Published in final edited form as:

Tob Regul Sci. 2018 March ; 4(2): 30–43. doi:10.18001/TRS.4.2.3.

Whose Post Is It? Predicting E-cigarette Brand from Social Media Posts

Elizabeth A. Vandewater, PhD*,

Director of Data Science and Research Services, The University of Texas at Austin, Population Research Center, Austin, TX

Stephanie L. Clendennen, MPH,

Predoctoral Fellow, UTHealth School of Public Health in Austin, TX

Emily T. Hébert, DrPH,

Postdoctoral Research Fellow, The University of Oklahoma Health Sciences Center, Oklahoma City, OK

Galya Bigman, PhD,

Graduate Research Assistant, UTHealth School of Public Health in Austin, TX

Christian D. Jackson, MS,

Statistician, UTHealth School of Public Health in Austin, TX

Anna V. Wilkinson, PhD, and

Associate Professor, UTHealth School of Public Health in Austin, TX

Cheryl L. Perry, PhD

Professor and Regional Dean, UTHealth School of Public Health in Austin, TX

Abstract

Objectives—E-cigarette advertisers know that 76% of youth use social media, yet little is known about the nature of e-cigarette advertising on social media most favored by youth. We utilized text-mining to characterize e-cigarette advertising and marketing messages from image-focused social media brand sites, and to construct and test an algorithm for predicting brand from brand-generated social media posts.

Methods—Data comprised 5022 unique posts accompanied by an image from Facebook, Instagram or Pinterest e-cigarette brand pages for Blu, Logic, Metro, and NJOY from February 2012 to April 2015. Text-tokenization was used to quantify text for use as predictors in analyses.

Results—Blu had the largest social media presence (65%), followed by Logic (16%), NJOY (12%) and Metro (7%). Blu's average post length was significantly shorter than all other brands. Words most commonly used in posts differed by brand. Regression analyses successfully differentiated Blu and NJOY brands from other brands.

*When work was completed, Correspondence Dr Vandewater; elizvan@prc.utexas.edu.

Conflicts of Interest

The authors have no conflicts of interest to report.

Conclusions—Analyses revealed e-cigarette brands used different types of messages to appeal to social media users. Whereas words used by Blu and NJOY sold a “lifestyle,” words used by Logic and Metro relied on device and product identification.

Keywords

text analytics; text-mining; text-tokenization; social media; marketing; e-cigarette brands; youth

Since becoming available on the US market in 2007, electronic cigarettes (e-cigarettes) quickly became a multi-billion dollar industry.¹ The US Food and Drug Administration (FDA) was given regulatory authority over e-cigarettes in 2016.² However, unlike cigarettes and smokeless tobacco, the advertising and marketing of e-cigarettes remains unrestricted.

The introduction of e-cigarettes to the US market was accompanied by a huge wave of online marketing efforts. E-cigarettes are widely promoted online,³⁻⁵ particularly on social media websites such as Facebook, Twitter, YouTube, and Instagram. Huang et al⁶ documented extensive e-cigarette marketing on Twitter, with more than 73,000 tweets related to e-cigarettes captured in a 2-month period. Content analyses of e-cigarette-related tweets suggest that the majority of tweets are commercial in nature,^{6,7} and that price promotions and discount codes are common.⁷ E-cigarette content is also popular on Instagram, where the total number of posts with the hashtag “#vape” grew by 4 million from 2014 to 2015,⁸ and, by 2015, advertisement-themed e-cigarette images outnumbered personal photos of products or vaping.⁹ E-cigarette brand-sponsored pages are also widespread on Facebook,⁵ and a content analysis of 365 e-cigarette videos on YouTube revealed that 85% were industry-sponsored.¹⁰

In the absence of marketing regulation, e-cigarette manufacturers regularly make unsubstantiated claims about the safety and use of e-cigarettes for smoking cessation.^{11,12} Tactics known to appeal to youth, such as mention of flavors,¹³ use of celebrity endorsements,¹² and use of cartoons or animation,¹⁴ are also commonplace in e-cigarette marketing.¹⁵

The unregulated and widespread marketing of e-cigarettes on social media is especially troublesome due to its broad appeal to youth. In 2015, 92% of youth aged 13-17 reported using the Internet daily, and 76% reported using social media.¹⁶ Most youth use multiple social media platforms, with a majority using Facebook (71%) and Instagram (52%).¹⁶ Marketers are well aware of the appeal of social media to youth, and have adopted new marketing strategies to incorporate interactive material that encourages users to share branded content and brand preferences to their peers, thereby eluding policies prohibiting corporate promotion of tobacco products on many social media websites.

Marketing on social media encourages user-interaction and engagement in ways that traditional marketing does not,¹⁷ and as such, has the potential for greater influence on youth. E-cigarette brands exploit features unique to social media in their marketing, such as mentioning consumers directly by username on Twitter or allowing users to comment on posts,⁵ thereby blurring the line between industry and peers.¹⁸ Thus, this new wave of digital and social media marketing is particularly effective at targeting adolescents and young

adults, as it utilizes peer acceptance, social interaction, and identity development through the use of “likes” on Facebook pages, “re-tweets” on Twitter, and positive product review videos on YouTube.

Research has demonstrated that youth and young adult exposure to pro-tobacco social media is associated with more positive attitudes towards smoking,¹⁹ increased positive perceptions of peer smoking norms,^{19,20} greater smoking intentions,¹⁹ and smoking behavior.^{6,21} Exposure to e-cigarette marketing is associated with e-cigarette use and increased susceptibility to e-cigarettes among middle and high school students,²² with stronger associations between exposure to marketing online and both susceptibility and use versus exposure to print, retail, and television marketing.^{22,23}

Given the lack of regulation of e-cigarette marketing and the potential influence of social media marketing on youth, more research is needed to improve understanding of the nature, extent, and impact of e-cigarette marketing via social media. However, analyzing social media data presents challenges. First, the data are largely text based, which until recently has been the purview of qualitative analytic techniques, mainly thematic or content coding. In theory, identifying themes or content in social media data is no different than identifying themes in other textual data, such as are generated from focus groups, in-depth interviews, or open-ended survey responses. Yet, the sheer volume of data generated by social media has rendered such coding strategies difficult if not logistically impossible. Whereas traditional content coding strategies tend to be both labor-intensive and time consuming, the emerging field of text-mining offers an array of novel tools and methods suitable for analysis of large amounts of textual data such as those typically generated by social media platforms.²⁴⁻²⁶ A third problem presented by social media data pertains to identifying content relevant posts. Text-mining techniques include a range of simple and sophisticated applications, all of which transform text into a structured format suitable for qualitative and/or quantitative analysis.²⁴

