HHS Public Access
Author manuscript
*Int J Med Inform*. Author manuscript; available in PMC 2022 November 01.

Published in final edited form as:
*Int J Med Inform*. 2021 November ; 155: 104574. doi:10.1016/j.ijmedinf.2021.104574.

# Using a Mixed Methods Approach to Identify Public Perception of Vaping Risks and Overall Health Outcomes on Twitter During the 2019 EVALI Outbreak

**Erin Kasson, MS**[1], **Avineet Kumar Singh, MS**[2], **Ming Huang, PhD**[3], **Dezhi Wu, PhD**[4,*], **Patricia Cavazos-Rehg, PhD**[1]

[1]Department of Psychiatry, Washington University School of Medicine, 660 S Euclid Ave, St. Louis, MO, USA

[2]Department of Computer Science and Engineering, University of South Carolina, Columbia, SC, USA

[3]Department of Artificial Intelligence and Informatics Research, Mayo Clinic, Rochester, MN, USA

[4]Department of Integrated Information Technology, University of South Carolina, Columbia, SC, USA

## Abstract

**Introduction:** Vaping product use (i.e., e-cigarettes) has been rising since 2000 in the United States. Negative health outcomes associated with vaping products have created public uncertainty and debates on social media platforms. This study explores the feasibility of using social media as a surveillance tool to identify relevant posts and at-risk vaping users.

**Methods:** Using an interdisciplinary method that leverages natural language processing and manual content analysis, we extracted and analyzed 794,620 vaping-related tweets on Twitter. After observing significant increases in vaping-related tweets in July, August, and September 2019, additional human coding was completed on a subset of these tweets to better understand primary themes of vaping-related discussions on Twitter during this time frame.

**Results:** We found significant increases in tweets related to negative health outcomes such as acute lung injury and respiratory issues during the outbreak of e-cigarette/vaping associated lung injury (EVALI) in the fall of 2019. Positive sentiment toward vaping remained high, even across the peak of this outbreak in July, August, and September. Tweets mentioning the public

[*]**Corresponding Author** Dezhi Wu, PhD, Department of Integrated Information Technology, University of South Carolina Columbia, SC, USA, dezhiwu@cec.sc.edu.

**Authors' Statement:** All the authors substantially contributed to the conceptualization or design of the study, or data extraction, or analysis and interpretation of data; drafting the article or revising it critically for important intellectual content; and final approval of the version to be published.

**Conflicts of interest:** None of the authors has any relevant conflicts of interest or financial disclosures that should be reported here.

**Ethics and Institutional Review Board approval:** Ethics and Institutional Review Board (IRB) approval was obtained from Washington University – IRB Approval #202101009.

perceptions of youth risk were concerning, as were increases in marketing and marijuana-related tweets during this time.

**Discussion:** The preliminary results of this study suggest the feasibility of using Twitter as a means of surveillance for public health crises, and themes found in this research could aid in specifying those groups or populations at risk on Twitter. As such, we plan to build automatic detection algorithms to identify these unique vaping users to connect them with a digital intervention in the future.

### Keywords

tweets; public health surveillance; vaping; tobacco use; EVALI; youth

## 1 Introduction

From 2011 to 2018, the number of e-cigarette users increased from approximately 7 million to nearly 40 million users with approximately 3.2% of US adults reporting use of vaping products in 2018 [1, 2]. Of further concern, 4.7% of middle school and 19.6% of high school students reporting use of these products in 2020 [3, 4] and the US Food and Drug Administration (FDA) declared youth vaping an epidemic in 2018 [5, 6]. In the fall of 2019, there was a spike in acute health consequences related to vaping, including lung injury, respiratory infections, seizures, and even death [7-9]. There was a rapid increase in cases and severity of E-cigarette or Vaping Associated Lung Injuries (now referred to as EVALI) during this period, with the CDC reporting 2,807 cases and 68 deaths related to this EVALI outbreak in 2019 [5]. This outbreak started in the spring of 2019, with US public health officials opening formal investigations to EVALI cases in July 2019 [10], increasing new coverage on this issue. On September 27, 2019, the CDC released information linking the use of unregulated and marijuana vaping products with the EVALI [11].

Social media platforms are often used to share personal opinions and views related to substance use, including the misuse of prescription drugs [12], substance effects and adverse side effects [13, 14], and substance use recovery [15]. This content is rich with both public health and clinical information, is publicly accessible, and is a viable way to examine substance use trends. As a social media platform heavily covering both media and news perspectives as well as personal opinions on a variety of topics, Twitter has the capacity to be used as a means of surveillance for public health crises [16], such as the EVALI crisis of 2019. Prior studies have documented that Twitter can be used to characterize and monitor public and media discussions on mental health and substance use topics [17-19], including vaping [20, 21] through providing valuable insights and informative strategies for detection and outreach for prevention and intervention. Numerous technical strategies have been leveraged to investigate content and public perceptions surrounding mental health and substance use specifically on Twitter, a social media platform heavily covering both media and news perspectives as well as personal opinions on a variety of topics. For example, text mining and statistical analysis have been used in the examination of depression and suicidal content on the Twitter platform [18, 22] and in the understanding of vaping flavors worldwide [21]. Further, sentiment analyses have been used on Twitter to better understand tobacco use attitudes and public perception [20, 23]. Additionally, natural language processing (NLP) and machine learning (ML) have been applied to text mining of Twitter content to inform detection models aimed to identify tweets relevant to vaping, tweets with pro-vape sentiments [24] and purchasing and use behaviors surrounding illicit opioids and marijuana access [27].

These studies, however, were conducted prior to the EVALI outbreak in 2019. Recent literature related to the EVALI crisis on Twitter has used social network analysis to track the spread of information [26] related to negative health outcomes and infoveillance strategies to correlate hospital admissions and death related to vaping during this time [27]. Nevertheless, previously limited integration of both clinical and technical expertise provides an opportunity to expand possible feasibility evaluations toward a more detailed and clinically informed investigation of vaping and the EVALI outbreak on Twitter.

In this study, we leverage text mining techniques (e.g., NLP) along with traditional behavioral content analysis approaches to examine tweets related to the EVALI outbreak in fall 2019 on Twitter [28]. NLP techniques enable automatic analyses of a large amount of Twitter data via term frequency and sentiment analysis, while traditional behavioral content analyses allow for deeper insights into themes within a small sample of this Twitter data. The combined approach aims to better understand the context with which keywords, phrases, sentiments, and themes shifts occurred within these discussions. Our results may inform later development of an advanced detection algorithm based on NLP and ML for identifying unique Twitter users who may be at risk for vaping-related negative health outcomes. The current study seeks (1) to identify trends of the EVALI crisis discussions on Twitter across 2019 and generate main themes using NLP and human annotation methods, and (2) to examine sentiment surrounding vaping on Twitter in relation to negative health outcomes discussed during this time frame.

