

Original Paper

E-Cigarette Social Media Messages: A Text Mining Analysis of Marketing and Consumer Conversations on Twitter

Allison J Lazard^{1,2*}, PhD; Adam J Saffer^{1*}, PhD; Gary B Wilcox^{3*}, PhD; Arnold DongWoo Chung^{3*}, MS; Michael S Mackert^{3*}, PhD; Jay M Bernhardt^{2*}, MPH, PhD

¹School of Media and Journalism, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States

²Center for Health Communication, Moody College of Communication, The University of Texas at Austin, Austin, TX, United States

³Center for Health Communication, Stan Richards School of Advertising and Public Relations, The University of Texas at Austin, Austin, TX, United States

* all authors contributed equally

Corresponding Author:

Allison J Lazard, PhD

Center for Health Communication

Moody College of Communication

The University of Texas at Austin

CMB 3.120

Austin, TX

United States

Phone: 1 317 331 7807

Fax: 1 919 962 0620

Email: lazard@unc.edu

Abstract

Background: As the use of electronic cigarettes (e-cigarettes) rises, social media likely influences public awareness and perception of this emerging tobacco product.

Objective: This study examined the public conversation on Twitter to determine overarching themes and insights for trending topics from commercial and consumer users.

Methods: Text mining uncovered key patterns and important topics for e-cigarettes on Twitter. SAS Text Miner 12.1 software (SAS Institute Inc) was used for descriptive text mining to reveal the primary topics from tweets collected from March 24, 2015, to July 3, 2015, using a Python script in conjunction with Twitter's streaming application programming interface. A total of 18 keywords related to e-cigarettes were used and resulted in a total of 872,544 tweets that were sorted into overarching themes through a text topic node for tweets (126,127) and retweets (114,451) that represented more than 1% of the conversation.

Results: While some of the final themes were marketing-focused, many topics represented diverse proponent and user conversations that included discussion of policies, personal experiences, and the differentiation of e-cigarettes from traditional tobacco, often by pointing to the lack of evidence for the harm or risks of e-cigarettes or taking the position that e-cigarettes should be promoted as smoking cessation devices.

Conclusions: These findings reveal that unique, large-scale public conversations are occurring on Twitter alongside e-cigarette advertising and promotion. Proponents and users are turning to social media to share knowledge, experience, and questions about e-cigarette use. Future research should focus on these unique conversations to understand how they influence attitudes towards and use of e-cigarettes.

(*JMIR Public Health Surveill* 2016;2(2):e171) doi: [10.2196/publichealth.6551](https://doi.org/10.2196/publichealth.6551)

KEYWORDS

e-cigarettes; social media; tweet; Internet

Introduction

Since the introduction of electronic nicotine delivery systems or electronic cigarettes (e-cigarettes) less than a decade ago, awareness and use of these products has risen dramatically [1-5]. E-cigarettes are battery-powered devices that deliver varying amounts of nicotine, propylene glycol, and flavorings, among other things, through an aerosolized liquid solution (commonly referred to as vaping) [6]. Introduced as a substitution for cigarettes, e-cigarette use has grown rapidly despite little knowledge about the short- and long-term health risks or the broad impact on public health given high levels of dual use of e-cigarettes and conventional cigarettes [6-9]. Concerns for increased e-cigarette use are not limited to dual-users. Lifetime or ever use of e-cigarettes has increased exponentially among adults, and more importantly, those most susceptible to addiction, adolescents and young adults [10-13].

As the prevalence of e-cigarettes increases, researchers must understand how individuals acquire knowledge about these products and consider how the different sources of information might influence decisions of use. Prior research has consistently found that individuals are exposed to, search for, and share information about e-cigarettes on social media platforms like Twitter and Facebook, among others [14-17]. Interestingly, e-cigarette users are “more likely to be exposed to e-cigarette information via TV viewing sites (eg, Hulu), Twitter, Facebook, email, and Internet search engines” than nonusers [15]. Therefore, it is important to understand what information e-cigarette users and potential users are exposed to on social media.

Twitter is particularly unique among social media platforms because users can broadcast their messages in the form of tweets and retweets to large audiences [18-20], often sharing information that may contribute to conceptualizations of new health products and phenomena [17,21,22]. Broadly speaking, there are 3 types of users tweeting about e-cigarettes: individual users representative of the public, e-cigarette proponents, and commercial marketers. Individual users are average people who tweet or retweet messages about e-cigarettes that can directly reach their followers or indirectly reach other users following hashtags or key terms in a tweet. Proponents are a diverse group of representatives of e-cigarette organizations, vaping advocates, and those who identify as e-cigarette users in their online profiles [23]. Notably, proponents have been shown to tweet e-cigarette content 5 times as often as individual users, much of which is positive toward e-cigarette use [23]. Commercial marketers, such as e-cigarette retailers and manufacturers, are those who use Twitter to promote and advertise their products. Individuals can be exposed to information about e-cigarettes from individual, proponent, or commercial sources.

Many studies have found that exposure to e-cigarette messages on Twitter has increased significantly in recent years [24-26]. As of 2016, 1 study found tweets about e-cigarettes have increased 5 times since 2012 [26]. Further investigations into the sources of e-cigarette tweets indicate that most tweets originate from commercial users as marketing or advertising promotional messages. According to Huang and colleagues

[27], 90% of e-cigarette tweets are from commercial users, and Kim et al [24] reported that 93.43% of tweets were for advertising.

