keywords were drawn from vaping-specific keywords used in systematic and scoping reviews around vaping [30, 31]. To validate the accuracy of the textual query in retrieving data regarding vaping, we handcoded a randomly generated sample of 100

text fragments, maintaining a distribution similar to the distribution of sources. Two coders coded whether these text fragments regarded vaping and we achieved >90% retrieval accuracy. We did not code data for misinformation as our goal was to determine the sentiment of text, not whether it was misinformation or otherwise around vaping. The time period August 1 2019 - April 21 2020 was chosen as it included several key vaping events, especially those related to the US outbreak of e-cigarette product use-associated lung injury (EVALI) in late 2019. It also provided sufficient data to detail news events around COVID-19. As we used a broad range of sources, we likely captured both organic and commercial posts around vaping. The content of such posts are significantly different [32]. We did not account for the difference in content between these different sources as our goal was to capture as much of the online vaping arena as possible and see how sentiment was associated across various online environments, not to differentiate the types of online content. We then processed the data for analysis as follows: 1) duplicate entries were removed; 2) keeping informative text by filtering out entries not in the 10 - 8000 character range. Text shorter than 10 words did not normally contain useful information. Text longer than 8000 characters tended to be short stories that contained information about vaping; 3) using the key words in the Appendix we further subset the entries to ensure that retained content was related to vaping; 4) text in non-English languages, emojis, punctuation, room reviews from Tripadvisor.com were removed.

#### Outcomes of interest

We first assembled a preliminary list of key vaping events based on a review of online vaping forums, editorials, and peer-reviewed vaping research, and consulted experts on tobacco cessation, to result in a final list of seven events as below. Criteria for event selection were as follows: 1) Minimization of redundancy. If there were related events on September 1, 5, 19, we selected the September 1 event; 2) Relative importance. We dropped events that were discussed minimally compared to other events on vaping forums and other related media. Events were as follows: 1) CDC announcing an investigation into vaping-related illnesses on August 17 2019 following an outbreak in 14 states [33]; 2) Trump administration plan to ban some vaping products on September 11 2019 [13]; 3) Massachusetts's ban on vaping products announced September 24 2019 [34]; 4) FDA's warning to consumers against the use of vape products containing THC on October 4 2019 [35]; 5) CDC's announcement on the link between Vitamin E acetate and EVALI on November 8 2019 [36]; 6) Chinese government disclosure of COVID-19 to the World Health Organization (WHO) [37] on December 31 2019; 7) WHO declaring COVID-19 a pandemic on March 11 2020 [37]. As COVID-19 seems to affect perceptions around vaping, both as a risk factor and a form of treatment, it is possible that the announcement of major developments in the pandemic would be related to sentiment around vaping [17].

#### Sentiment analysis

Sentiment analysis was performed on the dataset. The unit of analysis was text fragments aggregated by day, regardless of source. Sentiment analysis is the computational study of people's opinions towards entities such as products, services, and

events [38]. Sentiment analysis can detect polarity (positive or negative sentiment) in text, in this case within the online vaping environment, but not specific emotions (happiness or sadness)[39]. Sentiment analysis can use a lexicon of positive and negative words and phrases to automatically, through an algorithm, classify expressions within the data [40]. For example, “Beautiful” and “I like vaping” have positive valences and “Ugly” and “Vaping is bad” have negative valences. In our sentiment analysis model output, a sentiment score from -1 to 1 was provided for each segment of text. Note that sentiment analysis does not code text fragments as pro- or anti-vaping, but whether text has a negative or positive valence. A score close to -1 indicates a highly negative sentiment, and a score close to 1 represents a very positive sentiment. A score around +/- 0.5 details moderately positive or negative sentiment. A neutral text fragment would have a score of 0 [41]. Sentiment analysis determines the valence of a text without regard to the context, i.e. it does not determine whether the sentiment of the text fragment is related to vaping or a news event. As our data concerned only vaping-related posts, sentiment analysis provided an overview of sentiment within vaping, despite the technique’s inability to determine whether text was pro- or anti-vaping. Analysis was conducted using R with the `sentimentr` package [42]. `sentimentr` incorporates valence shifters (e.g. negators, amplifiers), de-amplifiers (downtoners) [42]. For example, `sentimentr` can code the following as having a negative valence: I like it but it’s not worth it. To validate the sentiment analysis results, we handcoded a randomly generated sample of 100 text fragments, maintaining a distribution similar to the distribution of sources. Two coders coded the valence of these fragments and we achieved >80% similarity with the sentiment analysis results. Below are examples of text fragments coded as positive and negative by the algorithm.