## 2 Methods

We extracted a random sample of 794,620 tweets from January 2019 to December 2019. This strategy of sampling pre-, during- and post-EVALI outbreak tweets gathered roughly 40% of the extracted 2019 vaping-related tweets using specified keywords, thus providing a representative sample for analysis. Specific keywords for this study were clinically relevant keywords based on our manual reviews of nearly 200 randomly sampled tweets across the 2019 timeframe by our clinical team. Keyword list (See Appendix 1) included "vaping/vape" and 60 other specifying terms from several categories guided by our primary research questions on exploring: (1) negative health outcomes from vaping, (2) positive health outcomes from vaping, (3) sentiments toward vaping products, (4) marijuana/weed or related unregulated product, and (5) youth and young adult age groups mentioned.

After data cleaning, we used three approaches for analysis of this dataset extracted from Twitter. *First,* we conducted a sentiment analysis to determine positive or negative sentiment toward vaping in tweets during each of the respective months in 2019. *Second,* we utilized frequency counts from the different keywords identified to generate trend results for these main themes. *Third,* we completed manual coding on a subset of 600 tweets (200 from July, August, September 2019) to more thoroughly explore those themes identified in the trend and sentiment analyses from a clinical point of view. An overview of these methods is provided in Figure 1.

The methods of data extraction and analysis were reviewed by the Washington University Institutional Review Board (IRB# 202101009). As the data examined were publicly available social media data, this study was determined to be non-human subjects research and exempt from review.

## 2.1 Data Collection

We used an open-sourced Python library, GetOldTweets3 [29] with identified vaping keywords provided by our clinical team for accessing the Twitter APIs to extract a random sample of 794,620 tweets from January 2019 to December 2019. The Python module GetOldTweets3 allows us to access and extract historical tweets of any date and topic, specified with a constraint of the Twitter APIs providing limited access to data more than 1 week old. There were no restrictions for scraping a large number of tweets or accessing historical tweets using this method, therefore providing a representative sample of tweets for subsequent analysis [29, 30].

## 2.2 Sentiment Analyses

Sentiment analysis is a common computational technique to measure the subjectivity, opinions, attitudes, and emotions in texts [31]. This approach quantifies the sentiment contents in the given texts in a continuum scale (e.g., [−1, 1]) [32]. We deployed a Python library for processing textual data, namely TextBlob, for sentiment analysis [33]. TextBlob provides a simple API for diving into common NLP tasks [34, 35]. We used the PatternAnalyzer module implemented in the TextBlob with the pattern library [36] to evaluate the sentiment of a tweet. The PatternAnalyzer reports the sentiment score and subjectivity of a tweet in a named tuple of the form sentiment (polarity, subjectivity). The polarity score is a float within the range [−1.0, 1.0]. The subjectivity is a float within the range [0.0, 1.0], where 0.0 is very objective and 1.0 is very subjective. Based on the sentiment score, our approach used in this study further classified tweets into positive, negative, and neutral sentiment toward vaping. We then calculated the distribution of tweets in terms of the three sentiment types per month.

## 2.3. Keyword Categorization and Trend Analysis of Primary Themes

After tokenizing each tweet into words, we removed the stop words, irrelevant words and special characters and calculated the frequency of each word and its monthly trend. We used a word cloud to visualize the top discussed words in the tweet collection, a common way to represent themes and frequencies by depicting specific words related to this topic found in the dataset in different sizes and color contrast. The bigger and brighter the word appears, the more often it is mentioned within all tweets.

We further manually categorized the frequently mentioned words into three important themes/topics relevant to our research questions and study aims. We then analyzed the frequencies of top mentioned words linked to the topics in the tweet corpus and their trends over time on: (1) negative health outcomes, (2) marijuana or unregulated products, and (3) youth and young adult age groups mentioned.

### 2.4 Manual Coding and Analyses

During the EVALI outbreak, July, August, and September of 2019 were identified as months during and just prior to the dramatic increase in vaping-related discussions in September on Twitter based on both the content and sentiment analyses outlined above. As such, a random sample of 200 tweets were pulled from each of these months for in-depth human coding toward contextual content analysis. Specifically, members of our clinical team with experience in substance use research (students in psychology, social work, or public health at the graduate level and/or with relevant experience coding qualitative social media data led by author PCR, a licensed clinical psychologist) used inductive and deductive methods to construct a codebook based on a review of sample tweets and informed by previous literature [37, 38]. Three primary coding categories were used: (1) type of tweet, including personal, marketing, or media/news/other [39]; (2) sentiment toward vaping [40]; and (3) health outcomes mentioned, including both positive (e.g., quitting combustible smoking) and negative (e.g., lung injury, death, addiction/dependence) [39, 41]. Secondary concepts that were coded as either present/not present including (1) Mentions teens/adolescents/young adults [42]; and (2) Mentions marijuana/weed/CBD/THC [43, 44]; Two independent human coders reviewed each tweet and assigned applicable codes based on text content with an average KAPPA score of 0.62, which is deemed as a substantial agreement [45]. A third coder then reviewed the coding from each preliminary coder and provided final codes for those tweets on which there was disagreement [46], which is a third-party resolution method used in previous qualitative analysis literature [47]. Both frequency and qualitative themes were then compared to the preliminary results from the computer science analyses to aid in the conceptualization of the clinical themes reflected in the dataset.

## 3 RESULTS

Overall trends in vaping-related tweets used for both sentiment and content analyses and involved unique users are shown in Figure 2. The vaping-related tweets and involved unique users had a similar trend in 2019. Results from analysis of the 794,620 tweets collected in 2019 indicated a spike in vaping-related discussions on Twitter in August and September, with a slight increase already in July.