However, the data collected for these studies may have had some confounding variables. Huang et al [27] noted that “the time frame during which our data were collected was just prior to the launch of major e-cigarette TV marketing.” Kim et al [24] analyzed data from the same time frame. This may have led to the finding that most e-cigarette tweets come from commercial users. Some evidence is emerging that challenges previously reported prevalence of e-cigarette tweets from commercial users and for advertising. Dai and Hao [28], who analyzed individual users’ organic tweets about e-cigarettes, found that 10.8% supported, 17.7% were against, and 19.4% were neutral toward the use of e-cigarettes [28], indicating there is variance in the public conversation about e-cigarettes on Twitter.

An additional confound is the types of users tweeting about e-cigarettes. To understand individual exposures, it is necessary to reveal the type of accounts generating e-cigarette-related tweets. Researchers must first recognize the prevalence of automated social bot accounts (also referred to as cyborgs) run by third parties to influence and promote e-cigarettes and related products but resembling average users [29,30]. Tweets from bot accounts are reducing the signal-to-noise ratio where individuals’ e-cigarette tweets are far fewer.

For instance, Clark et al [31] looked at a 10% sample of e-cigarette tweets from “Twitter’s garden hose” over a 2-year period and found that 80% of tweets were “automated and promotional.” Thus, it is necessary to increase the signal-to-noise ratio to accurately examine e-cigarette-related Twitter discussions among individuals, and denoising techniques allow researchers to do so [29]. However, Allem and Ferrara [29] cautioned researchers against using “crude” and “blunt” techniques such as removing tweets with links solely relying on community detection or methods solely relying on innocent-by-association paradigms. While denoising techniques are still emerging, this work highlights an important need for awareness of the signal-to-noise issue for surveillance of social media data.

Searching for the signal, themes of e-cigarette tweets have been studied to reveal salient topics and sentiment of publically accessible conversations on Twitter. Using a priori themes, Kavuluru and Sabbir [23] detected largely positive sentiment for themes of flavors, harm reduction, smoke-free aspects, and smoking cessation. Similarly and on a small scale relative to the current data-mining abilities, Myslin and colleagues [32] used a combination of a priori themes and iterative machine learning to identify a preponderance of first- or second-hand experience tweets about tobacco-related themes of hookah, cessation, and pleasure. E-cigarette content was not among the top themes, likely given the newness of these products in 2011-2012; however, a trend of positive sentiment with e-cigarette content was detected (contrary to more negative sentiment with smoking tweets) [32]. Cole-Lewis and colleagues [33,34], also using themes developed from previous research, revealed that advertising/promotion, policy/government, and

health/safety are among the most dominant themes. Much of this content had a positive sentiment skewing favorably toward e-cigarettes [33]. In a purposeful sample specific for smoking cessation, van der Tempel and colleagues [26] found similar dominant themes—marketing, news, personal experiences—from an a priori theme list. Together, the trends in the literature indicate that as e-cigarette content becomes more popular, the majority of this content is positive promotion. However, missing still is large-scale, inductive analysis of the topics and themes of the tweets from all categories of users. Our study presents the topics and themes of tweets about e-cigarettes from individual, proponent, and commercial users.

An accurate understanding of the types of users tweeting, as well as what they are tweeting, about e-cigarettes can provide a better understanding of what individuals are being exposed to on social media. Looking specifically at the nonadvertising tweets in their data, Kim et al [24] found that organic conversations are occurring online among individuals about e-cigarettes. This is particularly important because organic conversations may affect individual exposure. That is, if individuals are exposed to an e-cigarette tweet from another individual they know, they may be more easily persuaded given the relational closeness and potentially stronger levels of source credibility [35,36]. Thus, as the use of e-cigarettes continues to rise, public awareness of these products is likely shaped by the proliferation of messages shared and reshared on social media. To understand this media landscape, this study used a textual analysis method to examine the public conversation on Twitter and determine overarching themes and trending topics from commercial and consumer contributors.

Methods

Text Mining and Data Acquisition

This study used a text-mining approach to uncover key patterns and relationships within unstructured data to understand and evaluate information important to the audience. Text mining is the term used to describe either a single process or a collection of processes in which software tools actively engage in the “discovery of new, previously unknown information by automatically extracting information from different written (or text) sources” [37]. Text mining provides an opportunity to uncover key patterns and relationships within both structured and unstructured data and allows researchers to more easily understand and evaluate information important to the audience.

In the area of public health, text mining of social media has been used to detect and track disease outbreaks and estimate the level of public knowledge regarding health issues [38]. Twitter was selected for data collection due to its popularity as a microblog as well as the active nature of its users in sending messages to create conversations regarding the use of new products and related social issues [18,21]. The central challenge of text mining is the analysis of unstructured data in order to extract meaningful associations, trends, and patterns in large amounts of text. The increasing availability and magnitude of unstructured digital data available in social media such as Twitter, Facebook, blogs, and other online environments offers

new opportunities for researchers to investigate social, cultural, and health issues.

The methodology and workflow for this study depended upon a combination of human and technological analysis of the Twitter messages and employed 5 steps. First, the search term and time period were determined. Next, Twitter’s streaming application programming interface (API) was used to acquire the data. Third, researchers cleaned the data and removed duplication and unrecognized characters. Fourth, data were processed using text-mining software and fifth, the findings were interpreted.