Positive:

*THC is safe to vape people been doing it for ages.*

*The advantage of vaping that I can tell is it being odorless, meaning, you can’t smell the weed/cigarette smoke.*

Negative:

*This is no surprise. Anyone who has walked through a cloud of vape knows how bad it is.*

*I think e-cigs are worse than cigarettes and I’m not opposed to a ban.*

### Statistical analysis

We used an interrupted time series design with segmented regression analysis to determine immediate and longer-term impacts of individual news events on sentiment, controlling for pre-existing trends. The unit of analysis (dependent variable) was sentiment scores per day. Interrupted time series is the strongest quasi-experimental design to assess longitudinal effects of time-delimited treatments or interventions [43]. This design was appropriate as data was collected at multiple time points and we wanted to detect if a treatment (news events) had a significantly greater effect than another underlying trend [44]. We first conducted a visual examination on the pattern of the time series by plotting it and generating auto-correlation and partial correlation plots. No seasonal patterns were identified. Auto-correlation was tested with the Durbin-Watson test. Nonstationarity was identified using the augmented

Dickey-Fuller test and corrected through differencing. An autoregressive moving average (ARIMA) model was fit against a white noise series generated from the stationarized data to determine optimal model parameters. The model included binary variables for events (0=dates before the event, 1=dates after the event), time (1 was denoted for the first day and numbered sequentially after), and interaction terms between each event and time, as is standard for interrupted time series analysis using segmented regression [45]. We used Google Trends and US-based hospitalizations from vaping as control variables. These control variables may address underlying factors possibly influencing perceptions around vaping. By considering a broader picture of what may influence sentiment around vaping, we can better test the claims relation to the association between specific events and sentiment. We used internet search query trends as these may reveal what people are potentially thinking or doing based on the content and timing of their queries [46]. Regarding Google Trends, we tracked Google search interest (trends.google.com) originating from the US that mentioned terms regarding vaping (see Appendix for full list of terms). Searches were obtained from August 1 2019 - April 21 2020 to allow for historical trends to inform our sentiment analysis. Search interest represented search interest relative to the highest point in the given region and time period. A value of 100 represented peak popularity for the term. A value of 50 meant that the term was half as popular. A score of 0 meant there was not enough data for that term. Google search interest was derived using the rates provided by Google. We summed search interest for all search terms to result in a search interest value for each day. We derived hospitalization from vaping data by summing the number of individuals hospitalized with lung injury associated with e-cigarette use or vaping in the US on a particular week, from CDC data dated March 31 2019 - February 15 2020 [47].

We then repeated the above analyses dividing the data by online environment (blogs, comments, Facebook posts, forums, news), to determine if there was a differential effect of news events on sentiment score for each online environment. We calculated 95% confidence intervals for the association of each event with sentiment score. We only reported results where the key independent variable and its corresponding interaction term had  $P < 0.05$ . Analysis was conducted using R with the following packages: tseries, forecast and lmtest [48, 49, 50].

## Results

Post-processing, we had 3,753,248 documents (56,337,638 words) comprised of text from forums (51%), blogs (22%), Facebook posts (13%), news media (10%), and comments (5%). All data was US-based.

INSERT FIGURE 1 HERE

The mean sentiment score for all online environments summed together from August 1 2019 - April 21 2020 was -0.021 (SD=0.025). Mean sentiments by online environment were as follows: blogs (0.038, SD=0.063), comments (-0.051, SD=0.018), Facebook (-0.047, SD=0.037), forums (-0.003, SD=0.029) and news (-0.081, SD=0.089). Figure 1 illustrated sentiment scores of the online vaping arena over time, for all online environments. Aggregate sentiment for all online environments were initially negative and began to increase positively September 11 2019 onward, perhaps indicating an improvement around vaping perceptions. Sentiment for blogs and forums

increased positively after September 2019. Comments and news demonstrated an overall negative sentiment for the time period. Facebook posts were overall negative but increased positively in March 2020.

#### INSERT TABLE 1 HERE

Table 1 reported the estimates of the interrupted time series design with segmented regression analysis across various key vaping events. For all online environments, we observed a statistically significant negative association of 15% (Estimate: -0.16; 95% CI: -0.29, -0.03; P: 0.01) between sentiment score and the Trump administration's move towards a ban on flavored vaping products, possibly representing a deterioration in sentiment around vaping. This meant that the valence of text was negative at the time of the Trump vaping ban, and the text valence was correlated with that news event. Conversely, we observed a statistically significant positive association of 7% between sentiment score (Estimate: 0.07; 95% CI: 0.01, 0.14; P: 0.02) when COVID-19 was first reported to the WHO, perhaps indicating increased positive sentiment around vaping. The valence of the text was negative when COVID-19 was reported to the WHO, but rose to eventually become positive shortly after. To ensure the statistically significant findings were not chance results, we conducted validity checks. For each key vaping event, we examined a randomly generated sample of 50 associated text fragments, maintaining a distribution similar to the distribution of sources, for the time period around the event. We found that text fragments tended to mirror the results of the sentiment analysis. For example, when detailing text samples around the Trump administration's plan to ban flavored vapes, we found several documents, coded negative by sentiment analysis, detailing the negative stance toward the Trump decision:

