Original Paper

Topics of Nicotine-Related Discussions on Twitter: Infoveillance Study

Jon-Patrick Allem¹, MA, PhD; Allison Dormanesh¹, MS; Anuja Majmundar², PhD; Jennifer B Unger¹, PhD; Matthew G Kirkpatrick¹, PhD; Akshat Choube³, MSc; Aneesh Aithal³, MSc; Emilio Ferrara³, PhD; Tess Boley Cruz¹, PhD

¹Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA, United States
²American Cancer Society, Washington, DC, United States
³Department of Computer Science, University of Southern California, Los Angeles, CA, United States

Corresponding Author:
Jon-Patrick Allem, MA, PhD
Department of Preventive Medicine
Keck School of Medicine
University of Southern California
2001 N Soto Street, 3rd Floor, SSB 312D
Los Angeles, CA, 90032
United States
Phone: 1 323 442 7921
Email: allem@usc.edu

Abstract

Background: Cultural trends in the United States, the nicotine consumer marketplace, and tobacco policies are changing.

Objective: The goal of this study was to identify and describe nicotine-related topics of conversation authored by the public and social bots on Twitter, including any misinformation or misconceptions that health education campaigns could potentially correct.

Methods: Twitter posts containing the term “nicotine” were obtained from September 30, 2018 to October 1, 2019. Methods were used to distinguish between posts from social bots and nonbots. Text classifiers were used to identify topics in posts (n=300,360).

Results: Prevalent topics of posts included vaping, smoking, addiction, withdrawal, nicotine health risks, and quit nicotine, with mentions of going “cold turkey” and needing help in quitting. Cessation was a common topic, with mentions of quitting and stopping smoking. Social bots discussed unsubstantiated health claims including how hypnotherapy, acupuncture, magnets worn on the ears, and time spent in the sauna can help in smoking cessation.

Conclusions: Health education efforts are needed to correct unsubstantiated health claims on Twitter and ultimately direct individuals who want to quit smoking to evidence-based cessation strategies. Future interventions could be designed to follow these topics of discussions on Twitter and engage with members of the public about evidence-based cessation methods in near real time when people are contemplating cessation.

(J Med Internet Res 2021;23(6):e25579) doi: 10.2196/25579

KEYWORDS

nicotine; electronic cigarettes; Twitter; social media; social bots; cessation

Introduction

While combustible tobacco product use is declining in the United States, electronic cigarette (e-cigarette) use has risen in recent years among youth and young adults [1]. Nicotine is the primary psychoactive substance responsible for the abuse potential (ie, the likelihood that a substance will cause addiction) of combustible tobacco products and many e-cigarettes [2]. Like several other psychoactive drugs, including caffeine and amphetamines, nicotine produces acute central nervous system effects including increased heart rate, blood pressure, alertness, and decreased appetite [3], and both animal and human studies suggest that the drug may produce long-term deleterious effects on cognitive development among youth [3,4].
Research has repeatedly shown that there is substantial misunderstanding regarding the health risks of nicotine use [5]. While nicotine is the psychoactive component that sustains tobacco dependence [6], the primary carcinogenic harms are due to combustion of the tobacco leaf [3]. Nevertheless, one study demonstrated that 54% of smokers incorrectly believed that reductions in nicotine made cigarettes less dangerous [7]. Additionally, young adults (a priority population for tobacco control) commonly have misperceptions about the safety profile and nicotine content in e-cigarettes [8], including the unsubstantiated belief that e-cigarettes are relatively safe despite the burgeoning evidence indicating the products’ nicotine-related abuse potential [9,10] and associations with progression to regular combustible cigarette use [11].

Availability of different e-cigarette products like those compatible with multiple substances (eg, open-system pod mods) [12-14] or products that facilitate customization may contribute to youth experimentation and transitions to combustible cigarette use. Such nicotine-use trajectories among youth make it crucial to characterize the public’s experiences with, and perceptions of, nicotine.