Data Collection

Twitter was selected for data collection due to its popularity as a microblogging service and the active nature of its users in sending messages regarding news and social issues, including health-related issues. Tweets were collected from March 24, 2015, to July 3, 2015, using a Python script in conjunction with Twitter’s streaming API. A total of 18 keywords related to e-cigarettes, vaping, and e-cigarette brands were used (ecigs, ecigarettes, e-cigarettes, electronic cigarettes, vaping, vapeystick, ehookah, ejuice, Blu eCigs, E-Swisher, Ezsmoker, Fin, NJOY/NJOY, Smoke Assist, V2 Cigs, MarkTen, Vuse, and Tryst) which resulted in a total of 872,544 tweets and retweets. Tweets and retweets were separated into 2 files for the analysis, a tweet file containing 546,651 entries and a retweet file containing 325,893 entries.

Data Analysis

Analysis of the textual content of the tweets was conducted using SAS Text Miner version 12.1 (SAS Institute Inc). SAS Text Miner allowed the researchers to parse and extract information from text, filter, and assemble documents into related topics allowing the researchers to discover topics and understand the data. This software was used for descriptive text-mining purposes to uncover the primary topics that were being discussed during the 100 days examined.

Following the collection of the data, the initial step was to extract, clean, and create a dictionary of words from the data using a natural language processor. A node process flow was created in SAS Enterprise Miner Workstation version 12.1 (SAS Institute Inc). It began with a Text Parsing node where each tweet is divided into tokens (terms). Specifically, this includes identifying sentences, determining parts of speech, and stemming words. Words were spell-checked and parsed to identify entities and remove stop words. The identified tokens or terms were listed in a “term by frequency” matrix via a text transformation of the numerical representation of the text using linear algebra-based priority models. To ensure that words that contribute little were not overly emphasized by the algorithm, the following parts of speech were ignored: auxiliary verbs, conjunctions, determiners, interjections, participles, prepositions, and pronouns.

Next in the Text Filter node, terms that appeared in fewer than 10 messages were ignored. The data were filtered using Entropy as the term weight and Log as the frequency weighting. The term filtering alters the term-by-document matrix, which contains the frequency of the occurrence of the term in the

documents as the value of each cell. From this frequency matrix, a weighted term-by-document matrix was generated using software-driven term-weighting techniques. Within the Text Filter, the Filter Viewer was employed to visually inspect the individual terms. Unrecognizable symbols and letter groupings were manually excluded. Next, a check was made of the terms that were ignored to ascertain if any should be included in the analysis. A single author who had knowledge of the subject matter manually excluded irrelevant terms. Finally, the Text Filter node was used to reduce the total number of parsed terms, thereby eliminating extraneous information and retaining the most relevant parts of the text.

The Text Topic node was then employed to combine terms into topic groups. SAS Text Miner uses 2 types of clustering algorithms: expectation maximization (EM) and hierarchical clustering. EM clustering was used because it allows for and automatically selects between 2 versions of the algorithm—1 for small data files (standard) and 1 for larger (scaled) data files. Since there were over 800,000 tweets and retweets, the EM option was preferred. Options were selected within the software to create topic groups to include all topics that contained more

than 1% of the total tweets or retweets. Topics with less than 5467 tweets or 3259 retweets were excluded from the analysis given they represented less than 1% of the data.

Last, the Topic Viewer option in Text Topic was used to further refine and interpret the topic groups. Individual tweets and retweets were reviewed and used to create summaries of each topic group. One author evaluated the results by completing several different iterations of SAS Text Miner, comparing the different results, and selecting what appeared to be the optimum solution after careful inspection of the output. After visual examination of each topic list (9 tweet topics and 14 retweet topics), topics that clearly did not illustrate the main themes were removed to reduce noise. Individual review of the actual topics generated by the software was undertaken to further exclude topics that appeared from automated accounts. This was accomplished by individually reviewing the actual messages from each topic to produce the final grouping of topics for tweets (8 topics) and retweets (5 topics) with the description in [Tables 1 and 2](#). This process aided in noise reduction among the tweets as called for by Allem and Ferrar [29].

Table 1. E-cigarette tweets by topic.

	SAS Text Miner topic	n	Description of topic	Category
1	+E-cigarette, +smoke, ecigs, +quit, vaping	29,556	Pro-vaping discussion whether e-cigarettes are cessation devices or gateway products for youth to establish nicotine addictions	Proponent and individual user conversation
2	Ejuice, eliquid, vape, vapelife, ecig	24,064	Pro-vaping reactions to e-cigarette policies and questions about e-cigarette health risk claims	Proponent and individual user conversation
3	Vaping, vape, ecigs, vapelife, vapecommunity	21,555	News and updates from the vaping community, primarily through a vaping blog's daily updates	Proponent and individual user conversation
4	Vapeporn, +tree, vape, vaping, reddit	16,414	Vaping advocacy encouraging the uptake of vaping and tips for users	Proponent and individual user conversation
5	Electronic, +cigarette, +electronic cigarette, +employee, +relieve	12,694	Employees may use e-cigarettes at work as relief from smoking bans	Proponent and individual user conversation
6	+Juice, vapejuice, +vaporizer, +vapor, vape	9092	Promotion for e-juice	Marketing/advertising
7	+Well price, +vapour, +price, +good,	6860	Price promotion for e-cigarettes	Marketing/advertising
8	+Win, cigbuyer, vapegiveaway, +enter, +sampler	5892	Promotion for an e-liquid give-away	Marketing/advertising

Table 2. E-cigarette retweets by topic.