*Instead banning flavored e-cigarettes we need to ban Donald J. Trump as president. Along with Moscow Mitch.*

*Precisely... freedom to choose. Cigarettes and alcohol are legal. Six people die from THC vapes and Trump wants to ban all vaping. It's stupid.*

*Yea Trump ban e-cigarettes but keep up approval of the sale of automatic weapons their only causing multiple deaths. Great priorities. People using e-cigarettes do it by choice. But those massacred by automatic weapons don't have a choice or say. Get a brain Trump.*

Similarly, when detailing a sample of text around Chinese government disclosure of COVID-19 to the WHO, we found multiple documents, positively coded via sentiment analysis, indicating vaping cannabidiol (CBD) as a possible COVID-19 cure or protective agent, possibly indicative of the improved sentiment when COVID-19 was first reported to the WHO. Example text fragments regarding how vaping CBD can prevent/treat COVID-19:

*COVID-19 deaths invariably involve a "cytokine storm", an excessive, unchecked immune system response. Cannabinoids from cannabis, CBD in particular, can lower cytokine production naturally. Research needed asap!*

*Why hemp cbd flowers and a vaporizer are the best covid-19 prepping tools. REDACTED applauds the use of CBD during the coronavirus outbreak.*

No statistical differences were found between the other five outcomes of interest and sentiment scores for all online environments together. While not statistically significant, there seemed to be a decrease in sentiment scores after Massachusetts banned vaping products (September 24 2019).

Regarding forums, there was a statistically significant 40% negative association between sentiment score and the Trump administration's move towards a ban on flavored vaping products (Estimate: -0.37; 95% CI: -0.52, -0.22;  $P < 0.001$ ). There was also a 9% positive association between sentiment score and COVID-19 being reported to the WHO (Estimate: 0.09; 95% CI: 0.02, 0.15;  $P: 0.01$ ). Similarly, there was a 95% positive association between sentiment score and the WHO declaring COVID-19 a pandemic for news media (Estimate: 0.67; 95% CI: 0.10, 1.24;  $P: 0.02$ ). We verified that the statistically significant findings for individual online environments were not chance results using similar techniques as above. News events seemed to have similar associations with sentiment scores of different online environments.

## Discussion

We reported two main findings. First, for all online environments taken together there was a relationship between increased negative sentiment in the online vaping environment and the Trump administration's move to ban flavored vaping products. Second, the association between improved sentiment in the vaping environment and COVID-19 report to the WHO. Recent research indicated that vaping related news events can increase risk perceptions around vaping [24, 23], but has not detailed how events are related to sentiment in online vaping spaces. The possible Trump ban on vaping may have increased risk perceptions around vaping and thus decreased sentiment in the online vaping environment. We believe that misinformation around vaping devices as a possible COVID-19 treatment [17] may have improved sentiment in the online vaping sphere during COVID-19. We propose that only these two events were salient to individuals in the online vaping arena and thus associated with statistically significant shifts in sentiment. The other events, such as the CDC investigation into vaping-related illnesses may not have produced a similar effect on sentiment as individuals did not see an immediate threat to vaping products, unlike the planned Trump ban on vaping products. Similarly, the Massachusetts ban on vaping products may not have been salient as individuals were possibly already affected by the previous Trump decision on vaping products.

There is limited research on how news events relate to sentiment around vaping, especially in response to COVID-19 news. Previous work suggested that disproportionately negative news around vaping may dissuade people who smoke unwilling to quit from moving to vaping, likely impacting health outcomes [4]. We expand on such studies, bolstering the need for balanced views on vaping, demonstrating that news events casting vaping in a negative light may increase negative sentiment around vaping, perhaps discouraging some people who smoke from switching to vapes. A recent review detailed the role of misinformation on public health outcomes [51], and we expand on past work by providing evidence on how large-scale events may create misinformation in the health sphere. We also provide evidence that some news events may be associated with improved sentiment around vaping, particularly on social media, perhaps encouraging some individuals who smoke to make the switch. However, these improvements in sentiment may be related to misinformation around COVID-19. Improvements in sentiment around vaping may come with misinformation around the subject, buttressing the need for evidence-based public health messaging around vaping. While improved sentiment around



vaping may be beneficial for those who smoke and are unable to quit, we must ensure that such sentiment is not drawn from inaccurate information, perhaps impacting the effectiveness of vaping-centric tobacco control mechanisms.