Indeed, a nascent but growing body of literature has begun to examine the prevalence, nature, and content of e-cigarette marketing on social media using text-mining techniques. Despite the difficulties and time involved, many studies to date have used manual content coding to identify themes or topics in e-cigarette related posts.^{9,27,28} Subsequently, several studies have applied text-mining techniques to the study of e-cigarette related content on social media. Some have focused on developing automatic detection algorithms for identifying e-cigarette related content,²⁹ detection of robotic (automated) versus organic (human) posts,^{30,31} and detection of e-cigarette proponents.³² Others have used text-mining to identify key patterns, topics, genres, themes or sentiment of e-cigarette related posts.³³

However, although use of text analytic techniques to examine e-cigarette marketing empirically on social media platforms is growing, most existing research has focused largely on Twitter.^{6,7,29-41} Far fewer have focused on Youtube^{10,11,42} or Instagram.^{8,9,43} Finally, though some studies have used Facebook to characterize traditional tobacco marketing,²⁷ or to recruit survey participants regarding e-cigarette use,⁴⁴ no study appears to have analyzed e-cigarette related Facebook posts at the time of this writing.

As a result, there is a dearth of evidence about the nature and quality of e-cigarette advertising on the image-focused social media platforms most favored by youth and young adults (such as Facebook and Instagram), particularly given their popularity in these populations.¹⁶ In addition, there is a general lack of knowledge about the content of e-cigarette advertising on image-focused social media platforms generated by the brands themselves. Analysis of brand-generated posts can reveal information relevant to several critical questions about e-cigarette marketing on social media platforms known to be heavily used by youth. In this study, we aimed to address 2 such questions: (1) Can the nature of e-cigarette advertising be characterized based on the words used by specific e-cigarette brands in their social media posts? and (2) How well can we predict the e-cigarette brand based on words used in brand-generated posts? Relative to existing studies, our study is novel in that: (1) we rely on brand-generated (not user or retail outlet) posts; (2) we focus exclusively on posts from image-focused social media known to be highly popular among youth (rather than Twitter); and (3) we utilize a text-analytics approach which allows us to convert text to predictors suitable for use in quantitative analysis.²⁵

METHODS

Data Collection

Data for this study included posts made by 4 e-cigarette brands (Blu, NJOY, Logic, and Metro) on their Facebook, Instagram, and Pinterest pages between February 3, 2012 and April 16, 2015. These 4 brands were chosen because they had relatively large market shares in the US, had a presence on image-focused social media, and were available for purchase in Texas. We focused on brands available for purchase in Texas because this study complemented a larger data collection effort regarding tobacco product use among youth and young adults undertaken by scholars at the Texas Center for Tobacco Regulatory Science (TCORS).^{45,46} Data were purchased from Curalate (www.Curalate.com), a visual commerce platform that collects and compiles posts from brand-owned social media accounts accompanied by a brand-generated image. Information for each post included date and time stamps, post text, post URL, image URL, social network type (Facebook, Instagram, or Pinterest), and brand (Blu, NJOY, Logic, or Metro).

Approach

To prepare the text for quantitative analysis, we utilized an algorithm for converting (featurizing) text into conventional quantitative explanatory variables proposed by Foster, Stine, and Liberman,²⁵ who used it to build a regression model to predict the selling price of real estate from listing content. This approach relies on vector space models (VSM) for deriving semantic meaning as detailed in Turney and Pantel.²⁶ VSMs represent each document in a collection as a point in space (ie, a vector in a vector space). Points closer together in a vector space are semantically similar, whereas points further apart are semantically different. The values of the elements in a VSM are derived from event frequencies for example the number of times a particular word appears in a particular context (such as a document).^{26,47} In this study, the elements in the VSM reflect the number of times a word appears in a single brand-generated post.

The algorithm for featurizing text proposed by Foster et al²⁵ involves 3 steps. First, all source text is converted into word tokens, a process known as tokenization. A word token is an instance of a word, defined as a unique sequence of characters delimited by white space. Next, a matrix is computed that counts the number of times a word token appears within each document (or post in this case). This matrix is known as a document-term matrix.²⁶ Finally, a truncated singular value decomposition (SVD) of the resulting matrix of counts is computed. The leading singular vectors of this decomposition (ie, unique words that appear most frequently across all documents) are then used as explanatory variables in quantitative models.²⁵ Though other approaches are possible, we chose to use a unigram or “bag of words” approach in this study. In this approach, the importance of words to semantic meaning are considered individually, without consideration of word-adjacencies (bi-grams), or phrases (tri-grams or greater).²⁶

Data Preparation

The open source platform R (v.3.2.3) was used for all 3 steps required to featurize text as described above. To compute the document term matrix, word tokens based on text must be made uniform. For example, without this step, a capitalized word (eg, E-cigarette) would be counted as a separate word token from the same lower-case version of the word (eg, e-cigarette). Therefore, all posts were cleaned using the text-mining (tm) package in R (v. 3.2.3). Duplicate posts, meaningless and unrecognizable characters, and all punctuation were removed. All words were made lower-case. Finally, English stop words (ie, extremely common English language words including articles and pronouns) were removed.

The final sample included 4479 unique words (ie, terms) from 5022 unique posts (ie, documents) (Blu posts N = 3243, NJOY posts N = 590, Logic posts N = 816, Metro posts N = 373). These posts and unique word tokens were used to compute a document-term matrix containing 5022 rows (each representing an individual post), and 4479 unique words (as columns), with cells containing the frequency of each word used in each post. The final data frame for analysis combined information from the computed document-term matrix with information regarding social network platform (Facebook, Instagram, or Pinterest), brand (Blu, NJOY, Logic or Metro), and post length (number of words per post).