### 3.1 Sentiment Keywords and Sentiment Analysis

All vaping-related tweets increase in frequency in July, August, and September, and positive sentiment tweets were found to outnumber negative sentiment tweets even throughout the peak of the EVALI crisis (see Figure 3A). As an example, in September 2019, positive sentiment tweets reached their peak with 98,334 posts comparing to 62,202 negative sentiment posts. In order to better understand how sentiment may have differed across types of users/organizations posting this content, we further categorized these vaping-related tweets into three categories: (1) Tweets associated with vaping brand advertisements (using keywords such as “JUULvapor,” “blucigs,” “csvape,” “ijoyglobal” etc.), for which sentiments are illustrated in Figure 3B; (2) Tweets mentioned news and government agencies toward vaping (using keywords including "WSJ," "time," "washingtonpost," "nytimes," "ABC," "CNN," "CDCgov," "US_FDA," "FDACommissioner," "FDAMedWatch," "FDATobacco" and so on), and (3) Tweets mentioned vaping posted by general Twitter users for which sentiments are shown in Figure 3C. It is expected that vaping brands posted far more positive sentiment tweets toward vaping, for example, in September 2019, only 18 negative sentiment tweets in comparison to 136 positive vaping-related tweets. Notably, among the vaping-related tweets mentioned in news and government agencies, positive sentiment tweets are also far exceeding negative sentiment tweets toward vaping in 11 out of 12 months with a peak month in September 2019 indicating 136 positive vaping-related tweets in comparison to 61 negative tweets. The postings from government agencies such as the CDC were mostly to advocate for people to quit vaping and understand risks. It was further observed that positive sentiments from vaping business brands (2535 tweets, 0.319%) and the news and government agencies (1020 tweets, 0.128%) merely represented a minor portion of overall extracted tweets (totally 794,620 tweets) in 2019. During the EVALI outbreak period, the positive sentiment trend was similar to the whole year's trend that only 0.29% of positive tweets were from vaping brands and 0.14% of them were from news and government agencies.

Human coding reflected a bias toward positive sentiment tweets in July and August but noted a shift toward negative sentiment in September 2019. Tweets with positive sentiment were most often defending the use of vaping products during the crisis, stating that tobacco vaping is distinctly different and safer than marijuana vaping and continuing to assert that vaping is an essential alternative to combustible tobacco smoking (see Table 1).

### 3.2 Content Analysis of Primary Themes

We examined the mentions of keywords on three primary themes overtime in 2019, including 1) negative health outcomes (NHO), 2) marijuana/unregulated products (MJ), and 3) youth /adolescents (ADOL). Figure 4 outlines the frequency of *keyword mentions* for each of these themes across each month and visual depiction of top keywords appearing in July, August, and September 2019 are provided in a word cloud in Figure 5.

Human coding at the *tweet level* delineated positive and negative health outcomes mentioned related to vaping, demonstrating that positive health outcomes (e.g., mentions of quitting combustible smoking) were mentioned more frequently than negative health outcomes (e.g., EVALI) in July and August, with a shift in September during the peak of the EVALI crisis in 2019 (see Table 2). Nearly all tweets mentioning positive health outcomes due to vaping were related to the use of these products to quit combustible tobacco smoking. Further, mentions of health workers, researchers, and medical outcomes were most often discussed in tweets sharing media or news articles related to EVALI. Due to public health considerations about the contribution of additives and unregulated marijuana vaping products to the EVALI crisis, tweets mentioning marijuana and black-market products were also examined across 2019 (see Table 3). Although trends indicate youth keywords were mentioned along the same trend line as negative health outcomes and unregulated product keywords, human coding reflected a higher rate of youth-related tweets in July, with a decreased in August and September (Table 4). July tweets most often mentioned adolescents and young adults as a group at higher risk for vaping product use and for the need for prevention efforts among this group, while August and September tweets outlined public health messages toward all age groups regarding acute EVALI concerns.

## 4 Discussion

Investigations into discussions on Twitter during the EVALI crisis of fall 2019 support the use of Twitter as a means of public health surveillance for substance use concerns, including health outcomes surrounding vaping. The text mining and manual content analysis approaches used in this study provide insights on not only important themes and keywords post-crisis, and sentiment of these tweets were also examined. Furthermore, in-depth human coding analysis from a clinical perspective in this study also informs the design and development of future detection algorithms for vaping-related risks on social media platforms, strategies which have particular importance during the COVID-19 pandemic in which access to traditional recovery supports may be limited due to social distancing and public health precautions. A detailed understanding of public and media perceptions on Twitter as outlined in this study is crucial toward developing accurate ML models for intelligently detecting those themes most related to health risks and informing subsequent outreach strategies to unique Twitter users.

Notably, multiple computer science techniques as well as human coding identified higher rates of positive sentiment tweets than negative sentiment tweets, even when the EVALI concerns including death increased during July and August of 2019. Further, mentions of an overall teen vaping epidemic couched in long-term effects were replaced with keywords and discussions on the acute EVALI symptoms emerging. Human coding analysis further identified highly polarized viewpoints (i.e., pro-vaping vs. anti-vaping) on Twitter with regard to vaping, even in July prior to the dramatic increases in August and July. As negative health outcomes were increasing and becoming more severe, pro-vaping groups appeared to defend vaping as an essential form of harm reduction for combustible smoking. Policies in the news during this time frame regarding banning vaping products received several critiques, with tweets suggesting that restriction of regulated tobacco vaping products would drive vaping product users into the black market (where risk is higher) and would remove a necessary harm reduction tool for those dependent on nicotine pushing them back toward combustible tobacco use. While marijuana and unregulated products were later identified as drivers of the EVALI crisis [9], these investigations and detailed examinations of public perceptions inform future public health outreach efforts among the vaping community or those at-risk for uptake, suggesting receptivity of risk information from peers or others who vape may be higher than from public health or medical research sources [48, 49].

The mixed methods approach used in this study suggest the feasibility of developing algorithms for detecting users at risk for negative health outcomes, or those who are at risk for vaping product uptake. Computer science strategies allowed for a representative and surveillance-based overview of public perception, primary topics, and sentiment shifts surrounding vaping on Twitter, while more detailed clinical coding and interpretation of content at the individual level can inform extension of these results toward the development of feasible and effective methods of outreach and intervention. In July 2019, there were few mentions of negative health outcomes due to vaping, and most of these tweets were from media or news organizations. However, in August 2019, the number of unique users increased, suggesting that individual users were now reacting and sharing opinions, both rejecting or accepting, the emerging news surrounding EVALI incidences. This increase in individual users suggests that such time frames may allow for targeted and timely detection and outreach efforts for such health crises, which has been shown in previous literature for behavioral health including substance use [18, 23]. Further, tweets during the peak of the EVALI crisis also mentioned adolescent and young groups thought to be most at risk, which could be used to inform algorithm development to employ strategies to directly target users of this age group for intervention.