The strength of our work is the use of computational methods to explore how news events are associated with sentiment in the online vaping arena, in a range of online environments. Such outcome measurement is central to understanding how news events shift sentiment regarding vaping, allowing for accurate public health messaging around vaping when such events arise. Accurate public health messaging around vaping may augment health outcomes in two modes. Firstly, it can encourage some people who smoke in switching to vaping, improving health outcomes, even when news outlets display inordinately negative coverage regarding vaping. Secondly, such messaging can prevent misinformation, possibly leading to reduced health outcomes, such as in the case of COVID-19. For example, relying on COVID-19 misinformation, vapers who develop the condition may use vape-administered treatments, associated with reduced immune system functioning [21] perhaps heightening COVID-19 disease progression. Without accurate information, those who vape may share devices to administer unsubstantiated COVID-19 treatments - a possible site of transmission [17], perhaps increasing COVID-19 spread. We also note that messaging by public health authorities which supports vaping may inadvertently decrease risk perceptions among young people or non-smokers, who may then be more likely to try vaping, increasing their risk [52, 53]. We do not support discouraging news warning about the risks of e-cigarettes, but a careful consideration of the balance of risks related to promoting vaping as a harm reduction approach.

Public health authorities can also conduct interventions to balance the rhetoric of news events. Interventions that ask respondents to judge information accuracy around vaping [54, 55], may nudge individuals toward accurate information regarding vaping during news events which possibly distort vaping perceptions. Thus, our results may aid those who smoke and cannot quit in switching to vaping, minimize COVID-19-related misinformation around vaping, and mitigate further mischaracterizations of vaping, perhaps improving health outcomes of tobacco users. Future research can detail how some news events have a greater effect on sentiment around vaping compared to others, and address how best to intervene around disproportionate responses to vaping news events.

Our findings relied on the validity of data collected through the textual query. We searched a broad range of online media and our data contained text fragments which represented the news events in our analysis. We are thus confident in the comprehensiveness of our data. A key limitation is that we cannot say with certainty that vaping-related news events caused a shift in sentiment in the online vaping space or whether there were other underlying factors. We provide strong correlational evidence, but cannot make causal claims. We also were not able to adjust for other possible confounders. A large proportion of vaping posts on social media may be generated by bots [56], and we did not account for such posts as our goal was not to identify misinformation but code for vaping sentiment. We may have missed slang terms for vaping and perhaps underestimated the online vaping arena. Our data was drawn from August 1 2019 - April 21 2020 and we were not

able to explore the influence of news events before or after this period. Results suggested a shift in sentiment in the online vaping arena correlated with certain relevant events reported in the news, but the method did not allow a differentiation between sentiments and perceptions of people who smoke and those of dual users, those who vape, and people who previously smoked etc.

## Conclusions

Overall, we indicated that news events may be associated with either positive or negative sentiment in the online vaping environment. Depending on the nature of the event, we suggest that public health messaging may either ensure that those who smoke and wish to quit are not dissuaded from switching to vaping, or reduce incidences where those who vape injure themselves through unsubstantiated vape-related COVID-19 treatments. Findings have implications for the management of risk perceptions around vaping to improve health outcomes of tobacco users. Information-based policy instruments can be applied to balance the negative effects of news events that may create disproportionately negative vaping perceptions.

## Appendix

### Full list of search terms for textual query and Google search

e-cigarette, electronic cigarette, electronic cigarettes, electronic nicotine delivery, vape, vaping, electronic nicotine delivery system, personal vaporizer, vape pen, electric cigarette, electric nicotine delivery system, electric nicotine delivery device, e-hookah, e-juice, e-liquid.

### Declarations

Ethical Approval and Consent to participate

Approval and informed consent were not needed as we used an anonymized dataset. Yale University IRB committee guidelines waived the need for informed consent and ethical approval. Research was performed in accordance with the Declaration of Helsinki. This study was pre-registered on the Open Science Framework (OSF.IO/HZVJB).

Consent for publication

Not applicable

Availability of data and materials

The datasets used and analyzed during the current study available from the corresponding author on reasonable request.

Competing interests

Navin Kumar declares financial support through a grant from the Foundation for a Smoke-Free World, a US nonprofit 501(c)(3) private foundation with a mission to end smoking in this generation. The Foundation accepts charitable gifts from PMI Global Services Inc. (PMI); under the Foundation's Bylaws and Pledge Agreement with PMI, the Foundation is independent from PMI and the tobacco industry. There are no financial relationships with any other organizations that might have an interest in the submitted work in the previous three years; and no other relationships or activities that could appear to have influenced the submitted work. All remaining authors do not declare any conflicts of interest.