Data Analysis

Descriptive analyses included the frequency distributions of words by brand and the mean length of posts by brand. To assess the reach and popularity of the brand-generated social media posts in this sample, we examine the number of comments, likes, and re-shares for the 5022 posts by brand. To provide a preliminary characterization of the nature of Blu, NJOY, Logic, and Metro brand marketing messages, word clouds were created, producing a visual representation of the 25 most frequently used words by each brand in their social media posts. In these word clouds, it was apparent that 6 pairs of words had the same meaning (battery/batteries, cartridge/cartridges, cigarette/cigarettes, e-cig/e-cigs, e-cigarette/e-cigarettes, and smoke/smokes). Thus, these words were re-tokenized to represent a single word in the ensuing analyses. In addition, the preliminary word clouds revealed that words unique to particular brands reflected brand-sponsored events that happened to occur during data collection. For example, Blu sponsored an Indy Car Race, and NJOY sponsored a

World Series of Poker event during data collection. Words clearly related to these events (such as “Indycar” for the Indy Car Race and “WSOP” for the World Series of Poker, among others) were excluded in the quantitative analyses detailed below.

To characterize the nature of the posts generated by each brand, we first conducted a multinomial logistic regression with brand as the outcome and the most frequently used words across all posts as predictors. Brand was a categorical outcome with 4 categories (Blu [chosen as the reference category based on alphabetic order], NJOY, Logic and Metro). Following the procedure outlined in Foster et al.,²⁵ the 100 most frequently used words were examined for potential use as predictors in the regression model.

Because we were interested in our ability to differentiate the brands without the use of obvious brand identifiers, we excluded references to the brand (ie, brand names), words used exclusively used by a particular brand (eg, blufreedom, NJOY, etc), as well as words related to events sponsored by a particular brand that happened to occur during data collection. The final model included 59 predictors representing the number of times each of the 59 words were used across all 5022 posts – battery(ies), beach, buy, can, cartridge(s), charger, cherry, cigarette(s), coupon, cuban, customer, day, disposable, e-cig(s), e-cigarette(s), electric, electronic, eliquid, enjoy, espresso, event, flavor, find, free, freedom, friends, get, good, happy, kit, king, lounge, love, menthol, never, new, switch, one, order, pack, platinum, premium, saturday, smoking, smoke(s), starter, super, today, tobacco, using, use, vanilla, vape, vapor, vaping, vapelyfe, vaporlounge, and world. Thus, the multinomial logistic regression provides information on the odds that a post was generated from a particular brand (relative to the reference brand) based upon the frequency of use of each of the 59 words, respectively. The multinomial logistic regression also generates predicted probabilities, which is the probability that each of the 5022 posts was from a particular brand based upon the model.

Following the multinomial logistic regression, we conducted a discriminant function analysis (DFA).⁴⁸ DFA provides additional information not provided by multinomial logistic regression, including whether clusters of words are most responsible for brand differentiation, and the order of importance of words within each cluster responsible for group differences. Specifically, DFA produces a maximum of N groups-1 (in this case linear discriminant functions, which is a linear combination of variables calculated to maximize the proportion of variance explained by group differences. Functions are derived such that the first function represents the linear combination of variables most responsible for group differentiation, and so on. Thus, DFA identifies whether clusters of variables significantly predict group membership (in this case brand), how many clusters significantly predict group membership, which variables (in this case words) belong to each cluster, and the proportion of group variation accounted for by each cluster of words. In addition, DFA allows identification of the order of importance of the words within each cluster for group differentiation. DFA also produces information regarding the percentage of posts correctly classified, and provides group centroids. Group centroids represent the group (brand) mean on each discriminant function, providing additional information on the relative score of each brand on each discriminant function. Thus, group centroids serve to indicate the extent to

which a brand is characterized by the words comprising each discriminant function.⁴⁸ All statistical analyses were conducted using R (v. 3.2.3).

RESULTS

Descriptive Analyses

Blu had the largest social media presence, comprising 65% of posts, followed by Logic (16%), NJOY (12%), and Metro (7%). Although Blu had the largest number of posts, its average post length was significantly shorter than all other brands (Blu $M = 10.66$ words, $SD = 9.73$; Logic $M = 14.08$, $SD = 9.32$; Metro $M = 15.01$, $SD = 13.87$; NJOY $M = 16.81$, $SD = 9.93$). Blu posted most frequently on Facebook (73% of posts), followed by Instagram (21%), and Pinterest (6%). Logic posted on Pinterest (96%) and Instagram (4%). NJOY posted on Facebook (46%) and Instagram (54%). Metro posted almost entirely on Pinterest (99%) and only twice on Instagram (1%). In terms of reach, Blu posts ($N = 3243$) were commented on 15,831 times, liked 83,024 times, and re-posted 286 times. Logic posts ($N = 816$) were commented on 20 times, liked 374 times, and re-posted 550 times. NJOY posts ($N = 590$) were commented on 991 times and liked 7499 times. Data for re-posts of NJOY posts were unavailable. Metro posts ($N = 373$) were commented on 27 times, liked 897 times, and re-posted 2115 times.

Word Clouds

Preliminary word clouds were created to reveal the most frequently used words by each brand, as well as to get a sense of the different marketing approaches or brand voices of Blu, NJOY, Logic, and Metro brands. As Figure 1 shows, each word cloud contains the 25 most frequently used words for each brand. The most frequently used words appear in the center of the word cloud and are the largest in size. Word frequency decreases as words become smaller in size and move out from the center of the word cloud. In the word clouds, it appears that both Blu and NJOY posts focused on promoting a lifestyle, with Blu being more “laid back” and NJOY more “hard partying.” Blu frequently and uniquely used words including “lounge,” “beach,” and “vapor,” whereas NJOY used words including “vapelif,” “poker,” and “vegas.” Logic and Metro posts were more pedestrian. Logic appeared to focus on product purchase, while Metro focused on e-cigarette devices and their use. Logic frequently used words including “e-cig(s),” “tobacco,” “electric,” and “premium.” Metro used words including “e-cig(s),” “electronic,” “battery(ies),” and “cartridge(s).” Interestingly, whereas the single most frequently used word by Blu, NJOY, and Logic was the brand name, this was not true of Metro. Metro’s most frequently used word was in fact “e-cigs,” followed by “electronic,” and then “metro.” The preliminary word clouds provided an initial visualization of the nature of posts for each brand, while also aiding identification words that were essentially brand proxies. Such words represented brand sponsored events including “poker,” “vegas,” and “wsop” reflecting a World Series of Poker event sponsored by NJOY, and “indycar,” “grand,” “prix,” and “sxsw” reflecting Blu’s sponsorship of an Indycar Grand Prix and South by Southwest (SXSW) music festival during the time of data collection.