Publicly accessible data from people who post to social media platforms, like Twitter, can be used to describe perceptions of nicotine and the social and environment context surrounding nicotine use [15]. Twitter is used by 22% of US adults (distributed fairly evenly through racial and gender groups), with 42% of users on the platform daily [16]. Twitter is also used by 32% of adolescents (13 to 17 years old) in the United States [17]. Previous analyses of posts to Twitter have provided insight about what the public organically discusses regarding tobacco, including the frequency of use, co-use with other substances (eg, alcohol, marijuana), mentions of tobacco product appeal, and the locations where tobacco is often used [18,19]. Past literature also highlights the role of social bots (ie, automated accounts created to produce content and interact with human accounts on Twitter) in spreading unsubstantiated health claims and misinformation on health-related topics such as vaping and vaccines [20,21]. The goal of this study was to identify and describe nicotine-related topics of conversation authored by the public and social bots on Twitter, including any misinformation or misconceptions that health education campaigns could potentially correct.

Methods

Twitter posts containing the term “nicotine” (“#nicotine” would also be included in this search) were obtained from Twitter’s Streaming Application Program Interface (API; the filtered stream using the Twitter4J library for collecting tweets with no gaps in the collection time) from September 30, 2018 to October 1, 2019. There was a total of 1,203,466 posts containing this term during this time. Similar to prior research [15,18], we removed all retweets (n=786,327) and non-English tweets (n=45,497), resulting in 371,642 unique tweets. Removing retweets allowed us to treat each observation as independent. Posts that contained the term “nicotine” but were determined to be unrelated to our research objectives were identified and removed. This included tweets containing the phrases, “bad nicotine,” “nicotine heroin,” “nicotine stain,” and “silver spoon,” as these were references to popular song lyrics. As a result of this filtering process, we were left with 364,430 unique tweets.

Next, we identified social bots [20]. Social bots may bias the data, reducing our ability to dependably describe the public’s recent experience with nicotine [22]. We used Botometer [23] to distinguish between nonbots and social bots. Botometer analyzes the characteristics of a Twitter account and scores it based on how likely the account is to be a social bot. It is considered a state-of-the-art machine learning algorithm and has been used in prior research revolving around social bots and public health [15,21,24]. The Botometer threshold was set to >4 on the scale out of 5 of English scores and similar to prior research [25]. Each Twitter account was screened after posts were collected (ie, not in real time). During this process, Twitter accounts (n=27,186) responsible for posts in our data had been deleted. Because these Twitter accounts ceased to exist and could not be processed through Botometer, we removed the posts (n=42,890) from these accounts from our data. The final sample contained 321,540 posts, with 300,360 posts from 181,439 unique nonbot accounts, and 21,180 posts from 5889 social bots.

All analyses relied on public, anonymized data; adhered to the terms and conditions, terms of use, and privacy policies of Twitter; and were performed under the institutional review board approval from the authors’ university. To protect privacy, no tweets were reported verbatim in this article. To promote full transparency and foster reproducibility, all data and code are available from the lead author and posted on his website and data repository.

To prepare tweets for analysis, we conducted a number of transformations, including (1) basic normalization (ie, lower casing all tweets; removing extra spaces, punctuation, and special characters such as brackets), (2) stop word removal (ie, removing words such as “a,” “the”), (3) normalizing Twitter account mentions (ie, @account_name occurrences in the tweets were replaced by @person — a common token for all accounts), (4) lemmatization (ie, the removal of inflections and variants of words), (5) nonprintable character removal (ie, removing emoticons or as symbols from non-English languages), and (6) removal of hashtags and URLs.

To find topics within the tweets, we generated n-grams for n=1 (ie, unigrams) and n=2 (ie, bigrams) from each tweet. An n-gram is simply a sequence of n words. For example, the phrase “Player breaks record” contains the unigrams “player,” “breaks,” “record” and the bigrams “player breaks” and “breaks record.” By generating frequency counts of the most common unigrams and bigrams, we obtained an initial sense of the commonly discussed topics. From this assessment of the most common words and phrases, 4 of the authors reviewed posts in their entirety and arrived at a consensus on 15 commonly occurring topics. This strategy was used to summarize the raw text-based data, documenting the patterns that were present. Topics included person tagging (@person), addiction (mentions of being addicted to nicotine or craving nicotine), appeal (mentions of liking or loving nicotine), nicotine replacement therapies (NRT; mentions of the patch, gum, nicotine replacement),
vaping (mentions of using e-cigarettes, vaping, JUUL), smoking (mentions of smoking cigarettes, using other combustible tobacco), nicotine health risks (mentions of nicotine effects on the brain, respiratory health, the amount of nicotine in products), withdrawal (mentions of nicotine withdrawal), quit nicotine (mentions of quitting nicotine or going nicotine free), cessation (mentions of quitting or stopping smoking), polysubstance use (mentions of alcohol and nicotine use), caffeine (mentions of coffee and nicotine use), underage use (mentions of children and teens using nicotine, use of nicotine at high schools), and new products (mentions of a “nicotine shot” or a supplement to boost the amount of nicotine in e-liquids). Nicotine is safe (mentions of nicotine not being harmful by itself) was a topic established a priori since these posts may reflect misconceptions that could be addressed by health education campaigns [26].