Funding

This study was funded with a grant from the Foundation for a Smoke-Free World, a US nonprofit 501(c)(3) private foundation with a mission to end smoking in this generation. The Foundation accepts charitable gifts from PMI Global Services Inc. (PMI); under the Foundation's Bylaws and Pledge Agreement with PMI, the Foundation is independent from PMI and the tobacco industry. The contents, selection, and presentation of facts, as well as any opinions expressed herein are the sole responsibility of the authors and under no circumstances shall be regarded as reflecting the positions of the Foundation for a Smoke-Free World, Inc.

Author's contributions

KJ wrote the first draft. NK contributed to the manuscript write-up and review.

#### Acknowledgements

We thank the editor and reviewers for their comments. This study was pre-registered on the Open Science Framework (OSF.IO/HZVJB). This study was funded with a grant from the Foundation for a Smoke-Free World, a US nonprofit 501(c)(3) private foundation with a mission to end smoking in this generation. The Foundation accepts charitable gifts from PMI Global Services Inc. (PMI); under the Foundation's Bylaws and Pledge Agreement with PMI, the Foundation is independent from PMI and the tobacco industry. The contents, selection, and presentation of facts, as well as any opinions expressed herein are the sole responsibility of the authors and under no circumstances shall be regarded as reflecting the positions of the Foundation for a Smoke-Free World, Inc.

#### Author details

<sup>1</sup>, Yale School of Medicine, New Haven, CT, USA. <sup>2</sup>, Yale College, New Haven, CT, USA. <sup>3</sup>210 Prospect St, New Haven, CT 06511.

#### References

- McNeill, A., Brose, L.S., Calder, R., Bauld, L., Robson, D.: Evidence review of e-cigarettes and heated tobacco products 2018. A report commissioned by Public Health England. London: Public Health England **6** (2018)
- US Food and Drug Administration and others: Fda announces comprehensive regulatory plan to shift trajectory of tobacco-related disease, death. Agency to pursue lowering nicotine in cigarettes to non-addictive levels and create more predictability in tobacco regulation. Maryland (2017)
- Nyman, A.L., Huang, J., Weaver, S.R., Eriksen, M.P.: Perceived comparative harm of cigarettes and electronic nicotine delivery systems. *JAMA Network Open* **2**(11), 1915680–1915680 (2019)
- Tattan-Birch, H., Brown, J., Shahab, L., Jackson, S.E.: Association of the us outbreak of vaping-associated lung injury with perceived harm of e-cigarettes compared with cigarettes. *JAMA Network Open* **3**(6) (2020). doi:10.1001/jamanetworkopen.2020.6981
- Polosa, R., Morjaria, J.B., Caponnetto, P., Prosperini, U., Russo, C., Pennisi, A., Bruno, C.M.: Evidence for harm reduction in copd smokers who switch to electronic cigarettes. *Respiratory Research* **17**(1) (2016). doi:10.1186/s12931-016-0481-x
- Polosa, R., Cibella, F., Caponnetto, P., Maglia, M., Prosperini, U., Russo, C., Tashkin, D.: Health impact of e-cigarettes: a prospective 3.5-year study of regular daily users who have never smoked. *Scientific Reports* **7**(1) (2017). doi:10.1038/s41598-017-14043-2
- CDC. Centers for Disease Control and Prevention (2021). [https://www.cdc.gov/tobacco/basic\\_information/e-cigarettes/about-e-cigarettes.html](https://www.cdc.gov/tobacco/basic_information/e-cigarettes/about-e-cigarettes.html)