	SAS Text Miner topic	n	Description of topic	Category
1	Vapinxsomker, vaping, ecigs, +smoke, +e-cigarette	42,430	Discussion of policies banning e-cigarette use, e-cigarettes as (or not as) smoking cessation devices, and differentiation of e-cigarettes from traditional cigarettes	Proponent and individual user conversation
2	Vape, ejuice, ecigs, +win, vaping	33,767	Promotions for an e-juice give-away	Marketing/advertising
3	Amp, +win, +follow, +chance, http	18,305	Promotions for a chance to win an e-cigarette	Marketing/advertising
4	Vape, realdonnadevane, vaping, granny, ur	13,752	E-cigarette blogger suggests ways to “grow your #Vape business”	Marketing/advertising
5	+Starter, hookahcoals, http://t.co/5amkbytrp3, newyork, eastcoast	6197	Promotions for e-hookah starter kits	Marketing/advertising

Results

Overview

Of the 872,544 tweets and retweets captured from March to July 2015, 240,578 were included in the final topic groups that each represented more than 1% of the conversation on Twitter during this time period for tweets and retweets. These were divided among 126,127 tweets sorted into 8 unique topics, shown in Table 1, and 114,451 retweets in 5 topics, shown in Table 2. Each table shows the topics featuring the prevalent keywords as generated by SAS Text Miner, number of tweets per topic, a description of the topic, and whether the topic represents proponent and individual user conversations or marketing/advertising. All topics with diverse tweets not directly linked to marketing were labeled proponent and individual user conversations. Topics dominated by promotions or branded persuasive appeals were categorized as marketing/advertising; often a single, repeated tweet comprised the entire topic for marketing/advertising.

Tweet Topics

Of the 8 topics generated from the tweets, the top 5 topics, determined by number of tweets contained in the topic, included diverse proponent and individual user conversations. The most popular topic in the analysis represented a diverse public conversation that covered whether e-cigarettes are cessation devices or gateway products to get youth addicted to nicotine. This topic was dominated by pro-e-cigarette content and included comments that e-cigarettes may help people quit smoking, that the rise in e-cigarette use among adolescents may be deterring them from traditional tobacco experimentation, how e-cigarette bans may have unintended consequences, and claims and questions about whether the science showing the risk of e-cigarettes is flawed. Anti-e-cigarette tweets within this topic were limited but did include warnings for adolescent use.

The second most popular topic was a proponent and individual user conversation in reaction to e-cigarette bans and proposed taxes along with further questioning of whether there is evidence to support health risk claims about e-cigarette use. This topic represents a conversation around efforts to differentiate e-cigarettes from traditional tobacco products by pointing to the lack of evidence for the harm or risks of e-cigarettes along with the position that e-cigarettes should be promoted as smoking cessation devices, aside from the youth discussion above, by generally taking the stance that the use of e-cigarettes can save lives. Additionally, concerns expressed that products would be too expensive or unavailable because of regulation were also common.

The third topic from the tweets contained e-cigarette news and updates from proponents in the vaping community, although this community was dominated by one pro-e-cigarette news outlet. Many of these tweets were generated through the “share” option from the daily vaping news website and covered a range of topics that included coverage of policies (eg, bans, taxes), promotions from the organization, product reviews, tips for social and culture practices of vaping, and diverse articles that highlight e-cigarettes as cessation devices, along with the uncertainty of risk with e-cigarette use (eg, “Is a daily dose of

nicotine as benign as coffee?”). This topic also included notices of new products available for sale from a variety of distributors.

The fourth topic consisted of vaping advocacy comments from a variety of proponent and individual user angles. Tweets included tips for e-cigarette users, discussions of flavors, encouragement for expanding the social practices of vaping (via new people and new places), and information about the use of specific devices as well as using devices for marijuana consumption. The fifth proponent and individual user topic consisted of a discussion about how employees may use e-cigarettes as a way to find relief from smoking bans at workplaces, as well as providing information about e-cigarettes and vaping.

Topics 6 through 8, the least populated topics, contained marketing promotions from 3 unique vendors. Each of these topics contained a single repeated tweet or tweets with only slight variations. Topic 6 contained just over 9000 original tweets that promoted a single distributor’s e-juice in a variety of flavors, such as kettle corn, grape, vanilla, and menthol. Topic 7 was entirely the repetition of one price promotion tweet from a different distributor, merely mentioning they had the best prices. Topic 8 was similarly the repeat of a single tweet; the tweet was a give-away promotion for an e-liquid sampler from a third distributor. The sampler promotion contained four 30 mL bottles in flavors *gravel pit*, *lime cola*, *strawberry blonde*, and *trail mix*.

Retweet Topics

The 5 retweet topics consisted of 1 proponent and individual user conversation topic and 4 marketing/advertising topics. The most popular topic—the proponent and individual user conversation—contained references to policy bans for using e-cigarettes in public places and raising age restrictions for the legal purchase of tobacco products; comments about the likelihood that e-cigarette use does (or does not) lead to smoking cessation, with a heavy emphasis on how switching to e-cigarettes may not help users quit; and arguments for the differentiation between vaping and smoking, often framed as a pro-vaping argument focused on the reduced risk of e-cigarette use compared to smoking cigarettes.

The remaining topics, which make up over 60% of the retweeted content, were comprised exclusively of 4 unique messages retweeted over 70,000 times. Most (3 of the 4) retweeted messages were promotions to win a free bottle of e-juice, a variable wattage mod style e-cigarette, or an e-hookah starter kit. The remaining marketing/advertising topic of retweeted messages was a promotion for a single pro-e-cigarette blogger as an endorser and consultant for vape businesses.