The current study has several strengths, including the EVALI timeframe from which the data was extracted, and the interdisciplinary strategies used for analysis. Many of the studies using Twitter for surveillance of vaping-related content were conducted prior to the EVALI outbreak, and the use of 2019 data in the current study allows for identification of specific topics with regard to health risks and outcomes, uncertainty surrounding public health communications, and distinct public sentiment toward vaping during this crisis. Further, our inclusion of human coding and analysis of themes provides a more in-depth evaluation of public opinions on vaping and experiences of those who vape during such an outbreak. Topics identified in this study may be both unique to this 2019 timeframe, as well as highlight signals of on-going vaping discussions and debates on this platform. There are also several limitations to note. *First,* we were only able to delineate personal and marketing tweets from all other types (e.g., media/news, others) within a smaller sample using human coding. *Second,* we were unable to categorize each tweet based on demographics or user age and were only able to analyze age groups mentioned within the text of each tweet to provide an overall understanding of public perceptions of risk groups. Additionally, a more comprehensive extraction of tweets and a large sample of human coding throughout the entire 2019 timeframe could aid in better theme representation and more nuanced analyses to provide context for surveillance feasibility even prior to July 2019.

Overall, our interdisciplinary approach combining text mining and qualitative clinical coding supported the use of Twitter as a platform for surveillance of public sentiment and discussion surrounding vaping health outcomes. Such surveillance methods could be extended to other substance use and mental health concerns on both Twitter and other social media platforms, to better understand and inform strategies to improve public health outcomes at large.

## ACKNOWLEDGMENTS

Funding for this work was provided by the National Institutes of Health (NIH) [Grant No: K02 DA043657 AWARD (Cavazos-Rehg), the University of South Carolina [Grant No: 80002838 (Wu)], and with partial support from the USC Big Data Health Science Center, a USC excellence initiative program [Grant No: BDHSC-2021-14]. The content is solely the authors' responsibility and does not necessarily represent the official views of the funding agencies. We would also like to acknowledge Nina Kaiser, Melissa Vazquez, and Raven Riordan for their work in manual coding tweets for this study, as well as Lijuan Cao who assisted with conceptualization and early data collection. The authors have no conflicts of interest to report.

# Appendix

## APPENDIX I

### Keyword lists used to extract vaping-related tweets:

lung, injury, health, illness, popcorn, epidemic, disease, die, dying, breathe, breathing, cough, coughing, trouble, problem, sick, safer, addicted, pneumonia, hospital, healthy, healthcare, ecigs, ecigarette, electronic, vaper, vapor, vapers, quit, quitting, smoke, smoking, safe, better, cigarette, tobacco, lungs, injuries, illnesses, popcornlung, death, deaths, evali, respiratory, fatal, case, cases, crisis, danger, dangers, dangerous, concern, concerns, cause, causes, severe, threatening, doctor, risk, addiction, addictive, addicting.

Above words were combined with the words ‘vape’ and ‘vaping’.

## APPENDIX II

| Manual Codebook Details | Instructions |
|---|---|
| Positive Sentiment toward Vaping | Choose one |
| Negative Sentiment toward Vaping | |
| Neutral/Unclear Sentiment toward Vaping | |
| Teens | Present/Not Present |
| Mentions Marijuana/Weed or Related Unregulated Product (including THC, CBD) | |
| **Mentions Negative Health Outcomes (NHO)** | Choose Parent Code, Subcodes as Applicable |
| Mentions Illness (Respiratory or other) as NHO | |
| Mentions Hospitalization/Death as NHO | |
| Mentions Dependence/Addiction as NHO | |
| **Mentions Positive Health Outcomes** | |
| Mentions Cessation or reduction in combustible smoking (tobacco or other substance) as PHO | |
| **Mentions Other/Neutral Health Outcomes** | |
| No Clinical Codes | Choose if No Codes Marked Above |

## APPENDIX III

**Keywords Manually Grouped for Negative Health Outcomes:** death, illness, lung, die, health, kill, issue, dangerous, disease, risk, crisis, danger, sick, doctor, cancer, warn, severe, injury, epidemic, addiction, damage, deadly, addict, harm, respiratory, harmful, patient, medical, hospitalize, dead, hurt, hospital, cough, breathe, addictive, emergency, breathing, ill, suffer, symptom, breath, toxic, sicken, panic, pulmonary, infection, physician, diagnose, popcorn, pain, collapse, sickness, oxygen, copd, asthma, abuse

**Keywords Manually Grouped for Marijuana/Unregulated Products:** thc, oil, black_market, illegal, cartridge, market, cannabis, chemical, drug, high, vitamin, hit, weed, street, cbd, pot, substance, unregulated, unregulated, counterfeit, illegally, bootleg, cannabi, contaminate, dispensary, stonerlife

**Keywords Manually Grouped for Teens/Youth/Adolescents:** kid, teen, child, young, youth, student, teenager, minor, underage, high_schooler

## References


[1]. World Health Organization. WHO report on the global tobacco epidemic, 2019: Offer help to quit tobacco use. World Health Organization; 2019.

[2]. Creamer MR, Wang TW, Babb S, et al. Tobacco product use and cessation indicators among adults—United States, 2018. Morbidity and Mortality Weekly Report. 2019;68(45):1013.

[3]. Wang TW, Neff LJ, Park-Lee E, Ren C, Cullen KA, King BA. E-cigarette use among middle and high school students—United States, 2020. Morbidity and Mortality Weekly Report. 2020;69(37):1310. [PubMed: 32941408]

[4]. Miech R, Johnston L, O'Malley PM, Bachman JG, Patrick ME. Trends in adolescent vaping, 2017–2019. New England Journal of Medicine. 2019;381(15):1490–1491.

[5]. Medical News Today. Are e-cigarettes a safe alternative to smoking? 2020; https://www.medicalnewstoday.com/articles/216550.

[6]. Delnevo CD, Giovenco DP, Steinberg MB, et al. Patterns of electronic cigarette use among adults in the United States. Nicotine & Tobacco Research. 2016;18(5):715–719. [PubMed: 26525063]

[7]. Krishnasamy VP, Hallowell BD, Ko JY, et al. Update: characteristics of a nationwide outbreak of e-cigarette, or vaping, product use–associated lung injury—United States, August 2019–January 2020. Morbidity and Mortality Weekly Report. 2020;69(3):90. [PubMed: 31971931]

[8]. King BA, Jones CM, Baldwin GT, Briss PA. The EVALI and youth vaping epidemics—implications for public health. New England Journal of Medicine. 2020;382(8):689–691.

[9]. Gordon T, Fine J. Cornering the suspects in vaping-associated EVALI. Mass Medical Soc; 2020.

[10]. Pray IW, Atti SK, Tomasallo C, Meiman JG. E-cigarette, or Vaping, Product Use-Associated Lung Injury Among Clusters of Patients Reporting Shared Product Use-Wisconsin, 2019. Morbidity and Mortality Weekly Report 2020;69:236–240. [PubMed: 32134907]

[11]. Leas EC, Nobles AL, Caputi TL, Dredze M, Zhu SH, Cohen JE, Ayers JW. News coverage of the E-cigarette, or Vaping, product use Associated Lung Injury (EVALI) outbreak and internet searches for vaping cessation. Tobacco Control. 2020 10 13.