Multinomial Logistic Regression

Prior to running the multinomial regression, we examined the correlation coefficients among the 59 words examined as predictor variables; we found no indication that collinearity was present.⁴⁹ Results of the multinomial regression analysis with Blu as the reference category are shown in Table 1. Model results for the remaining comparisons, specifically NJOY versus Logic, NJOY versus Metro, and Logic versus Metro, were similar to model results with Blu as the reference category. Odds ratios presented in Table 1 can be interpreted such that for each unit increase in the use of a word, the odds that the post was generated from a particular brand relative to the reference brand, increase by a factor of x , where x is the odds-ratio.

Relative to Blu, the odds that a post was generated by NJOY was significantly greater for 15 words and significantly lower for 10 words; the odds that a post was generated by Logic was significantly greater for 16 words and significantly lower for 8; and the odds that a post was generated by Metro was significantly greater for 19 words, and significantly lower for 8. Relative to Blu, 9 words were related to greater odds that a post was generated by NJOY exclusively (“buy,” “day,” “friends,” “get,” “happy,” “order,” “today,” “vapelite,” and “world”), and one (“electric”) was related to lower odds. Seven words were related to greater odds that a post was generated by Logic exclusively (“customer,” “event,” “menthol,” “premium,” “tobacco,” “use,” and “vapor”), and 4 were related to greater odds that a post was generated by Metro exclusively (“battery(ies),” “coupon,” “disposable,” and “never”).

Figure 2 shows the predicted probabilities generated from the logistic regression with Blu as the reference category. The predicted probabilities for each brand are represented in 3 panels because they are plotted in a 3-dimensional space yet represented in 2 dimensions. The predicted probabilities represent the probability that each of the 5022 posts belongs to a particular brand based upon the model. As such, they represent how well the model was able to predict which brand a post was from because of the words used across all posts. Posts predicted to be from Blu are shown in blue, Logic in orange, Metro in purple, and NJOY in green. The clustering of posts by brand assigned colors indicates that the model was reasonably successful in predicting which brand made a particular post. However, the model did a better job of predicting which posts belonged to Blu and NJOY than predicting which posts belonged to Logic and Metro. Further details regarding how well the model predicted brand is provided via the discriminant function analysis.

Discriminant Function Analysis

Three significant discriminant functions (ie, clusters of words) were identified, accounting for 100% of the variance between brands. The first function ($F = 45.84$; $p < .000$) accounted for 47% of between-group variance; the second function ($F = 37.43$; $p < .000$) accounted for 33% of between-group variance; and the third function ($F = 28.38$; $p < .000$) accounted for the remaining 20% of the between-group variance.

Table 2 presents the function labels, total structure coefficients, and standardized structure coefficients. Total structure coefficients are akin to factor loadings, and represent the bivariate correlation between each word and each discriminant function. As such, they

indicate which words load on which function. Standardized structure coefficients are akin to beta coefficients in regression, and represent the relationship between the word and the discriminant function controlling for all other words in the model. Thus, they reveal the relative importance of each word within the cluster of words loading on the function. Similar to latent factors, discriminant functions are labeled based on their content. We labeled the first function “Lounging” because it was comprised of the words “lounges,” “beach,” “vapor,” and “e-cig(s).” The second function was labeled “Vaping” as it was comprised of “vape,” “vapelite,” “vapelyfe,” “world,” “cigarette(s),” “electronic,” “super,” and “tobacco.” We called the third function “Device” because it was comprised of “cartridge(s),” “battery(ies),” “coupon,” “charger(s),” and “premium” (listed in order of importance to group discrimination, respectively).

Discriminant function analysis produces 2 other estimates relevant to our purposes, group centroids and percent correctly classified. Group centroids indicate the average score for each brand on each linear discriminant function. As such, they indicate how much a particular brand is characterized by each function. Figure 3 presents the group centroids of each brand on each of the 3 functions. As Figure 3 shows, Blu is positively characterized by the Lounging function only. NJOY is positively characterized by Vaping, negatively characterized by the Lounging function, and slightly negatively characterized by the Device function. Logic scored negatively on all 3 functions, indicating that it is decidedly not reflected by the content of any of the 3 functions. Finally, Metro is positively characterized by the Device function, but negatively characterized by the Lounging and Vaping functions. With respect to percent correctly classified, the results of the DFA indicated that the model correctly classified 91.40% of Blu posts, 61.63% of NJOY posts, 50.87% of Logic posts, and 45.55% of Metro posts.

Thus, although clearly successful at predicting Blu posts, and somewhat successful predicting NJOY posts, the model was less successful at predicting either Logic or Metro posts. Whereas at least some of this difference can be attributed to the fact that Blu had many more posts relative to the other brands, NJOY had fewer posts than Logic, indicating that the success of the predictive model was not based purely on the sheer number of posts.

DISCUSSION

To our knowledge, this is the first study to convert text into conventional quantitative explanatory variables to characterize the nature of e-cigarette marketing messages generated on image-focused social media brand pages. Our findings reveal that the 4 major e-cigarette brands used rather different types of messages to appeal to users on image-focused social media platforms. Whereas terms used by both Blu and NJOY sold a “lifestyle,” words used by Logic and Metro relied on device and product identification (eg, “charger,” “cigarette(s)”). This difference was apparent in the regression models, where words associated with Blu and NJOY social media posts could be discerned, but there was less differentiation between words used in Logic and Metro posts.

This study is novel in several ways. First, we examined image-focused social media platforms, including Facebook, Instagram, and Pinterest. Although most research in this area

has utilized text-based social media such as Twitter,^{6,7,29-41} Facebook and Instagram are the most common platforms among US teens who use social media.¹⁶ Given that e-cigarette use among youth has more than tripled since 2011,⁵⁰ it is critical to evaluate the nature and extent of e-cigarette-related marketing on social media platforms most likely to have a broad youth audience. Second, we specifically analyzed brand-generated social media marketing posts. To date, most studies have analyzed user-generated content or have been unable to identify the source of the social media content analyzed.⁵¹ However, research has suggested that a large proportion of e-cigarette related social media is brand-sponsored, with one study finding that 80% of all tweets were automated or promotional in nature.³⁰ The analytic methods utilized in this study may enable future researchers to distinguish brand-sponsored social media generated by individual users who are paid to promote the brand from content generated by non-paid users. Moreover, we were able to correctly identify over 90% of Blu posts, and over 60% of NJOY posts. Thus, the analytic approach used here may be used in future studies to identify whether or not a social media post was industry-sponsored, as well as the specific brand from which it was generated. The ability to distinguish industry-sponsored from user-generated messages is important not only for identifying potential targets for regulation, but also for filtering out third-party posts from studies that aim to evaluate consumer attitudes and behaviors based on social media patterns.⁵¹

Limitations

This study has certain limitations. Although we believe that the data provided were a census of all brand-generated posts accompanied by a brand-generated image, we are not able to verify this. Thus, whereas we know that the posts were brand-generated, it is possible that some posts were missed. In addition, any characterization of brand marketing is based upon posts from a specific time period, and thus, may or may not be relevant to posts from other time periods, given the dynamic nature of advertising. Future research could use text-mining techniques to track marketing trends in real time and examine how marketing trends change over time. Furthermore, although we were able to characterize the content of marketing messages posted by e-cigarette brands, we do not know who was exposed to these messages. An individual's exposure to social media messaging is dependent on their individualized online experiences. Because all social media posts included in this analysis were public, individuals did not have to be followers of the e-cigarette brand social media accounts to view marketing messages. This makes it difficult to characterize the reach or impact of these messages completely, even with the information presented regarding the number of comments, likes, and re-shares.