Discussion

Principal Findings

As the use of e-cigarettes continues to rise, public awareness and perception of these products are likely shaped by the proliferation of messages shared and reshared on social media [26]. This study examined the public conversation on Twitter to determine overarching themes and trending topics. Topics found in this study included whether e-cigarettes are cessation

devices or gateway products for tobacco addiction, how e-cigarettes differ from traditional cigarettes, reactions to e-cigarette policies and health risk claims, news and updates from vaping communities, use of e-cigarettes where smoking bans exist, and a variety of marketing product promotions and giveaways. While past research has found a preponderance of marketing and advertising dominating the content on Twitter [24,27], this study revealed that proponents and individual users are participating in public conversations about e-cigarettes at a much larger scale than previously suggested. However, the diversity in content did not reveal a conversation with diverse perspectives. Individual user and proponent tweets intertwined with marketing messages, which still have a strong presence, to present a rather unbalanced, likely proponent-driven and perpetuated conversation about e-cigarettes use, norms, and policy.

In contrast to earlier studies, this study is the first to indicate that public conversations, from a mix of individual users and proponents, are now dominating the trending topics on Twitter for e-cigarettes, even with the inclusion of commercial activity [26]. These topics, with over 800,000 tweets generated in a 100-day window, provide rich insights into salient issues for Twitter users, especially for those who support the use of e-cigarettes. Advocates for e-cigarettes have taken to Twitter to share their thoughts and opinions, contributing to an unbalanced, likely proponent-driven conversation about e-cigarettes that heavily favors pro-e-cigarette arguments. Notably, although the most popular topics were labeled proponent and individual users conversations, these were not free of industry influence via a mixture of marketing tweets within as well as messages clearly influenced by marketing strategies from users and industry members alike. Content from e-cigarette proponents reflects the indirect marketing influence in these conversations.

It is perhaps not surprising that those passionate enough to tweet are talking about the benefits of e-cigarette use. Similar trends of Twitter conversations dominated with pro-e-cigarette content have been detected in response to e-cigarette educational campaigns and announcements of e-cigarette regulations [39,40]. Our findings, which mirror these trends of countercampaigns and antipolicy Twitter bombing to flood conversations with one perspective [39,40], highlight the pressing need for public health professionals to engage the public on social media. This finding reveals perhaps an even larger concern for all public health professionals: each person who goes online to do a little research about e-cigarettes is going to encounter a tilted conversation encouraging e-cigarette use, promoting vaping as a socially acceptable practice for all ages, discrediting scientific evidence for health risks, and rallying around the idea that e-cigarettes should largely be outside the bounds of policy. Thinking about how public health advocates can either more actively engage in this conversation or encourage a broader range of the public, inclusive of those with neutral or anti-e-cigarette positions, to post is necessary to create a more balanced conversation.

Marketing and advertising still have a strong presence on Twitter; however, our results only partially support what others have found. Previous studies have suggested that marketing

content saturates over 90% of the information about e-cigarettes on Twitter [24,27]. Although still present in this sample, this analysis revealed a much smaller proportion of marketing/advertising content as the most popular topics, especially for tweets. The fact that the explicitly marketing messages fall behind the individual user tweets indicates that they do not spread or influence the conversation as much as has been shown in previous years.

Marketing messages do spread when specific promotions are retweeted verbatim. A limited number of 140-character-or-less messages that focused on promotions and giveaways proliferated rapidly through retweets during the 100 days in this study. These retweets can be interpreted as public reactions [18], demonstrating that some attention toward promotion of regulations for e-cigarette prices and give-away promotions online is warranted. Taken together, however, explicit marketing content, although retweeted more, was a relatively small piece of the online conversation. While attention to policy and regulations for marketing content might be worthwhile, the more important public health effort is likely to focus on engaging in the conversation to create a more balanced perspective available to Twitter users.

Limitations

As with all social media research and analyses, there are several limitations to this study. While an analysis that captured a wide breadth of tweets and retweets allowed for insights about large-scale theme and topics, this does not represent the exposure for all Twitter users. Individuals customize their Twitter experiences by following accounts, thus not all users would experience the content shared on social media in the same way. Additionally, although insights for individual user and proponent opinions can be made from content shared in the topics, this analysis does not reveal the impact of the tweets or retweets on perception and attitudes toward e-cigarettes from users who see this content. Lastly, no automated denoising technique was applied prior to text mining the tweets captured in this study; we took steps to reduce noise manually. As emerging techniques become more reliable [29], future studies should consider applying automated denoising techniques before analysis.

Conclusions

As the prevalence of e-cigarettes continues to rise, it is important to know what messages about these products are potentially influencing consumer attitudes and use. This study is the first to uncover trending themes and topics from large-scale public conversations on social media.

While e-cigarette brands and distributors continue to use social media for e-cigarette marketing and promotion, these findings reveal that unique, large-scale consumer conversations are taking place on Twitter. Individuals are turning to social media to participate in discussions about policies, personal experiences, and the differentiation of e-cigarettes from traditional tobacco. Public health advocates should actively participate on social media to balance the conversation, and future research should investigate how these unique conversations influence attitudes toward and use of e-cigarettes.

Conflicts of Interest

None declared.