Each tweet was classified to one or more topics based on the occurrence of at least one topic-related pattern, which is similar to prior research [18,25]. This pattern could be a unigram, a bigram, or groups of words that must occur in the normalized tweets in a specific order. This was accomplished by using a rule-based classification algorithm developed in Python that inspects each tweet for the presence of a specified set of patterns representing a topic. Since a single post could discuss multiple topics, we report the percentage of overlap between each topic by utilizing a confusion matrix. Each cell in the matrix represents the intersection of 2 topics. The value of the cell represents the percentage of the total corpus that belongs to both topics. For example, a hypothetical post such as “Hey @person look who is nicotine free today” would be classified under “person tagging” and “quit nicotine.” The number of posts containing both would be found at the intersection of the matrix for these 2 topics.

**Results**

The total coverage of the 15 topics constituted 82.86% (248,893/300,360) of all tweets in the corpus from nonbots (Figure 1). The remaining 17.14% (51,467/300,360) of tweets were too diverse to be classified into a single topic with meaningful coverage (ie, coverage of each subsequent topic would be less than 1% of total tweets). The most prevalent topic in this corpus was “person tagging” at 40.27% (120,962/300,360), followed by “smoking” at 20.96% (62,956/300,360) and “vaping” at 20.89% (62,736/300,360). “Addiction” was the next most prevalent topic at 19.85% (59,634/300,360), followed by “NRT” at 12.14% (36,475/300,360) and “quit nicotine” at 10.23% (30,724/300,360). Among “quit nicotine,” posts suggested going “cold turkey,” a day without nicotine, trying to quit, and needing help in quitting. “Nicotine health risks” was a common topic at 7.95% (23,869/300,360), followed by “underage use” at 6.89% (20,683/300,360), “cafeine” at 5.40% (16,233/300,360), “appeal” at 4.68% (14,061/300,360), “new products” at 4.24% (12,747/300,360), and “cessation” at 3.43% (10,288/300,360). Among “cessation,” posts suggested quitting and stopping smoking. “Nicotine is safe” was an uncommon topic at 0.50% (1489/300,360).

The total coverage of the same 15 topics constituted 75.56% (16,004/21,180) of all tweets in the corpus from social bots (Figure 2). Comparing the 2 corpora, some topics had similar prevalences, while other topics stood out with large differences. For example, the largest difference in prevalence in topics between corpora was found in “person tagging” (nonbots at 40.27% [120,962/300,360] versus social bots at 20.47% [4336/21,180]), followed by “underage use” (nonbots at 6.89% [20,683/300,360], “cafeine” at 5.40% [16,233/300,360], “appeal” at 4.68% [14,061/300,360], “new products” at 4.24% [12,747/300,360], and “cessation” at 3.43% [10,288/300,360]. Among “cessation,” posts suggested quitting and stopping smoking. “Nicotine is safe” was an uncommon topic at 0.50% (1489/300,360).

The content found in each category was overall consistent between nonbots and social bots in all but “cessation.” Posts in “cessation” from social bots regularly included the use of hypnotherapy, acupuncture, magnets worn on the ears, and time spent in the sauna as effective ways to stop smoking.
**Figure 1.** Prevalence of topics from nonbot corpus. NRT: nicotine replacement therapy.
Figure 2. Comparison of prevalence of topics between nonbots and social bots. NRT: nicotine replacement therapy.