- Cantrell, J., Huang, J., Greenberg, M., Willett, J., Hair, E., Vallone, D.: History and current trends in the electronic cigarette retail marketplace in the united states: 2010–2016. *Nicotine and Tobacco Research* **22**(5), 843–847 (2020)
- Stratton, K., Kwan, L.Y., Eaton, D.L., et al.: Public health consequences of e-cigarettes: consensus study report. Public health consequences of e-cigarettes: consensus study report. (2018)
- Berry, K.M., Fetterman, J.L., Benjamin, E.J., Bhatnagar, A., Barrington-Trimis, J.L., Leventhal, A.M., Stokes, A.: Association of electronic cigarette use with subsequent initiation of tobacco cigarettes in us youths. *JAMA network open* **2**(2), 187794–187794 (2019)
- Hartmann-Boyce, J., McRobbie, H., Butler, A.R., Lindson, N., Bullen, C., Begh, R., Theodoulou, A., Notley, C., Rigotti, N.A., Turner, T., et al.: Electronic cigarettes for smoking cessation. *Cochrane database of systematic reviews* (9) (2021)
- Nyakutsikwa, B., Britton, J., Bogdanovica, I., Langley, T.: Vitamin e acetate is not present in licit e-cigarette products available on the uk market. *Addiction* **115**(4), 782–783 (2020). doi:10.1111/add.14920
- The White House: Remarks by President Trump in Meeting on E-Cigarettes. Accessed 2020-06-30 (2019). <https://www.whitehouse.gov/briefings-statements/remarks-president-trump-meeting-e-cigarettes/>
- Jaffe, S.: Will trump snuff out e-cigarettes? *The Lancet* **394**(10213), 1977–1978 (2019)
- Hartnett, K.P., Kite-Powell, A., Patel, M.T., Haag, B.L., Sheppard, M.J., Dias, T.P., King, B.A., Melstrom, P.C., Ritchey, M.D., Stein, Z., et al.: Syndromic surveillance for e-cigarette, or vaping, product use-associated lung injury. *New England Journal of Medicine* **382**(8), 766–772 (2020)
- Glantz, S.A.: Reduce your risk of serious lung disease caused by corona virus by quitting smoking and vaping. *Global Center for Good Governance in Tobacco Control (GGTC)* (2020). <https://ggtc.world/2020/05/06/reduce-your-risk-of-serious-lung-disease-caused-by-corona-virus-by-quitting-smoking-and-vaping/>
- Majmundar, A., Allem, J.-P., Cruz, T.B., Unger, J.B.: Public health concerns and unsubstantiated claims at the intersection of vaping and covid-19. *Nicotine & Tobacco Research* (2020)
- Lewis, T.: Smoking or vaping may increase the risk of severe coronavirus infection. *Scientific American* **17** (2020)
- Oh, A.Y., Kacker, A.: Do electronic cigarettes impart a lower potential disease burden than conventional tobacco cigarettes?: Review on e-cigarette vapor versus tobacco smoke. *The Laryngoscope* **124**(12), 2702–2706 (2014). doi:10.1002/lary.24750
- Esposito, G., Pesce, M., Seguela, L., Sanseverino, W., Lu, J., Corpetti, C., Sarnelli, G.: The potential of cannabidiol in the covid-19 pandemic: a hypothesis letter. *British Journal of Pharmacology* (2020)
- Brown, J.D.: Cannabidiol as prophylaxis for sars-cov-2 and covid-19? unfounded claims versus potential risks of medications during the pandemic. *Research in social & administrative pharmacy: RSAP* (2020)
- Notley, C., Ward, E., Dawkins, L., Holland, R.: The unique contribution of e-cigarettes for tobacco harm reduction in supporting smoking relapse prevention. *Harm Reduction Journal* **15**(1) (2018). doi:10.1186/s12954-018-0237-7
- Duong, H.T., Liu, J.: Vaping in the news: The influence of news exposure on perceived e-cigarette use norms. *American Journal of Health Education* **50**(1), 25–39 (2018). doi:10.1080/19325037.2018.1548315