[12]. Chary M, Genes N, Giraud-Carrier C, Hanson C, Nelson LS, Manini AF. Epidemiology from tweets: estimating misuse of prescription opioids in the USA from social media. Journal of Medical Toxicology. 2017;13(4):278–286. [PubMed: 28831738]

[13]. Nguyen T, Larsen ME, O'Dea B, Phung D, Venkatesh S, Christensen H. Estimation of the prevalence of adverse drug reactions from social media. International Journal of Medical Informatics. 2017;102:130–137. [PubMed: 28495341]

[14]. Cavazos-Rehg PA, Sowles SJ, Krauss MJ, Agbonavbare V, Grucza R, Bierut L. A content analysis of tweets about high-potency marijuana. Drug and Alcohol Dependence. 2016;166:100–108. [PubMed: 27402550]

[15]. D'Agostino AR, Optican AR, Sowles SJ, Krauss MJ, Lee KE, Cavazos-Rehg PA. Social networking online to recover from opioid use disorder: a study of community interactions. Drug and Alcohol Dependence. 2017;181:5–10. [PubMed: 29024875]

[16]. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. American Journal of Public Health. 2017;107(1):e1–e8.

[17]. Karami A, Dahl AA, Turner-McGrievy G, Kharrazi H, Shaw G Jr. Characterizing diabetes, diet, exercise, and obesity comments on Twitter. International Journal of Information Management. 2018;38(1):1–6.

[18]. McClellan C, Ali MM, Mutter R, Kroutil L, Landwehr J. Using social media to monitor mental health discussions-evidence from Twitter. Journal of the American Medical Informatics Association. 2017;24(3):496–502. [PubMed: 27707822]

[19]. Meng H-W, Kath S, Li D, Nguyen QC. National substance use patterns on Twitter. PloS One. 2017;12(11):e0187691. [PubMed: 29107961]

[20]. Visweswaran S, Colditz JB, O'Halloran P, et al. Machine learning classifiers for Twitter surveillance of vaping: Comparative machine learning study. Journal of Medical Internet Research. 2020;22(8):e17478. [PubMed: 32784184]

[21]. Dulam RVS, Murthy M, Luo J. Seeing through the smoke: a world-wide comparative study of e-cigarette flavors, brands and markets using data from Reddit and Twitter. arXiv preprint arXiv:200201575. 2020.

[22]. Szlyk H, Deng J, Xu C, Krauss MJ, Cavazos-Rehg PA. Leveraging social media to explore the barriers to treatment among individuals with depressive symptoms. Depression and Anxiety. 2020;37(5):458–465. [PubMed: 31943530]

[23]. Guiñazú MF, Cortés V, Ibáñez CF, Velásquez JD. Employing online social networks in precision-medicine approach using information fusion predictive model to improve substance use surveillance: A lesson from Twitter and marijuana consumption. Information Fusion. 2020;55:150–163.

[24]. Myslín M, Zhu S-H, Chapman W, Conway M. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. Journal of Medical Internet Research. 2013;15(8):e174. [PubMed: 23989137]

[25]. Mackey TK, Li J, Purushothaman V, et al. Big Data, Natural Language Processing, and Deep Learning to Detect and Characterize Illicit COVID-19 Product Sales: Infoveillance Study on Twitter and Instagram. JMIR Public Health and Surveillance. 2020;6(3):e20794. [PubMed: 32750006]

[26]. Ahmed W, Marin-Gomez X, Vidal-Alaball J. Contextualising the 2019 e-cigarette health scare: insights from Twitter. International Journal of Environmental Research and Public Health. 2020;17(7):2236.

[27]. Hoffman B, Colditz J, Sidani J, et al. Correlation of Twitter Data to Reported Cases of E-cigarette or Vaping Product Use-Associated Lung Injury (EVALI). A22. Current Topics in Tobacco Cessation and E-cigarettes: American Thoracic Society; 2020:A1071–A1071.

[28]. Zolnoori M, Balls-Berry JE, Brockman TA, Patten CA, Huang M, & Yao L (2019). A systematic framework for analyzing patient-generated narrative data: Protocol for a content analysis. JMIR Research Protocols, 8(8), e13914.

[29]. Python Programming Index. GetOldTweets3 0.0.11. 2019; https://pypi.org/project/GetOldTweets3/.

[30]. Chandrasekaran R, Mehta V, Valkunde T, Moustakas E. Topics, Trends, and Sentiments of Tweets About the COVID-19 Pandemic: Temporal Infoveillance Study. Journal of Medical Internet Research. 2020;22(10):e22624. [PubMed: 33006937]

[31]. Cambria E, Das D, Bandyopadhyay S, Feraco A. A practical guide to sentiment analysis. Springer; 2017.

[32]. Huang M, ElTayeby O, Zolnoori M, Yao L. Public opinions toward diseases: Infodemiological study on news media data. Journal of Medical Internet Research. 2018;20(5):e10047. [PubMed: 29739741]

[33]. TextBlob. Simplified Text Processing. 2020; https://textblob.readthedocs.io/en/dev/.

[34]. Madhu S An approach to analyze suicidal tendency in blogs and tweets using Sentiment Analysis. Int J Sci Res Comput Sci Eng. 2018;6(4):34–36.

[35]. Hermansyah R, Sarno R. Sentiment Analysis about Product and Service Evaluation of PT Telekomunikasi Indonesia Tbk from Tweets Using TextBlob, Naive Bayes & K-NN Method. Paper presented at: 2020 International Seminar on Application for Technology of Information and Communication (iSemantic)2020.

[36]. GitHub. Pattern. 2021; https://github.com/clips/pattern.

[37]. Lienemann BA, Unger JB, Cruz TB, Chu K-H. Methods for coding tobacco-related Twitter data: a systematic review. Jounral of Medical Internet Ressearch. 2017;19(3):e91.