References

- Pearson JL, Richardson A, Niaura RS, Vallone DM, Abrams DB. e-Cigarette awareness, use, and harm perceptions in US adults. *Am J Public Health*. Sep 2012;102(9):1758-1766. [FREE Full text] [doi: [10.2105/AJPH.2011.300526](https://doi.org/10.2105/AJPH.2011.300526)] [Medline: [22813087](https://pubmed.ncbi.nlm.nih.gov/22813087/)]
- Loukas A, Batanova M, Fernandez A, Agarwal D. Changes in use of cigarettes and non-cigarette alternative products among college students. *Addict Behav*. Oct 2015;49:46-51. [doi: [10.1016/j.addbeh.2015.05.005](https://doi.org/10.1016/j.addbeh.2015.05.005)] [Medline: [26046401](https://pubmed.ncbi.nlm.nih.gov/26046401/)]
- King BA, Patel R, Nguyen K, Dube SR. Trends in awareness and use of electronic cigarettes among US adults, 2010-2013. *Nicotine Tob Res*. 2015;17(2):219-227. [doi: [10.1093/ntr/ntu191](https://doi.org/10.1093/ntr/ntu191)]
- Ramo DE, Young-Wolff KC, Prochaska JJ. Prevalence and correlates of electronic-cigarette use in young adults: findings from three studies over five years. *Addict Behav*. Feb 2015;41:142-147. [FREE Full text] [doi: [10.1016/j.addbeh.2014.10.019](https://doi.org/10.1016/j.addbeh.2014.10.019)] [Medline: [25452058](https://pubmed.ncbi.nlm.nih.gov/25452058/)]
- Adkison SE, O'Connor RJ, Bansal-Travers M, Hyland A, Borland R, Yong H, et al. Electronic nicotine delivery systems: international tobacco control four-country survey. *Am J Prev Med*. Mar 2013;44(3):207-215. [FREE Full text] [doi: [10.1016/j.amepre.2012.10.018](https://doi.org/10.1016/j.amepre.2012.10.018)] [Medline: [23415116](https://pubmed.ncbi.nlm.nih.gov/23415116/)]
- Grana R, Benowitz N, Glantz SA. E-cigarettes: a scientific review. *Circulation*. May 13, 2014;129(19):1972-1986. [FREE Full text] [doi: [10.1161/CIRCULATIONAHA.114.007667](https://doi.org/10.1161/CIRCULATIONAHA.114.007667)] [Medline: [24821826](https://pubmed.ncbi.nlm.nih.gov/24821826/)]
- Cobb NK, Byron MJ, Abrams DB, Shields PG. Novel nicotine delivery systems and public health: the rise of the “e-cigarette”. *Am J Public Health*. Dec 2010;100(12):2340-2342. [doi: [10.2105/AJPH.2010.199281](https://doi.org/10.2105/AJPH.2010.199281)] [Medline: [21068414](https://pubmed.ncbi.nlm.nih.gov/21068414/)]
- Williams M, Villarreal A, Bozhilov K, Lin S, Talbot P. Metal and silicate particles including nanoparticles are present in electronic cigarette cartomizer fluid and aerosol. *PLoS One*. 2013;8(3). [FREE Full text] [doi: [10.1371/journal.pone.0057987](https://doi.org/10.1371/journal.pone.0057987)] [Medline: [23526962](https://pubmed.ncbi.nlm.nih.gov/23526962/)]
- Kosmider L, Sobczak A, Fik M, Knysak J, Zaciera M, Kurek J, et al. Carbonyl compounds in electronic cigarette vapors: effects of nicotine solvent and battery output voltage. *Nicotine Tob Res*. Oct 2014;16(10):1319-1326. [FREE Full text] [doi: [10.1093/ntr/ntu078](https://doi.org/10.1093/ntr/ntu078)] [Medline: [24832759](https://pubmed.ncbi.nlm.nih.gov/24832759/)]
- King BA, Patel R, Nguyen KH, Dube SR. Trends in awareness and use of electronic cigarettes among US adults, 2010-2013. *Nicotine Tob Res*. Feb 2015;17(2):219-227. [FREE Full text] [doi: [10.1093/ntr/ntu191](https://doi.org/10.1093/ntr/ntu191)] [Medline: [25239961](https://pubmed.ncbi.nlm.nih.gov/25239961/)]
- Centers for Disease Control and Prevention. Tobacco Use Among Middle and High School Students—United States, 2011-2015. *MMWR*. Apr 16, 2016;65(14):361-367. [FREE Full text]