<table>
<thead>
<tr>
<th>Topics</th>
<th>Nonbots</th>
<th>Bots</th>
<th>Delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addiction</td>
<td>19.85</td>
<td>9.22</td>
<td>10.63</td>
</tr>
<tr>
<td>Appeal</td>
<td>4.68</td>
<td>2.60</td>
<td>2.08</td>
</tr>
<tr>
<td>Caffeine</td>
<td>5.40</td>
<td>1.81</td>
<td>3.59</td>
</tr>
<tr>
<td>Cessation</td>
<td>3.43</td>
<td>5.07</td>
<td>-1.64</td>
</tr>
<tr>
<td>New Products</td>
<td>4.24</td>
<td>14.62</td>
<td>-10.38</td>
</tr>
<tr>
<td>Nicotine Health Risks</td>
<td>7.95</td>
<td>8.38</td>
<td>-0.43</td>
</tr>
<tr>
<td>Nicotine Safe</td>
<td>0.50</td>
<td>0.18</td>
<td>0.32</td>
</tr>
<tr>
<td>NRT</td>
<td>12.14</td>
<td>16.96</td>
<td>-4.82</td>
</tr>
<tr>
<td>Person Tagging</td>
<td>40.27</td>
<td>20.47</td>
<td>19.80</td>
</tr>
<tr>
<td>Polysubstance Use</td>
<td>2.66</td>
<td>1.20</td>
<td>1.46</td>
</tr>
<tr>
<td>Quit Nicotine</td>
<td>10.23</td>
<td>9.73</td>
<td>0.50</td>
</tr>
<tr>
<td>Smoking</td>
<td>20.96</td>
<td>18.79</td>
<td>2.17</td>
</tr>
<tr>
<td>Underage</td>
<td>6.89</td>
<td>5.00</td>
<td>1.89</td>
</tr>
<tr>
<td>Vaping</td>
<td>20.89</td>
<td>27.61</td>
<td>-6.72</td>
</tr>
<tr>
<td>Withdrawal</td>
<td>1.88</td>
<td>0.85</td>
<td>1.03</td>
</tr>
</tbody>
</table>

Discussion

Principal Findings

This study is one of the largest Twitter studies to date focused on nicotine-related conversations, describing over 300,000 unique posts from over 180,000 unique accounts and addressing the underlying questions of what the public discusses or perceives about nicotine (rather than focusing on one specific tobacco product). We identified a number of topics of conversation ranging from nicotine appeal to withdrawal to smoking cessation. Posts discussed addiction, NRT, health risks, and nicotine use in combination with alcohol and caffeine. This study also distinguished nicotine-related topics of conversations by social bots and nonbots, describing differences in prevalence of topics by account type.

In this study, Twitter posts mentioning new products represented a larger proportion of posts by social bots compared to nonbots, suggesting that companies or retailers or e-cigarette hobbyists may be using bots to promote new products. Social bots have previously been found to promote emerging products on Twitter; for example, in 2017, it was found that social bots were more than 2 times as likely to post about a new vaping product compared to nonbots [15]. Posts from social bots identified in the present study perpetuated a number of methods with very limited evidence as smoking cessation interventions, including hypnotherapy, acupuncture, trips to the sauna, and the use of magnets behind the ear. In contrast to front-line treatments such as tailored behavioral counseling (eg, individual, group, and phone) and medication (eg, varenicline, bupropion, NRT), these alternative methods have little to no empirical evidence to support their efficacy [27,28]. Unsubstantiated health claims perpetuated by social bots may have offline consequences, such as leaving Twitter users with the impression that these methods are good cessation strategies, thus diverting them from more effective approaches.

Unsubstantiated health claims on Twitter from social bots have been documented in prior research; for example, several studies have reported that social bots regularly make claims touting the effectiveness of e-cigarettes in smoking cessation [15,24] and claims propagating misinformation pertaining to vaccinations [21]. Recently, it was reported that social bots were responsible for disseminating unsubstantiated health claims pertaining to cannabis with posts suggesting cannabis could allay health concerns ranging from triple-negative breast cancer to plantar fasciitis [25]. Health education efforts are needed to correct misinformation and ultimately direct individuals who want to quit smoking to evidence-based cessation strategies [27,29]. Misperceptions or myths about cessation could be most persuasively countered with two-sided messages that provide a brief acknowledgement of the misconception, then refute it,
and followed by a stronger statement about the more effective intervention [30]. For example, Twitter posts could be circulated that state: “If you feel addicted to cigarettes, you could try quitting cold turkey or with hypnotherapy, but you are more likely to succeed if you work with a Quitline like 1-800-NO BUTTS.”

“Person tagging” was a predominant theme in the current study of nicotine-related posts to Twitter and in line with prior research [18,25]. Person tagging in this context is a social practice where Twitter users directly interact with one another to exchange their attitudes about and experiences with nicotine. Posts classified under “person tagging” regularly used @Person to engage others in discussions about nicotine. These online communications may impact nicotine use; for example, Unger and colleagues [31] demonstrated an empirical link between adolescents’ and young adults’ tobacco-related Twitter activity and their tobacco product use. The current study’s findings are highly relevant to the public health community, as repeated exposure to nicotine-related messaging and reported nicotine use by Twitter connections may influence the social norms of those exposed to the content and lead to imitation of the behaviors [32].