24. Dave, D., Dench, D., Kenkel, D., Mathios, A., Wang, H.: News that takes your breath away: Risk perceptions during an outbreak of vaping-related lung injuries (2020). doi:10.3386/w26977
25. Tanne, J.H.: Don't vape, cdc says, as us lung disease epidemic grows. *BMJ*, 5479 (2019). doi:10.1136/bmj.l5479
26. Ferrer, R.A., Portnoy, D.B., Klein, W.M.: Worry and risk perceptions as independent and interacting predictors of health protective behaviors. *Journal of health communication* **18**(4), 397–409 (2013)
27. Rivers, C.M., Lewis, B.L.: Ethical research standards in a world of big data. *F1000Research* **3** (2014)
28. Lamy, F.R., Daniulaityte, R., Sheth, A., Nahhas, R.W., Martins, S.S., Boyer, E.W., Carlson, R.G.: "those edibles hit hard": Exploration of twitter data on cannabis edibles in the us. *Drug and alcohol dependence* **164**, 64–70 (2016)
29. Kim, Y., Huang, J., Emery, S.: Garbage in, garbage out: data collection, quality assessment and reporting standards for social media data use in health research, infodemiology and digital disease detection. *Journal of medical Internet research* **18**(2), 41 (2016)
30. Glasser, A.M., Collins, L., Pearson, J.L., Abudayyeh, H., Niaura, R.S., Abrams, D.B., Villanti, A.C.: Overview of electronic nicotine delivery systems: a systematic review. *American journal of preventive medicine* **52**(2), 33–66 (2017)
31. McCausland, K., Maycock, B., Leaver, T., Jancey, J.: The messages presented in electronic cigarette-related social media promotions and discussion: scoping review. *Journal of Medical Internet Research* **21**(2), 11953 (2019)
32. Huang, J., Kornfield, R., Szczypka, G., Emery, S.L.: A cross-sectional examination of marketing of electronic cigarettes on twitter. *Tobacco control* **23**(suppl 3), 26–30 (2014)
33. CDC: CDC, states investigating severe pulmonary disease among people who use e-cigarettes. Accessed 2020-06-30 (2019). <https://www.cdc.gov/media/releases/2019/s0817-pulmonary-disease-ecigarettes.html>
34. Knowles, H.: Massachusetts to ban sale of all vaping products for 4 months in toughest state crackdown. Accessed 2020-06-30 (2019). <https://www.washingtonpost.com/health/2019/09/24/massachusetts-ban-all-vaping-product-sales-months-toughest-state-crackdown/>
35. Berke, J.: The FDA just leveled a stark warning against using any vapes containing THC amid an outbreak of lung disease. Accessed 2020-06-30 (2019). <https://www.businessinsider.com/fda-warns-consumers-not-to-use-marijuana-vapes-amid-lung-disease-crisis-2019-10>
36. Grady, D.: Vaping Illnesses Are Linked to Vitamin E Acetate, C.D.C. Says. Accessed 2020-06-30 (2019). <https://www.nytimes.com/2019/11/08/health/vaping-illness-cdc.html>
37. WHO: WHO Timeline - COVID-19. Accessed 2020-05-29 (2020). <https://www.who.int/news-room/detail/27-04-2020-who-timeline-covid-19>
38. Pang, B., Lee, L., et al.: Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval* **2**(1–2), 1–135 (2008)
39. Bae, Y., Lee, H.: Sentiment analysis of twitter audiences: Measuring the positive or negative influence of popular twitterers. *Journal of the American Society for Information Science and Technology* **63**(12), 2521–2535 (2012)
40. Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing contextual polarity in phrase-level sentiment analysis. In: *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing*, pp. 347–354 (2005)
41. Blankers, M., van der Gouwe, D., van Laar, M.: 4-fluoramphetamine in the netherlands: Text-mining and sentiment analysis of internet forums. *International Journal of Drug Policy* **64**, 34–39 (2019)
42. Rinker, T.: Package 'sentimentr'. Retrieved **8**, 31 (2017)
43. Peng, R.D., Dominici, F., Louis, T.A.: Model choice in time series studies of air pollution and mortality. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **169**(2), 179–203 (2006)
44. Kontopantelis, E., Doran, T., Springate, D.A., Buchan, I., Reeves, D.: Regression based quasi-experimental approach when randomisation is not an option: interrupted time series analysis. *bmj* **350**, 2750 (2015)
45. Bernal, J.L., Cummins, S., Gasparrini, A.: Interrupted time series regression for the evaluation of public health interventions: a tutorial. *International journal of epidemiology* **46**(1), 348–355 (2017)
46. Ayers, J.W., Althouse, B.M., Dredze, M.: Could behavioral medicine lead the web data revolution? *Jama* **311**(14), 1399–1400 (2014)
47. Centers for Disease Control and Prevention and others: Outbreak of lung injury associated with e-cigarette use, or vaping. Retrieved October **23**, 2019 (2019)
48. Trapletti, A., Hornik, K.: Tseries: Time Series Analysis and Computational Finance. (2019). R package version 0.10-47. <https://CRAN.R-project.org/package=tseries>
49. Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F., Razbash, S., Wang, E., Yasmeeen, F.: forecast: Forecasting Functions for Time Series and Linear Models. (2020). R package version 8.12. <http://pkg.robjhyndman.com/forecast>
50. Hothorn, T., Zeileis, A., Farebrother, R.W., Cummins, C., Millo, G., Mitchell, D.: lmtree: Testing Linear Regression Models. (2019). R package version 0.9-37. <https://cran.r-project.org/web/packages/lmtree>
51. Swire-Thompson, B., Lazer, D.: Public health and online misinformation: challenges and recommendations. *Annual Review of Public Health* **41**, 433–451 (2020)
52. Gendall, P., Hoek, J.: Role of flavours in vaping uptake and cessation among new zealand smokers and non-smokers: a cross-sectional study. *Tobacco Control* (2020)
53. Thorndike, A.N.: E-cigarette use by young adult nonsmokers: Next-generation nicotine dependence? *Annals of internal medicine* **170**(1), 70–71 (2019)
54. Pennycook, G., Rand, D.G.: Fighting misinformation on social media using crowdsourced judgments of news source quality. *Proceedings of the National Academy of Sciences* **116**(7), 2521–2526 (2019)
55. Barnett, P.A., Hoskins, C.E., Alhoti, J.A., Carpenter, L.J.: Reducing public misinformation about organ