[38]. Elo S, Kyngäs H. The qualitative content analysis process. Journal of Advanced Nursing. 2008;62(1):107–115. [PubMed: 18352969]

[39]. van der Tempel J, Noormohamed A, Schwartz R, Norman C, Malas M, Zawertailo L. Vape, quit, tweet? Electronic cigarettes and smoking cessation on Twitter. International Journal of Public Health. 2016;61(2):249–256. [PubMed: 26841895]

[40]. Martinez LS, Hughes S, Walsh-Buhi ER, 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. 2018;23(6):550–562. [PubMed: 29979920]

[41]. 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. 2019;21(2):e11953. [PubMed: 30720440]

[42]. Miech R, Patrick ME, O'Malley PM, Johnston LD. What are kids vaping? Results from a national survey of US adolescents. Tobacco Control. 2017;26(4):386–391. [PubMed: 27562412]

[43]. Budney AJ, Sargent JD, Lee DC. Vaping cannabis (marijuana): parallel concerns to e-cigs? Addiction. 2015;110(11):1699–1704. [PubMed: 26264448]

[44]. Sowles SJ, Krauss MJ, Connolly S, Cavazos-Rehg PA. Peer reviewed: A content analysis of vaping advertisements on Twitter, November 2014. Preventing Chronic Disease. 2016;13.

[45]. McHugh ML. Interrater reliability: the kappa statistic. Biochemia Medica. 2012 ;22(3):276–82. [PubMed: 23092060]

[46]. Syed M, Nelson SC. Guidelines for establishing reliability when coding narrative data. Emerging Adulthood. 2015;3(6):375–387.

[47]. Lee JL, DeCamp M, Dredze M, Chisolm MS, Berger ZD. What are health-related users tweeting? A qualitative content analysis of health-related users and their messages on twitter. Journal of Medical Internet Ressearch. 2014;16(10):e237.

[48]. Schmiege SJ, Klein WM, Bryan AD. The effect of peer comparison information in the context of expert recommendations on risk perceptions and subsequent behavior. European Journal of Social Psychology. 2010;40(5):746–759.

[49]. Duong HT, Liu J. Vaping in the news: the influence of news exposure on perceived e-cigarette use norms. American journal of health education. 2019;50(1):25–39.

Author Manuscript

Author Manuscript

Author Manuscript

**Highlights:**

- Positive sentiment toward vaping outweighed negative sentiment across 2019
- Mentions of lung injury and negative health outcomes peaked in September 2019
- Manual coding identified marijuana products and youth as primary themes
- Twitter data can serve as a surveillance tool to identify changes in risk over time
- Themes identified on Twitter can inform public health interventions and outreach