- Regan A, Promoff G, Dube S, Arrazola R. Electronic nicotine delivery systems: adult use/awareness of the “e-cigarette” in the USA. *Tob Control*. 2013;22(1):19-23. [doi: [10.1136/tobaccocontrol-2011-050044](https://doi.org/10.1136/tobaccocontrol-2011-050044)]
- Sutfin EL, McCoy TP, Morrell HE, Hoepfner BB, Wolfson M. Electronic cigarette use by college students. *Drug Alcohol Depend*. Aug 1, 2013;131(3):214-221. [FREE Full text] [doi: [10.1016/j.drugalcdep.2013.05.001](https://doi.org/10.1016/j.drugalcdep.2013.05.001)] [Medline: [23746429](https://pubmed.ncbi.nlm.nih.gov/23746429/)]
- Dawkins L, Turner J, Roberts A, Soar K. “Vaping” profiles and preferences: an online survey of electronic cigarette users. *Addict*. 2013;108(6):1115-1125. [doi: [10.1111/add.12150](https://doi.org/10.1111/add.12150)]
- Emery S, Vera L, Huang J, Szczypka G. Wanna know about vaping? Patterns of message exposure, seeking and sharing information about e-cigarettes across media platforms. *Tob Control*. 2014;23(suppl 3):iii17-iii25. [doi: [10.1136/tobaccocontrol-2014-051648](https://doi.org/10.1136/tobaccocontrol-2014-051648)]
- Pepper J, Brewer N. Electronic nicotine delivery system (electronic cigarette) awareness, use, reactions and beliefs: a systematic review. *Tob Control*. Sep 2014;23(5):375-384. [FREE Full text] [doi: [10.1136/tobaccocontrol-2013-051122](https://doi.org/10.1136/tobaccocontrol-2013-051122)] [Medline: [24259045](https://pubmed.ncbi.nlm.nih.gov/24259045/)]
- Chu K, Sidhu A, Valente T. Electronic cigarette marketing online: a multi-site, multi-product comparison. *JMIR Public Health Surveill*. 2015;1(2):e11. [FREE Full text] [doi: [10.2196/publichealth.4777](https://doi.org/10.2196/publichealth.4777)] [Medline: [27227129](https://pubmed.ncbi.nlm.nih.gov/27227129/)]
- Boyd D, Golder S, Lotan G. Tweet, tweet, retweet: Conversational aspects of retweeting on twitter. 2010. Presented at: System Sciences (HICSS) 43rd Hawaii International Conference; 2010; Hawaii.
- Chen G. Tweet this: A uses and gratifications perspective on how active Twitter use gratifies a need to connect with others. *Computers in Human Behavior*. Mar 2011;27(2):755-762. [doi: [10.1016/j.chb.2010.10.023](https://doi.org/10.1016/j.chb.2010.10.023)]
- Wu S, Hofman J, Mason W, Watts D. Who says what to whom on twitter. 2011. Presented at: Proceedings of the 20th International Conference on World Wide Web; 2011; Hyderabad, India.
- Scanfeld D, Scanfeld V, Larson E. Dissemination of health information through social networks: twitter and antibiotics. *Am J Infect Control*. Apr 2010;38(3):182-188. [FREE Full text] [doi: [10.1016/j.ajic.2009.11.004](https://doi.org/10.1016/j.ajic.2009.11.004)] [Medline: [20347636](https://pubmed.ncbi.nlm.nih.gov/20347636/)]
- Radzikowski J, Stefanidis A, Jacobsen K, Croitoru A, Crooks A, Delamater P. The measles vaccination narrative in Twitter: a quantitative analysis. *JMIR Public Health Surveill*. 2016;2(1):e1. [FREE Full text] [doi: [10.2196/publichealth.5059](https://doi.org/10.2196/publichealth.5059)] [Medline: [27227144](https://pubmed.ncbi.nlm.nih.gov/27227144/)]
- Kavuluru R, Sabbir AK. Toward automated e-cigarette surveillance: Spotting e-cigarette proponents on Twitter. *J Biomed Inform*. Jun 2016;61:19-26. [doi: [10.1016/j.jbi.2016.03.006](https://doi.org/10.1016/j.jbi.2016.03.006)] [Medline: [26975599](https://pubmed.ncbi.nlm.nih.gov/26975599/)]