- donation: An experimental intervention. *Journal of Social Distress and the Homeless* **18**(1-2), 57–73 (2009)
56. Martinez, L.S., Hughes, S., Walsh-Buhi, E.R., Tsou, M.-H.: “okay, we get it. you vape”: an analysis of geocoded content, context, and sentiment regarding e-cigarettes on twitter. *Journal of health communication* **23**(6), 550–562 (2018)

#### Figures

**Figure 1** Online vaping arena sentiment score scatterplots and trends with a timeline of vaping-related news events, for various online environments.

- (a) August 17 2019: CDC announces investigation into vaping-related illnesses
- (b) September 11 2019: Trump administration considers ban on vaping products
- (c) September 24 2019: Massachusetts bans vaping products
- (d) October 4 2019: FDA warns against using vape products containing THC
- (e) November 8 2019: CDC announces relationship between Vitamin E acetate and lung injury outbreak
- (f) December 31 2019: COVID-19 first reported to the WHO
- (g) March 11 2020: WHO declares COVID-19 a pandemic

Note. Association between vaping sentiment online and exposure to news events. On the left vertical axis, a sentiment score of 0 indicated that on average the sentiment towards vaping is neither positive nor negative. A sentiment score  $>0$  indicated an averaged positive sentiment towards vaping. The solid line indicated the smoothed sentiment score with the corresponding confidence interval (in grey) and was obtained by fitting the date of the posts to the sentiments of the posts using an interrupted time series design with segmented regression analysis. Red boxes denoted news events that had statistically significant association ( $p < 0.05$ ) with shifts in sentiment scores.

#### Tables

**Table 1** Estimates of the interrupted time series design with segmented regression analysis across various key vaping events, for various online environments

Event	All online environments		Blogs		Comments		Facebook		Forums		News media	
	Estimate (95% CI)	p interaction	Estimate (95% CI)	p interaction	Estimate (95% CI)	p interaction	Estimate (95% CI)	p interaction	Estimate (95% CI)	p interaction	Estimate (95% CI)	p interaction
CDC announcing an investigation into vaping-related illnesses	0.017 (-0.014, 0.047)	0.282<0.001	-0.039 (-0.168, 0.090)	0.552 0.702	0.029 (0.001, 0.057)	<b>0.039</b> <b>0.002</b>	0.003 (-0.014, 0.068)	0.194 0.004	0.044 (0.010, 0.078)	<b>0.010</b> <.001	-0.018 (-0.176, 0.139)	0.819 0.626
Trump administration plan to ban some vaping products	-0.162 (-0.292, -0.032)	<b>0.014</b> <b>0.016</b>	-0.180 (-0.583, 0.223)	0.381 0.386	0.023 (-0.150, 0.090)	0.627 0.626	0.143 (-0.188, 0.160)	0.972 0.915	-0.371 (-0.519, -0.223)	<.001 <.001	-0.212 (-0.887, 0.462)	0.537 0.468
Massachusetts bans vaping products	0.104 (-0.142, 0.350)	0.408 0.300	0.182 (-0.479, 0.844)	0.589 0.495	0.002 (-0.222, 0.227)	0.984 0.968	0.121 (-0.206, 0.447)	0.467 0.440	0.212 (-0.104, 0.029)	0.134 0.042	0.230 (-1.039, 1.499)	0.722 0.671
FDA warns consumers against the use of vape products containing THC	0.005 (-0.213, 0.222)	0.043 0.959	-0.116 (-0.749, 0.518)	0.720 0.772	0.007 (-0.197, 0.211)	0.946 0.970	-0.107 (-0.400, -0.186)	0.473 0.469	0.126 (-0.127, 0.379)	0.270 0.489	-0.182 (-1.317, 0.952)	0.753 0.768
CDC announces link between Vitamin E acetate and EVALI	-0.005 (-0.069, 0.059)	0.884 0.558	0.007 (-0.444, 0.458)	0.975 0.766	0.028 (-0.027, 0.083)	0.313 0.245	0.032 (-0.051, 0.115)	0.449 0.300	-0.038 (-0.104, 0.029)	0.270 0.489	-0.107 (-0.431, 0.217)	0.518 0.942
COVID-19 first reported to the WHO	0.073 (0.012, 0.135)	<b>0.020</b> <b>0.026</b>	0.141 (-0.428, 0.709)	0.628 0.617	0.023 (-0.031, 0.077)	0.413 0.595	0.026 (-0.056, 0.107)	0.531 0.729	0.085 (0.017, 0.150)	<b>0.014</b> <b>0.013</b>	0.165 (-0.151, 0.482)	0.306 0.179
WHO declared COVID-19 a pandemic	-0.104 (-0.216, 0.009)	0.071 0.076	0.060 (-0.826, 0.945)	0.895 0.911	0.021 (-0.076, 0.118)	0.675 0.595	-0.868 (-1.014, -0.722)	<.001 <.001	-0.195 (-0.313, -0.078)	<b>0.001</b> <b>0.002</b>	0.670 (0.099, 1.241)	<b>0.021</b> <b>0.020</b>

Bold p-values indicate statistical significance ( $P < 0.05$ ) for both the news event (key dependent variable) and the corresponding interaction term ( $P < 0.05$ ).