CCS CONCEPTS • content analysis • sentiment analysis • text mining

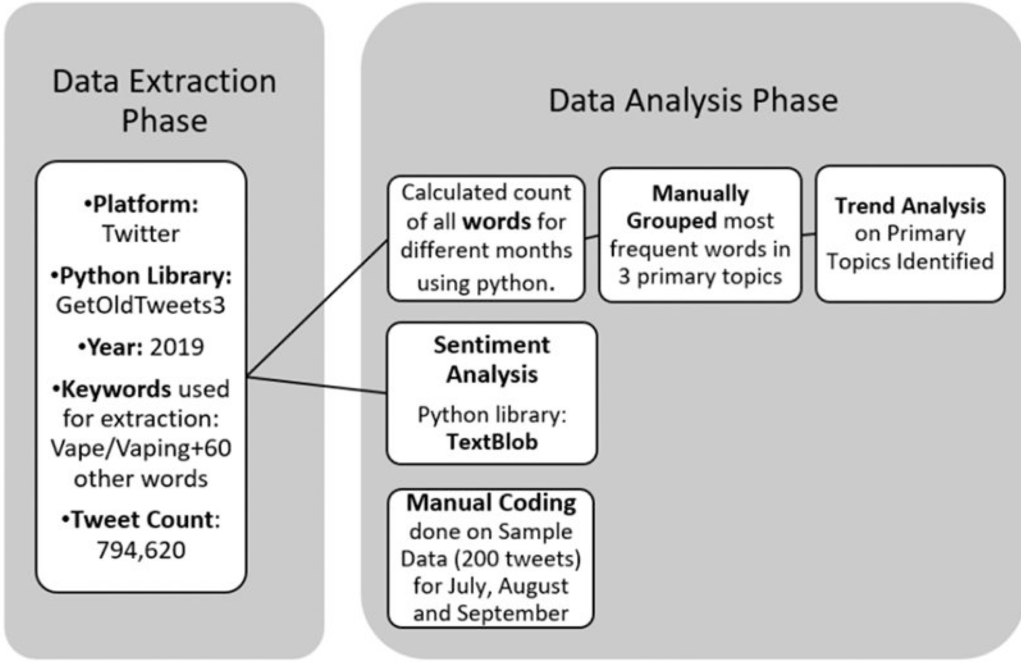


**Figure 1.**
Methods for Data Extraction and Data Analysis

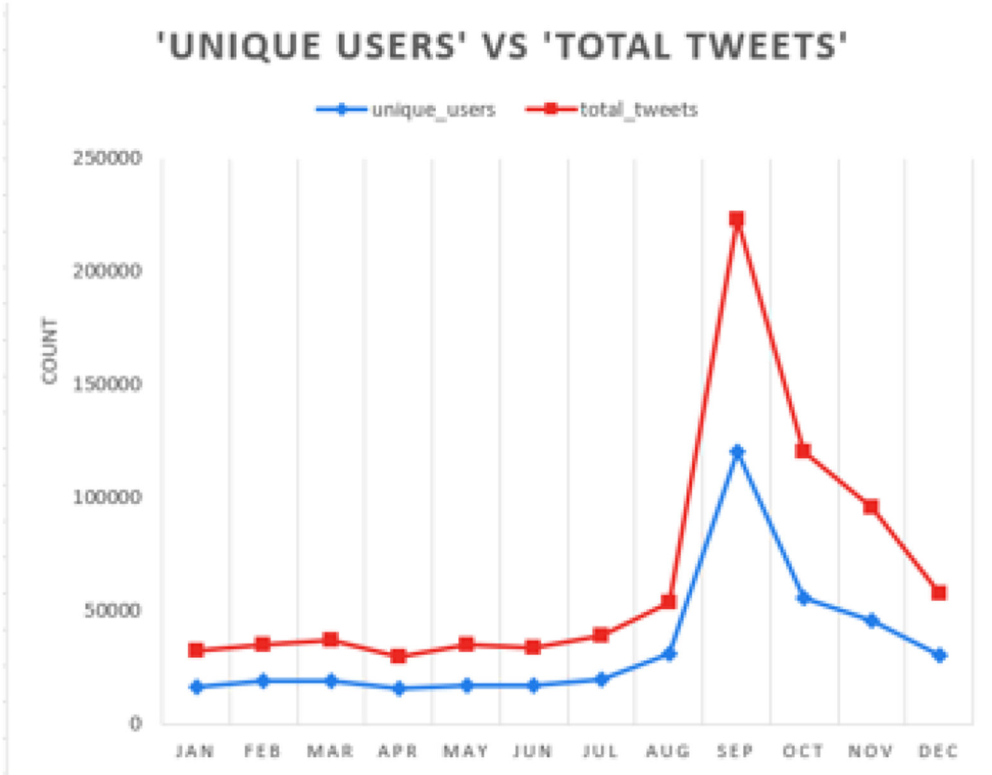


**Figure 2.**
Total Tweets and Tweets from Unique Users in 2019

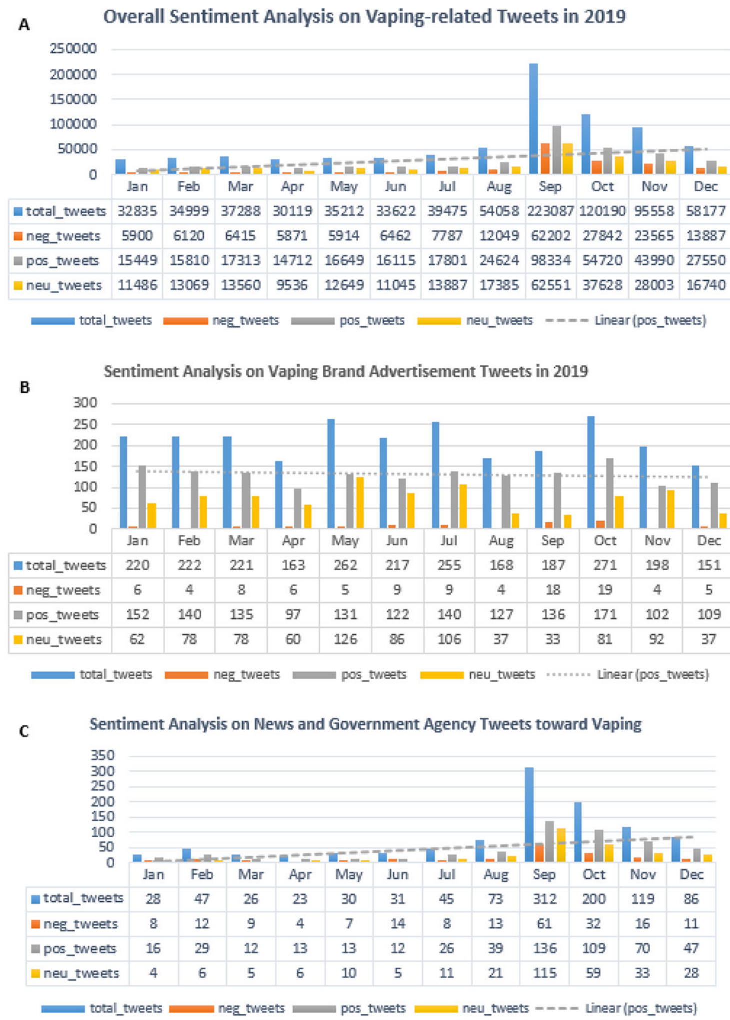


**Figure 3.**

Positive and Negative Sentiment Toward Vaping in 2019 Tweets

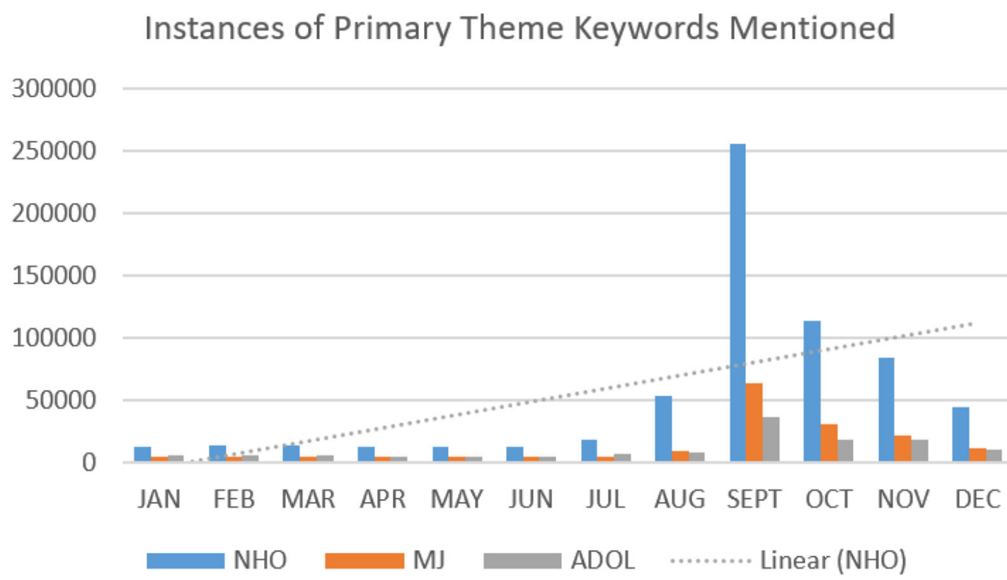


**Figure 4.**

Instances of Vaping-related keywords mentioned by primary themes across 2019


**Figure 5.**
Word Cloud for vaping-related tweets in July, August, and September 2019

**Table 1:**

Sentiments for Vaping Products Mentioned in 2019 Tweets

| Month | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sept | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | **N=188** | **N=192** | **N=198** | | | |
| Negative Sentiment | | | | | | | 84 | 93 | 69 | | | |
| Example Tweets: | •*(Jul) "Teenage vaping is an epidemic, and now youth are facing addiction to nicotine and some are seeing health issues after becoming dependent on e-cigarettes. https://wapo.st/2yjFEcC"*<br>• *(Jul) "Do you still want to take that hit? If you vape, you're four times more likely to start smoking. #JUULisOVER"*<br>• *(Aug) "I've heard from several smokers, vaping is more harmful than smoking.*<br>• *(Aug) "Yeah, I knew this was coming. What about vapor bubbles accumulated in lungs? DUH.* DIDN'T IT OCCUR TO PEOPLE THAT VAPING IS JUST AS DUMB AS SMOKING."<br>• (Sept) *"LMAO I don't support the vaping ban but you all are fools if you really believe vaping is a better alternative to smoking"*<br>• *(Sept) "At least smoking kills you over 40 years, evidently vaping kills you in four, I'll stick with my cigarettes"* | | | | | | | | | | | |
| Positive Sentiment | | | | | | | 82 | 53 | 95 | | | |
| Example Tweets: | • *(Jul) There are no circumstances in which vaping is more dangerous than smoking. None at all. Vaping is not a public health hazard because it is >95% safer than smoking and can replace smoking for many addicted to nicotine. It creates significant public health benefits."*<br>• *(Jul) "The benefit of vaping and other non-combustibles is not just that they reduce smoking, but that they do it by user choice and without relying on coercive measures like taxes or stigma on users. The 'how' matters as much as the 'what' - though the 'why' to quit is the same."*<br>• *(Aug) "…but if you smoke and are trying to quit, #vaping may be an excellent #harm reduction strategy—in which case: vape!"*<br>• (Aug) *"Vaping is: The most effective smoking cessation aid ever invented. 2-3 times more effective than patches, gums, sprays. It is no more addictive or harmful than patches. #VapingSavesLives STUDIES: https://www.nejm.org/doi/full/10.1056/NEJMoa1808779"*<br>• *(Sept) "Even if they take all vaping products away, I will continue to stand for the truth about vaping and use whatever resources I have to share the truth until vaping and vapor products are supported as the life saving cessation products they are."*<br>• *(Sept) "Your lack of empathy toward 3.3 million adult smokers who have quit by vaping, 5 million more adult vapers on the journey to quit, and 36 million US smokers, is awful. Talk to an adult vaper. They are not corporations. They are PEOPLE."* | | | | | | | | | | | |
| Neutral/ Unclear Sentiment | | | | | | | 22 | 47 | 34 | | | |