Finally, whereas we focused on messages generated from largely image-based social media platforms, we did not analyze the images themselves. It is important to note that text associated with images, although useful in itself, does not reflect the entire content of brand generated messages on image-focused social media. Additionally, it seems likely that our model was able to predict Blu brand posts most accurately due to the much larger number of posts from Blu relative to the other 3 brands examined. Future research should address the minimum number of posts needed to identify brand posts adequately. Thus, although this study represents a relatively rare examination of text from image-focused social media sites, our findings are limited to text, nonetheless.

IMPLICATIONS FOR TOBACCO REGULATION

The FDA's recent deeming rule to govern the use and sale of e-cigarettes fails to address the marketing of e-cigarettes.⁴ E-cigarette companies have flooded social media, which is inexpensive, wide-reaching, and highly popular among young people.¹⁶ Findings from this study highlight the use of "lifestyle" characterizations as a marketing strategy for both Blu and NJOY, the 2 brands with the largest market share in the US. Given the preoccupation with identity development among youth who are also heavy users of the social media platforms where such "lifestyles" are being promoted, these findings support a call for the regulation of e-cigarette industry-generated marketing, particularly on social media websites frequented by youth under the age of 18. Enacting comprehensive regulations on social media marketing is difficult given the blur between industry-sponsored and user-generated content. Freedom of speech enables individuals to create and share content freely on social media. However, research suggests most content is industry-sponsored,³⁰ and our study is among the first to analyze industry-sponsored content only. Industry-sponsored material is likely to be highly calculated and carefully planned with specific keywords. In this study, we were able to predict which brand made a particular post based on the most frequently used words across all posts. Although not an aim of this study, future research might use this approach to help identify social media posts on social media that may be industry-sponsored, but are not explicitly identified as such, thereby rendering regulations regarding industry-sponsored material potentially more enforceable. Even if regulatory implications prove to be unfeasible, the analytic technique presented here can help identify key promotional strategies used by brands, so that effective counter-marketing campaigns can be developed.

Human Subjects Statement

The Committee for the Protection of Human Subjects at UTHealth, The University of Texas Health Science Center at Houston, in Houston (reference HSC-SPH-13-0343) approved this study.

Acknowledgements

Research reported in this paper was supported by grant number [1 P50 CA180906] from the National Cancer Institute and the FDA Center for Tobacco Products (CTP), and by grant, P2CHD042849, Population Research Center, awarded to the Population Research Center at The University of Texas at Austin by the Eunice Kennedy Shriver National Institute of Child Health and Human Development. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or the FDA.

References

1. Giovenco DP, Hammond D, Corey CG, et al. E-cigarette market trends in traditional US retail channels, 2012–2013. *Nicotine Tob Res.* 2015;17(10):1279-1283. [PubMed: 25542918]
2. US Food and Drug Administration, HHS. Deeming tobacco products to be subject to the Federal Food, Drug, and Cosmetic Act, as amended by the Family Smoking Prevention and Tobacco Control Act; restrictions on the sale and distribution of tobacco products and required warning statements for tobacco products. Final rule. *Fed Regist.* 2016;81(90):28973-29106. To be codified at 21 CFR § 1100, 1400, 1143. Available at: <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=3&cad=rja&uact=8&ved=0ahUKEwicoMWSzpLYAhVG2mMKHamwCqoQFggzMAI&url=https%3A%2F%2Fwww.federalregister.gov%2Fdocuments>

%2F2016%2F05%2F10%2F2016-10685%2Fdeeming-to-bacco-products-to-be-subject-to-the-federal-food-drug-and-cosmetic-act-as-amended-by-the&usg=AOvVaw1h-l-ZVQOYz50teUdfIRnZ. Accessed December 17, 2017. [PubMed: 27192730]