Prior research has shown that a cessation program utilizing Twitter to deliver an intervention for smoking cessation can be successful in helping participants sustain abstinence [33]. The present study did not identify participants looking to quit smoking on Twitter; however, these findings suggest that Twitter may be a place where such participants could be found as people tweet about the difficulty of quitting nicotine. “Vaping,” “addiction,” “quit nicotine,” “withdrawal,” and “cessation” were all topics in the present study. Future interventions could be designed to follow these topics of discussions on Twitter and engage with potential participants about evidence-based cessation methods in near real time when people are contemplating cessation [34].

“Polysubstance use” and “caffeine” were identified as topics in the current study. Polysubstance use has been reported in several earlier Twitter-based studies; for example, a prior analysis of hookah-related posts to Twitter from 2017 to 2018 found that many posts described alcohol, marijuana, and other substance use along with hookah [18]. Similar findings were reported in Twitter studies focused on e-cigarettes [15] and cannabis [35]. Past work also raises concerns about the unknown health effects of caffeine in flavored e-liquids [36] and preference of e-liquids with active caffeine ingredients for weight loss [37]. The present findings supplement these previous studies and further awareness of the occurrence of polysubstance use. This is particularly important because alcohol and caffeine can potentiate the reinforcing effects of nicotine [38,39], potentially leading to escalation in use of one or both substances. Similar to prior Twitter studies focused on JUUL use [40], the current study found posts indicative of underage use of nicotine (ie, mentions of nicotine use at high schools and among teenagers). This finding is concerning because nicotine impairs adolescents’ and young adults’ brain development [2,3,41]. In addition, posts about underage use may normalize e-cigarettes in young viewers, with the potential to increase experimentation and regular use [42].

Limitations
This study focused on posts to Twitter, and findings may not extend to other social media platforms. The posts in this study were collected within a 12-month period and may not extend to other time periods. Data collection relied on Twitter’s Streaming API, which prevented collection of posts from private accounts. Findings may not generalize to all Twitter users or to the US population. Not all tweets were covered by the established categories, and topics of conversation were not segmented by geographic location, preventing this study from understanding the effect of different state tobacco policies on the public’s experience with nicotine. Prior research has shown that significant geographic biases can occur in the context of conversations over Twitter [43,44]. In some instances, unigrams and bigrams used to define topics may have multiple meanings that were ignored in the current study; for example, the word “school” in nicotine-related posts may not always indicate underage use, as college students or other educational professionals may be discussing nicotine use.

Conclusions
Common nicotine-related topics on Twitter included smoking, vaping, cessation, withdrawal, and appeal, among others. These results suggest that Twitter users often discuss grappling with quitting smoking, nicotine withdrawal, and nicotine cravings. Such topics of conversation warrant considerations by public health researchers in the future. Twitter may act as a platform to engage with those struggling with nicotine dependence, as well as those initiating use with nicotine-related products, by informing them of the potential for dependence and subsequent health consequences of use. Posts from social bots regularly included the use of hypnotherapy, acupuncture, magnets worn on the ears, and time spent in the sauna as effective ways to stop smoking. Misinformation regarding nicotine has been a component of tobacco industry marketing and has the potential to influence beliefs, perceptions, and use of tobacco; thus, it is important to provide a recent account of what posts discuss on Twitter about nicotine in hopes of correcting misinformation and directing tobacco users to more effective interventions.

Acknowledgments
Research reported in this publication was supported by the National Cancer Institute and the Food and Drug Administration Center for Tobacco Products (CTP) [Grant # U54 CA180905]. The funders had no role in study design; collection, analysis, and interpretation of data; writing the report; and the decision to submit the report for publication. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funders.
References


Abbreviations

API: application program interface
NRT: nicotine replacement therapy

©Jon-Patrick Allem, Allison Dormanesh, Anuja Majmundar, Jennifer B Unger, Matthew G Kirkpatrick, Akshat Choube, Aneesh Aithal, Emilio Ferrara, Tess Boley Cruz. Originally published in the Journal of Medical Internet Research (https://www.jmir.org), 07.06.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on https://www.jmir.org/, as well as this copyright and license information must be included.