**Table 2:**

Health Outcomes Due to Vaping Mentioned in 2019 Tweets

| Keyword | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sept | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Human Coding** | | | | | | | **N=188** | **N=192** | **N=198** | | | |
| Mentions Negative Health Outcomes | | | | | | | 74 | 38 | 118 | | | |
| Mentions EVALI | | | | | | | 38 | 9 | 58 | | | |
| Mentions Hospitalization/ Death | | | | | | | 0 | 6 | 51 | | | |
| Mentions Dependence/ Addiction | | | | | | | 13 | 11 | 6 | | | |
| Mentions Positive Health Outcomes | | | | | | | 57 | 46 | 31 | | | |
| Vaping to Quit Combustible Smoking | | | | | | | 47 | 41 | 30 | | | |
| Example Tweets: | *Negative Health Outcomes*<br>• *(Jul) "Doctors suspect vaping is behind severe lung damage found in eight Wisconsin teenagers http://hill.cm/binmXMs"*<br>• *(Jul) "It's addictive. It damages lungs. It limits brain development. It's a waste of money. It's not biodegradable. It can cause cancer. Please keep your kids away from it and… DON'T VAPE #ecig #vapes"*<br>• *(Aug) "Quit vaping people!!! You all will be collapsed lungs like me!!"*<br>• *(Sept) "In case you have not realized, vaping is not safer than smoking. YOU ARE BEING LIED TO!"*<br>• (Sept) "Do *any of you think that all this anti vaping shit lately is a scam by big tobacco to get people already addicted to nicotine to start smoking cigs?"*<br>• *(Sept) "I know vaping fans will tell you how much healthier it is compared to smoking, but something is seriously wrong. The suggestion is that you stop immediately until they can determine why people are dying."*<br>Positive Health Outcomes<br>• *(Jul) "Vaping is saving lives by helping people quit smoking traditional cigarettes"*<br>• *(Jul) "Evidence on the Safety of E-cigarette #eCig #eCigarette #vape #vapingSavesLives #TobaccoHarmReduction"*<br>• *(Aug) "My dad was a two pack a day smoker. Vaping is the only thing that worked for him to quit. Now he only vapes a few times each day. I consider that a miracle"*<br>• *(Aug) "Vaping is the only thing that works for me. I still have dreams of buying cigarettes six years later."*<br>• (Sept) *"I spent 10 years of my life smoking cigarettes. 10 years of addiction that I knew would kill me. I knew my kids would watch me suffer. I knew I'd leave them too early…. 2 years ago, vaping saved my life. I say we #RiseAndVape."*<br>• (Sept) *"I for one have no intention of stopping people from getting off of deadly cigarettes! I am not above civil disobedience & posting bail! I Cannot Allow Vaping To Die! If I have to make own mods & my own liquid and teach others how I will do it! I will not stop"* | | | | | | | | | | | | |

**Table 3:**

Marijuana and Unregulated Products Mentioned in 2019 Tweets

| Keyword | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sept | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | **N=188** | **N=192** | **N=198** | | | |
| Mentions Marijuana, Cannabis, CBD/THC | | | | | | | 8 | 31 | 38 | | | |
| Example Tweets: | • *(Jul) "The "vaping" hospitalizations seem entirely due to illicit drugs. This is *not the same thing* as vaping. Any journal or health professional who equates the two is dangerously irresponsible. & These illnesses *will happen more* as legal vaping is banned or restricted."*<br>• *(Jul) "They tell us 3.6 million teens have tried vaping, but now we are supposed to believe that 8 kids got lung damage and hospitalized from vaping regular e-cigs? No, this was black market stuff and could have contained other drugs and chemicals."*<br>• (Aug) *"vape usually has nicotine which is addictive so just know that before you start. weed is usually fine but make sure if you are smoking a joint that it isn't laced with another drug cause that's…. bad lol"*<br>• (Sept) *"all these vaping problems aren't from vapes they are from dab pens but nobody wants to tell that to their parents or doctors"*<br>• *(Sept) "Most of these vaping illnesses are coming from black-market THC products, but the solution is to make *legal* vaping products *less* available? Driving more people to go to the black market? How has this irrational story become the prevalent one?"*<br>• *(Sept) "It wasn't even vaping stuff they got sick on, but black market THC."* | | | | | | | | | | | |

**Table 4:**

Age Group and Communities Mentioned in 2019 Vape-related Tweets

| Month | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sept | Oct | Nov | Dec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | **N=188** | **N=192** | **N=198** | | | |
| Mentions Teens, Adolescents, Youth | | | | | | | 75 | 21 | 36 | | | |

Example Tweets:

- *(Jul) "The vape industry is playing kids the same exact way the cigarette companies did years ago."*
- *(Jul) "Officials say it is an "epidemic of youth use," and estimate that the number of high school students who vape has risen to about 3 million students. Many schools have changed their response from strict discipline to education and treatment."*
- *(Aug) "It's obvious why some teens think vaping is harmless. Because they see adults lie to them over & over and try to scare them away from vaping. People do it on a regularly. For example, you cited a bogus heart attack study."*
- *(Sept) "Data shows a high use of flavored vapes among youth. Parents have a responsibility to understand the dangers that come from vaping. Our organization supports the removal of vapes from stores until they are approved by @US_FDA. #BeBest"*
- *(Sept) "Can we relax with all the vaping shit? Leave us alone to kill ourselves. It is already against the law to smoke for teens under 18. Vaping companies can't help that bad parents let their kids vape. #hardpillstoswallow"*