3. Yamin CK, Bitton A, Bates DW. E-cigarettes: a rapidly growing Internet phenomenon. *Ann Intern Med.* 2010;153(9):607-609. [PubMed: 21041581]
4. Richardson A, Ganz O, Vallone D. Tobacco on the web: surveillance and characterization of online tobacco and e-cigarette advertising. *Tob Control.* 2015;24(4):341-347. [PubMed: 24532710]
5. Chu KH, Sidhu AK, Valente TW. Electronic cigarette marketing online: a multi-site, multi-product comparison. *JMIR Public Health Surveill.* 2015;1(2):e11. [PubMed: 27227129]
6. Huang J, Kornfield R, Szczypka G, Emery SL. A cross-sectional examination of marketing of electronic cigarettes on Twitter. *Tob Control.* 2014;23(Suppl 3):iii26-iii30. [PubMed: 24935894]
7. Jo CL, Kornfield R, Kim Y, et al. Price-related promotions for tobacco products on Twitter. *Tob Control.* 2016;25:476-479. [PubMed: 26130449] :
8. Laestadius LI, Wahl MM, Cho YI. #Vapelife: an exploratory study of electronic cigarette use and promotion on Instagram. *Subst Use Misuse.* 2016;51(12):1669-1673. [PubMed: 27484191]
9. Chu KH, Allem JP, Cruz TB, Unger JB. Vaping on Instagram: cloud chasing, hand checks and product placement. *Tob Control.* 2017;26:575-578.:
10. Paek HJ, Kim S, Hove T, Huh JY. Reduced harm or another gateway to smoking? Source, message, and information characteristics of e-cigarette videos on YouTube. *J Health Commun.* 2014;19(5):545-560. [PubMed: 24117370]
11. de Andrade M, Hastings G, Angus K. Promotion of electronic cigarettes: tobacco marketing reinvented? *BMJ.* 2013;347:f7473. [PubMed: 24361526]
12. Grana RA, Ling PM. "Smoking revolution": a content analysis of electronic cigarette retail websites. *Am J Prev Med.* 2014;46(4):395-403. [PubMed: 24650842]
13. Zhu SH, Sun JY, Bonnevie E, et al. Four hundred and sixty brands of e-cigarettes and counting: implications for product regulation. *Tob Control.* 2014;23(Suppl 3):iii3-iii9. [PubMed: 24935895]
14. Padon AA, Maloney EK, Cappella JN. Youth-targeted e-cigarette marketing in the US. *Tob Regul Sci.* 2017;3(1):95-101. [PubMed: 28083545]
15. Fairchild AL, Bayer R, Colgrove J. The renormalization of smoking? E-cigarettes and the tobacco "endgame." *N Engl J Med.* 2014;370(4):293-295. [PubMed: 24350902]
16. Lenhart A, Duggan M, Perrin A, et al. *Teens, Social Media Technology Overview 2015.* Washington, DC: Pew Research Center; 2015.
17. Tsai WHS, Men LR. Motivations and antecedents of consumer engagement with brand pages on social networking sites. *J Interact Advert.* 2013;13(2):76-87.
18. Griffiths R, Casswell S. Intoxicogenic digital spaces? Youth, social networking sites and alcohol marketing. *Drug Alcohol Rev.* 2010;29(5):525-530.
19. Yoo W, Yang J, Cho E. How social media influence college students' smoking attitudes and intentions. *Comput Human Behav.* 2016;64:173-182. [PubMed: 27956757]
20. Elmore KC, Scull TM, Kupersmidt JB. Media as a "super peer": how adolescents interpret media messages predicts their perception of alcohol and tobacco use norms. *J Youth Adolesc.* 2017;46(2):376-387. [PubMed: 27837371]
21. Wilkinson AV, Vandewater EA, Carey FR, Spitz MR. Exposure to pro-tobacco messages and smoking status among Mexican origin youth. *J Immigr Minor Health.* 2014;16(3):385-393. [PubMed: 23584711]
22. Mantey DS, Cooper MR, Clendennen SL, et al. E-cigarette marketing exposure is associated with e-cigarette use among US youth. *J Adolesc Health.* 2016;58(6):686-690. [PubMed: 27080732]
23. Cohn E, Reininger B, Fisher-Hoch S, et al. Exposure to e-cigarette marketing and product use among Mexican American young adults on the US-Mexico border. *National Conference on Tobacco or Health*; 3 22-24, 2017; Austin, TX Available from: https://www.dropbox.com/sh/earh91lbjlt4tl/AACSD0gZdJRXjHifFeqfe22la?dl=0&preview=Cohn+EA%2C+Fisher-Hoch+SP%2C+Reininger+BM%2C+Perry+CL%2C+Wilkinson+AV_NC-TOH_March+2017.pdf. Accessed December 17, 2017.

24. Benoit K, Herzog A. Text analysis: estimating policy preferences from written and spoken words. In Bachner J, Ginsberg B, Hill KW, eds. *Analytics, Policy and Governance*. New Haven, CT: Yale University Press; 2017:137-159.
25. Foster D, Stine M, Liberman RA. Featurizing text: converting text into predictors for regression analysis. The Wharton School of the University of Pennsylvania, Philadelphia, PA; 2013 Available at: <http://www-stat.wharton.upenn.edu/~stine/research/regressor.pdf> Accessed December 20, 2017.
26. Turney PD, Pantel P. From frequency to meaning: vector space models of semantics. *J Artif Intell Res*. 2010;37:141-188.
27. Liang Y, Zheng X, Zeng DD, et al. Exploring how the tobacco industry presents and promotes itself in social media. *J Med Internet Res*. 2015;17(1):e24. [PubMed: 25608524]
28. Myslin M, Zhu SH, Chapman W, Conway M. Using Twitter to examine smoking behavior and perceptions of emerging tobacco products. *J Med Internet Res*. 2013;15(8):e174. [PubMed: 23989137]
29. Aphinyanaphongs Y, Lulejian A, Brown DP, et al. Text classification for automatic detection of e-cigarette use and use for smoking cessation from Twitter: a feasibility pilot. Paper presented at the Pacific Symposium on Bio-computing; 1 4-8, 2016; Kohala Coast, HI Available at: <https://psb.stanford.edu/previous/psb16/conference-materials/proceedings.pdf>. Accessed December 20, 2017.
30. Clark EM, Jones CA, Williams JR, et al. Vaporous marketing: uncovering pervasive electronic cigarette advertisements on Twitter. *PLoS One*. 2016;11(7):e0157304. [PubMed: 27410031]
31. Clark EM, Williams JR, Jones CA, et al. Sifting robotic from organic text: a natural language approach for detecting automation on Twitter. *J Comput Sci*. 2015;16:1-7.
32. Kavuluru R, Sabbir A. Toward automated e-cigarette surveillance: spotting e-cigarette proponents on Twitter. *J Comput Sci*. 2016;61:19-26.
33. Lazard AJ, Saffer AJ, Wilcox GB, et al. E-cigarette social media messages: a text mining analysis of marketing and consumer conversations on Twitter. *JMIR Public Health Surveill*. 2016;2(21):e171. [PubMed: 27956376]
34. Ayers JW, Leas EC, Allem J-P, et al. Why do people use electronic nicotine delivery systems (electronic cigarettes)? A content analysis of Twitter, 2012-2015. *PLoS One*. 2017;12(3):e0170702. [PubMed: 28248987]
35. Cole-Lewis H, Pugatch J, Sanders A, et al. Social listening: a content analysis of e-cigarette discussions on Twitter. *J Med Internet Res*. 2015;17(10):e243. [PubMed: 26508089]
36. Kim Y, Huang J, Emery S. Garbage in, garbage out: data collection, quality assessment and reporting standards for social media data use in health research, infodemiology and digital disease detection. *J Med Internet Res*. 2016;18(2):e41. [PubMed: 26920122]
37. Prochaska JJ, Pechmann C, Kim R, Leonhardt JM. Twitter= quitter? An analysis of Twitter quit smoking social networks. *Tob Control*. 2012;21(4):447-449. [PubMed: 21730101]
38. Rose SW, Jo CL, Binns S, et al. Perceptions of menthol cigarettes among Twitter users: content and sentiment analysis. *J Med Internet Res*. 2017;19(2).
39. van der Tempel J, Noormohamed A, Schwartz R, et al. Vape, quit, tweet? Electronic cigarettes and smoking cessation on Twitter. *Int J Public Health*. 2016;61(2):249-256. [PubMed: 26841895]
40. Chu KH, Unger JB, Allem JP, et al. Diffusion of messages from an electronic cigarette brand to potential users through Twitter. *PLoS One*. 2015;10(12):e0145387. [PubMed: 26684746]
41. Sanders-Jackson A, Brown CG, Prochaska J. Applying linguistic methods to understanding smoking related conversations on Twitter. *Tob Control*. 2015;24(2):136-138. [PubMed: 24227540]
42. Willis E, Haught MJ, Morris II DL. Up in vapor: exploring the health messages of e-cigarette advertisements. *Health Commun*. 2017;32(3):372-380. [PubMed: 27309130]
43. Phua J, Jin SV, Hahn JM. Celebrity-endorsed e-cigarette brand Instagram advertisements: effects on young adults' attitudes towards e-cigarettes and smoking intentions. *J Health Psychol*. 2017 2 1:1359105317693912 doi: 10.1177/1359105317693912 [Epub ahead of print]
44. Berg CJ. Preferred flavors and reasons for e-cigarette use and discontinued use among never, current, and former smokers. *Int J Public Health*. 2016;61(2):225-236. [PubMed: 26582009]
45. Pérez A, Harrell MB, Malkani RI, et al. Texas Adolescent Tobacco and Marketing Surveillance system's design. *Tob Regul Sci*. 2017;3(2):151-167. [PubMed: 29098172]