24. Kim A, Hopper T, Simpson S, Nonnemaker J, Lieberman AJ, Hansen H, et al. Using Twitter data to gain insights into e-cigarette marketing and locations of use: an infoveillance study. *J Med Internet Res*. Nov 06, 2015;17(11):e251. [FREE Full text] [doi: [10.2196/jmir.4466](https://doi.org/10.2196/jmir.4466)] [Medline: [26545927](https://pubmed.ncbi.nlm.nih.gov/26545927/)]
25. Chu K, Unger J, Allem JP, Pattarroyo M, Soto D, Cruz TB, et al. Diffusion of messages from an electronic cigarette brand to potential users through Twitter. *PLoS One*. 2015;10(12). [FREE Full text] [doi: [10.1371/journal.pone.0145387](https://doi.org/10.1371/journal.pone.0145387)] [Medline: [26684746](https://pubmed.ncbi.nlm.nih.gov/26684746/)]
26. van der Tempel J, Noormohamed A, Schwartz R, Norman C, Malas M, Zawertailo L. Vape, quit, tweet? Electronic cigarettes and smoking cessation on Twitter. *Int J Public Health*. Mar 2016;61(2):249-256. [doi: [10.1007/s00038-016-0791-2](https://doi.org/10.1007/s00038-016-0791-2)] [Medline: [26841895](https://pubmed.ncbi.nlm.nih.gov/26841895/)]
27. Huang J, Kornfield R, Szczypka G, Emery SL. A cross-sectional examination of marketing of electronic cigarettes on Twitter. *Tob Control*. Jul 2014;23 Suppl 3:iii26-iii30. [FREE Full text] [doi: [10.1136/tobaccocontrol-2014-051551](https://doi.org/10.1136/tobaccocontrol-2014-051551)] [Medline: [24935894](https://pubmed.ncbi.nlm.nih.gov/24935894/)]
28. Dai H, Hao J. Mining social media data for opinion polarities about electronic cigarettes. *Tob Control*. 2016. [doi: [10.1136/tobaccocontrol-2015-052818](https://doi.org/10.1136/tobaccocontrol-2015-052818)]
29. Allem JP, Ferrara E. The importance of debiasing social media data to better understand e-cigarette-related attitudes and behaviors. *J Med Internet Res*. Aug 09, 2016;18(8):e219. [FREE Full text] [doi: [10.2196/jmir.6185](https://doi.org/10.2196/jmir.6185)] [Medline: [27507563](https://pubmed.ncbi.nlm.nih.gov/27507563/)]
30. Ferrara E, Varol O, Davis C, Menczer F, Flammini A. The rise of social bots. *Commun ACM*. Jun 24, 2016;59(7):96-104. [doi: [10.1145/2818717](https://doi.org/10.1145/2818717)]
31. Clark EM, Jones CA, Williams JR, Kurti AN, Norotsky MC, Danforth CM, et al. Vaporous marketing: uncovering pervasive electronic cigarette advertisements on Twitter. *PLoS One*. 2016;11(7):e0157304. [FREE Full text] [doi: [10.1371/journal.pone.0157304](https://doi.org/10.1371/journal.pone.0157304)] [Medline: [27410031](https://pubmed.ncbi.nlm.nih.gov/27410031/)]
32. Myslín M, Zhu S, Chapman W, Conway M. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. *J Med Internet Res*. Aug 29, 2013;15(8):e174. [FREE Full text] [doi: [10.2196/jmir.2534](https://doi.org/10.2196/jmir.2534)] [Medline: [23989137](https://pubmed.ncbi.nlm.nih.gov/23989137/)]
33. Cole-Lewis H, Pugatch J, Sanders A, Varghese A, Posada S, Yun C, et al. Social listening: a content analysis of e-cigarette discussions on Twitter. *J Med Internet Res*. Oct 27, 2015;17(10):e243. [FREE Full text] [doi: [10.2196/jmir.4969](https://doi.org/10.2196/jmir.4969)] [Medline: [26508089](https://pubmed.ncbi.nlm.nih.gov/26508089/)]
34. Cole-Lewis H, Varghese A, Sanders A, Schwarz M, Pugatch J, Augustson E. Assessing electronic cigarette-related Tweets for sentiment and content using supervised machine learning. *J Med Internet Res*. Aug 25, 2015;17(8):e208. [FREE Full text] [doi: [10.2196/jmir.4392](https://doi.org/10.2196/jmir.4392)] [Medline: [26307512](https://pubmed.ncbi.nlm.nih.gov/26307512/)]
35. Westerman D, Spence P, Van Der Heide B. Social media as information source: recency of updates and credibility of information. *J Comput-Mediat Comm*. Nov 08, 2013;19(2):171-183. [doi: [10.1111/jcc4.12041](https://doi.org/10.1111/jcc4.12041)]
36. Dubois E, Gaffney D. The multiple facets of influence: identifying political influentials and opinion leaders on Twitter. *American Behavioral Scientist*. Apr 08, 2014;58(10):1260-1277. [doi: [10.1177/0002764214527088](https://doi.org/10.1177/0002764214527088)]
37. Ampofo L, Collister S, O'Loughlin B, Chadwick A. Text mining and social media: When quantitative meets qualitative and software meets people. In: Halfpenny P, Proctor R, editors. *Innovations in Digital Research Methods*. Thousand Oaks. Sage Publications Inc; 2015;161-192.
38. Broniatowski D, Paul M, Dredze M. National and local influenza surveillance through Twitter: an analysis of the 2012-2013 influenza epidemic. *PLoS One*. 2013;8(12):e83672. [FREE Full text] [doi: [10.1371/journal.pone.0083672](https://doi.org/10.1371/journal.pone.0083672)] [Medline: [24349542](https://pubmed.ncbi.nlm.nih.gov/24349542/)]
39. Harris JK, Moreland-Russell S, Choucair B, Mansour R, Staub M, Simmons K. Tweeting for and against public health policy: response to the Chicago Department of Public Health's electronic cigarette Twitter campaign. *J Med Internet Res*. Oct 16, 2014;16(10):e238. [FREE Full text] [doi: [10.2196/jmir.3622](https://doi.org/10.2196/jmir.3622)] [Medline: [25320863](https://pubmed.ncbi.nlm.nih.gov/25320863/)]
40. Allem J, Escobedo P, Chu K, Soto D, Cruz T, Unger J. Campaigns and counter campaigns: reactions on Twitter to e-cigarette education. *Tob Control*. 2016. [doi: [10.1136/tobaccocontrol-2015-052757](https://doi.org/10.1136/tobaccocontrol-2015-052757)]

Abbreviations

API: application programming interface
e-cigarette: electronic cigarette
EM: expectation maximization

Edited by T Sanchez; submitted 26.08.16; peer-reviewed by JP Allem, H Yang; comments to author 19.09.16; revised version received 31.10.16; accepted 16.11.16; published 12.12.16

Please cite as:

Lazard AJ, Saffer AJ, Wilcox GB, Chung AD, Mackert MS, Bernhardt JM

E-Cigarette Social Media Messages: A Text Mining Analysis of Marketing and Consumer Conversations on Twitter

JMIR Public Health Surveill 2016;2(2):e171

URL: <http://publichealth.jmir.org/2016/2/e171/>

doi: [10.2196/publichealth.6551](https://doi.org/10.2196/publichealth.6551)

PMID: [27956376](https://pubmed.ncbi.nlm.nih.gov/27956376/)

©Allison J Lazard, Adam J Saffer, Gary B Wilcox, Arnold DongWoo Chung, Michael S Mackert, Jay M Bernhardt. Originally published in JMIR Public Health and Surveillance (<http://publichealth.jmir.org>), 12.12.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.