46. Loukas A, Chow S, Pasch KE, et al. College students' polytobacco use, cigarette cessation, and dependence. *Am J Health Behav.* 2016;40(4):514-522. [PubMed: 27338998]
47. Liping J, Ng MK, Huang, JZ. Knowledge-based vector space modeling for text clustering. *Knowl Inf Syst.* 2010;25(1):35-55.
48. Klecka WR. *Discriminant Analysis.* Newbury Park, CA: Sage Publications, Inc; 1980:1-73.
49. Midi H, Sarkar S, Rana S. Collinearity diagnostics of binary logistic regression model. *Journal of Interdisciplinary Mathematics.* 2010;13(3):253-267.
50. US Department of Health and Human Services (USD-HHS). *E-Cigarette Use Among Youth and Young Adults: A Report of the Surgeon General.* Atlanta, GA: USDHHS, US Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2016.
51. Allem JP, Ferrara E. The importance of debiasing social media data to better understand e-cigarette-related attitudes and behaviors. *J Med Internet Res.* 2016;18(8):e219. [PubMed: 27507563]

Figure 1. Word Clouds for the 25 Most Frequently Used Words for Each Brand

Note.

Top left word cloud = Blu, Top right word cloud = NJoy, Bottom left word cloud = Logic, Bottom Right word cloud = Metro.

Figure 2. Predicted Probabilities from Multinomial Logistic Regression with Blu as Reference Note.

Plotted predicted probabilities indicate how well the logistic regression model predicted which brand made a particular post based on the words used in that post. Blu posts are shown in blue, Logic in orange, Metro in purple, and NJoy in green.

Figure 3.
E-cigarette Brand Group Centroids on Discriminant Functions

Table 1
Multinomial Logistic Regression Predicting Brand with Blu as the Reference Category

Predictors:	NJOY		Logic		Metro	
	OR	(CI)	OR	(CI)	OR	(CI)
Battery(ies)	0.00	(0.00)	0.81	(−1.14,2.75)	2.90 ^{***}	(1.45,4.35)
Beach	0.04	(−0.95,1.04)	0.00	(0.00)	0.00	(0.00)
Buy	131.04 ^{***}	(128.61,133.48)	0.00	(0.00)	0.17	(−5.00,5.33)
Can	1.38	(0.68,2.09)	1.79 ^{***}	(1.22,2.36)	1.78 ^{***}	(1.11,2.45)
Cartridge(s)	0.00	(0.00)	0.00	(0.00)	4.57 ^{***}	(3.21,5.93)
Charger	0.001	(−314.57,314.57)	11.15 ^{***}	(8.34,13.96)	3.62 ^{***}	(11.35,15.89)
Cherry	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
Cigarette(s)	1.78	(0.74,2.83)	1.91 ^{***}	(1.14,2.68)	4.16 ^{***}	(3.42,4.89)
Coupon	0.00	(0.00)	0.00	(0.00)	502.63 ^{***}	(499.69,505.58)
Cuban	6.56e+5 ^{***}	(6.56e+5)	2.28e+11 ^{***}	(2.28e+11)	2.46e+4 ^{***}	(2.46e+4)
Customer	0.22	(−1.44,1.89)	4.86 ^{***}	(3.94,5.79)	0.00	(0.00)
Day	2.25 ^{***}	(1.75,2.75)	0.37	(−0.34,1.09)	0.27	(−0.96,1.50)
Disposable	0.00	(0.00)	0.26	(−1.29,1.81)	2.94 ^{***}	(1.89,3.98)
Ecig(s)	0.75	(0.23,1.26)	0.98	(0.56,1.40)	0.65	(0.12,1.18)
Ecigarette(s)	2.42e+12 ^{***}	(2.42e+12)	2.47e+13 ^{***}	(2.47e+13)	3.63e+12 ^{***}	(3.63e+12)
Electric	2.16e+5 ^{***}	(2.16e+5)	0.00	(0.00)	9.67	(−117.20,136.55)
Electronic	0.16	(−2.26,2.59)	3.21 ^{***}	(2.19,4.23)	8.14 ^{***}	(7.18,9.11)
Eliquid	1.29e+11 ^{***}	(1.29e+11)	1.80e+4 ^{***}	(1.80e+4)	380.24 ^{***}	(380.24)
Enjoy	1.11	(−1.24,3.46)	60.39 ^{***}	(58.44,62.35)	3.86 ^{***}	(1.93,5.80)
Espresso	4.38e+14 ^{***}	(4.38e+14)	0.06 ^{***}	(0.06)	1.30e+4 ^{***}	(1.30e+4)
Event	1.56	(0.83,2.30)	3.03 ^{***}	(2.51,3.55)	0.00	(0.00)
Flavor	0.83	(−0.46,2.13)	0.23	(−1.40,1.86)	1.82	(0.98,2.6)
Find	4.50 ^{***}	(3.33,5.66)	2.13	(0.98,3.28)	6.34 ^{***}	(5.27,7.41)

Predictors:	NJOY		Logic		Metro	
	OR	(CI)	OR	(CI)	OR	(CI)
Free	11.90 ^{***}	(10.76,13.03)	4.88 ^{***}	(3.61,6.16)	2.83 ^{***}	(1.26,4.40)
Freedom	0.12	(−1.93,2.18)	0.00	(0.00)	0.00	(0.00)
Friends	9.10 ^{***}	(7.88,10.33)	1.98	(0.30,3.67)	0.00	(0.00)
Get	2.97 ^{***}	(2.47,3.47)	1.50	(0.94,2.06)	0.37	(−0.57,1.30)
Good	1.59	(0.71,2.47)	0.40	(−0.82,1.62)	1.57	(0.63,2.51)
Happy	2.30 ^{***}	(1.75,2.86)	0.00	(0.00)	1.67	(0.90,2.44)
Kit	0.78	(−0.55,2.11)	0.00	(0.00)	0.45	(−1.80,2.70)
King	8.35e+19 ^{***}	(8.35e+19)	0.00	(0.00)	1.35e+11 ^{***}	(1.35e+11)
Lounge	0.00	(−101.49,101.49)	0.00	(0.00)	0.00	(−0.001,0.001)
Love	4.18 ^{***}	(3.45,4.92)	1.03	(0.07,2.00)	2.76 ^{***}	(1.90,3.62)
Menthol	1.20	(−0.47,2.86)	2.15 ^{***}	(1.01,3.29)	1.59	(0.48,2.70)
Never	0.50	(−1.34,2.35)	2.40	(0.87,3.93)	8.05 ^{***}	(6.72,9.39)
New	1.06	(0.42,1.70)	0.61	(−0.09,1.32)	0.70	(−0.11,1.50)
Switch	2.12e+15 ^{***}	(2.12e+15)	2.69e+14 ^{***}	(2.69e+14)	7.18e+14 ^{***}	(7.18e+14)
One	1.58 ^{***}	(1.09,2.07)	0.66	(0.03,1.28)	0.76 [*]	(0.02,1.50)
Order	7.74 ^{***}	(6.16,9.32)	0.00	(0.00)	0.35	(−2.90,3.60)
Pack	0.83	(−0.38,2.05)	1.42	(0.32,2.52)	1.38	(0.27,2.48)
Platinum	0.01 ^{***}	(0.01)	4.59e+29 ^{***}	(4.59e+29)	1.16e+11 ^{***}	(1.16e+11)
Premium	0.01	(−166.20,166.20)	23.24 ^{***}	(21.74,24.73)	0.00	(0.00)
Saturday	0.52	(−0.94,1.97)	0.00	(0.00)	0.00	(0.00)
Smoking	0.44	(−1.69,2.57)	3.17 ^{***}	(2.22,4.12)	4.48 ^{***}	(3.54,5.41)
Smoke(s)	2.23 ^{***}	(1.06,3.39)	2.96 ^{***}	(2.08,3.84)	5.12 ^{***}	(4.20,6.04)
Starter	0.05	(−2.59,2.69)	0.00	(0.00)	1.06	(−1.29,3.42)
Super	7.98e+13 ^{***}	(7.98e+13)	7.29e+9 ^{***}	(7.29e+9)	45.38 ^{***}	(45.37,45.38)
Today	2.57 ^{***}	(1.97,3.16)	0.11	(−1.27,1.48)	1.53	(0.74,2.32)

Predictors:	NJOY		Logic		Metro	
	OR	(CI)	OR	(CI)	OR	(CI)
Tobacco	1.20	(−0.02,2.43)	6.10 ^{***}	(5.27,6.94)	1.15	(0.23,2.06)
Using	0.60	(−1.57,2.77)	7.77 ^{***}	(6.83,8.71)	4.82 ^{***}	(3.61,6.02)
Use	1.10	(−0.07,2.27)	17.73 ^{***}	(17.05,18.42)	2.09	(0.99,3.20)
Vanilla	1.72	(0.11,3.34)	0.89	(−1.66,3.44)	0.00	(0.00)
Vape	43.13 ^{***}	(42.41,43.84)	0.10	(−2.53,2.72)	8.31 ^{***}	(7.35,9.28)
Vapor	0.00	(−660.10,660.10)	3.02 ^{***}	(2.26,3.79)	0.49	(−0.73,1.70)
Vaping	0.07	(−0.59,0.73)	0.00	(−2.00,2.01)	0.25	(−0.42,0.93)
Vapelife	6.07 ^{***}	(5.36,6.78)	0.00	(0.00)	0.16	(−1.90,2.22)
Vapelyfe	3.59e+20 ^{***}	(3.59e+20)	2.56e+5 ^{***}	(2.56e+5)	7.40e+9 ^{***}	(7.40e+9)
Vaporlounge	0.00	(0.00)	0.00	(0.00)	0.00	(0.00)
World	22.23 ^{***}	(21.38,23.08)	1.59	(0.33,2.86)	0.73	(−1.47,2.93)
Constant	0.18 [*]	(0.02,0.34)	0.34 ^{***}	(0.22,0.47)	0.16	(−0.01,0.33)

^{*}
p < .05;

^{**}
p < .01;

^{***}
p < .001

Note.

Single CI values within parentheses indicate identical upper and lower CI bounds.

Blu is the reference category. Each “word” predictor reflects the frequency with which that word appeared across all 5022 posts. Odds ratios (OR) indicate that for each increase in the use of the word, the odds that the post was generated from NJOY, Logic or Metro, relative to Blu, changes by a factor of x, where x is the odds ratio.

Table 2
Discriminant Function Analysis Predicting E-cigarette Brand

	Total Structure Coefficients	Standardized Structure Coefficients
Function 1: Lounging		
Lounge	.47	.63
Beach	.29	.54
Vapor	.32	.03
Ecig(s)	.20	.02
Function 2: Vaping		
Vape	.52	.47
Vapelite	.32	.17
Vapelyfe	.27	.15
World	.20	.27
Ecigarette(s)	-.27	-.23
Electronic	-.29	-.16
Super	.20	.02
Tobacco	-.25	-.09
Function 3: Device		
Cartridge(s)	.41	.30
Battery(ies)	.39	.20
Coupon	.38	.49
Charger(s)	.32	.15
Premium	-.24	-.62