

AJPM FOCUS

INCLUSIVITY IN PEOPLE, METHODS, AND OUTCOMES

PILOT DATA ANALYSIS

An Investigation of Age-Differentiated Conversations About Electronic Nicotine Delivery Systems on Reddit



Mario A. Navarro, PhD,¹ Andrea Malterud, PhD,¹ Zachary P. Cahn, PhD,¹ Laura Baum, MA, MPH,² Thomas Bukowski, MPA, MA,² Caroline Kery, MS,³ Robert F. Chew, MS,³ Annice E. Kim, PhD²

Introduction: This study analyzes age-differentiated Reddit conversations about ENDS.

Methods: This study combines 2 methods to (1) predict Reddit users' age into 2 categories (13–20 years [underage] and 21–54 years [of legal age]) using a machine learning algorithm and (2) qualitatively code ENDS-related Reddit posts within the 2 groups. The 25 posts with the highest karma score (number of upvotes minus number of downvotes) for each keyword search (i.e., query) and each predicted age group were qualitatively coded.

Results: Of 9, the top 3 topics that emerged were flavor restriction policies, Tobacco 21 policies, and use. Opposition to flavor restriction policies was a prominent subcategory for both groups but was more common in the 21–54 group. The 13–20 group was more likely to discuss opposition to minimum age laws as well as access to flavored ENDS products. The 21–54 group commonly mentioned general vaping use behavior.

Conclusions: Users predicted to be in the underage group posted about different ENDS-related topics on Reddit than users predicted to be in the of-legal-age group.

AJPM Focus 2023;2(1):100045. Published by Elsevier Inc. on behalf of The American Journal of Preventive Medicine Board of Governors. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

INTRODUCTION

Electronic nicotine delivery systems (ENDSs) are the most commonly used tobacco product among U.S. youth, with an estimated 4.43 million high school students and 860,000 middle school students having ever used an ENDS product as of 2021.¹ Youth ENDS use increased rapidly after 2016, in part due to the appeal of flavored products.² By 2019, the average number of days that high school students used nicotine tobacco products had nearly doubled in just 2 years,³ and National Youth Tobacco Survey estimated that past-30 day use of any tobacco product reached its highest level since 2000.³ In addition to the addictive properties of nicotine, reviews have identified several other harms and potential harms associated with ENDS use, including inhalation of toxins and decreases in lung function.^{4–6} Because the ENDS

product landscape is rapidly changing,⁷ social media listening provides unique methodologies to obtain rapid insights and surveillance on product discussions.

Recent qualitative studies using social media for tobacco prevention and control research rely heavily on thematic coding and content analysis of posted material.

From the ¹Office of Health Communication and Education, Center for Tobacco Products, Food and Drug Administration (FDA), Silver Spring, Maryland; ²Center for Health Analytics, Media, and Policy, RTI International, Research Triangle Park, North Carolina; and ³Center for Data Science, RTI International, Research Triangle Park, North Carolina

Address correspondence to: Mario A. Navarro, PhD, Office of Health Communication and Education, Center for Tobacco Products, Food and Drug Administration (FDA), 10903 New Hampshire Avenue, Silver Spring MD 20993. E-mail: Mario.Navarro@fda.hhs.gov.

2773-0654/\$36.00

<https://doi.org/10.1016/j.focus.2022.100045>

For example, Wang and colleagues posted on the social media site, Reddit, to investigate ENDS flavor mentions,⁸ and Brett and colleagues coded Reddit posts to find influences and barriers to use and perceptions of JUUL.⁹ However, the lack of publicly available demographic information on users is a limitation of social media data and may prevent researchers from understanding at-risk audiences via this route.¹⁰ To better understand conversations of tobacco public education target audiences on Reddit, Chew et al.¹¹ developed an algorithm that examines users' posts and metadata to predict and categorize Reddit users' ages into 1 of 2 groups: 13–20 years (i.e., underage [UA]) or 21–54 years (i.e., of legal age [OLA]). These 2 age groups were used to separate users' legal use of tobacco products and to provide an appropriate model because there were very few age references for those aged >54 years. This exploratory study, using the Chew and colleagues¹¹ algorithm, investigates ENDS conversations, with a focus on flavor restriction and Tobacco 21 policy discussions for posts originating from predicted UA and OLA groups.

METHODS

Study Population

Figure 1 summarizes the 3 overarching steps of identification undertaken in this study. First, Reddit posts about vaping in general, flavor restriction policies, and Tobacco 21 policies were identified and downloaded from Brandwatch.com, a social media listening platform. Multiple search keywords were used to identify relevant posts about general vaping (e.g., vape, vaping, E-cigarette), flavor restriction policies (e.g., flavor policy), and Tobacco 21 policies (e.g., minimum [min] age laws and tobacco-related words such as cigarettes, vapes, and cigars). These keyword groups formed 3 separate queries to pull the data. Searches were also restricted to English language-only posts.

Measures

A previously developed age prediction model was used to predict the age group for each author as either UA (13–20 years), OLA (21–54 years), or uncertain.¹¹ These categories were used to examine the differences in conversations depending on whether the user was OLA to use tobacco. The lower bound was selected

because Reddit users must be aged ≥ 13 years, and those aged >54 years could not be appropriately classified because of the small number of individuals who fell into this category during the development of the model. The age prediction model uses the gradient-boosted trees algorithm¹² to predict the probability that each user belongs to either the UA or OLA age groups. Analogous to logistic regression, predicted probabilities are generated by multiplying the trained model weights by the input variable values for each new observation, summing them together, and applying an inverse logit transformation. There are 15 input variables required for the model to generate predictions, spanning literary characteristics (e.g., sentences per comment) to subreddit posting frequencies (e.g., "proportion of user's posts or comments in the *r/teenagers* subreddit"). A full list of the variables used in the model, as well as further background on other variables considered, variable importance, and model performance, can be found in Chew et al.¹¹ Because the model does not produce perfect predictions (test set F1 score, ~ 0.79), we reduced the likelihood that the model returned false positives by only considering predictions with a predicted probability > 0.6 for either age group. This process of rejecting predictions for which the model is most uncertain is referred to as classification with a reject option¹³ in the literature. After applying the age prediction model to the posts from each query, we selected the 25 posts in each predicted age group and query with the highest karma scores (number of upvotes – number of downvotes). This resulted in 150 total posts across both age groups and 3 queries.

Data Analysis

Two coders were trained using a standardized codebook, and after achieving sufficient inter-rater reliability (percentage agreement reached at least 70%), they independently coded the study sample. All themes listed in the Results section were the themes in the codebook. Not all themes were present; more information is provided in the Results. Posts were excluded if they mentioned marijuana/tetrahydrocannabinol/cannabidiol, were not in the English language, or were not relevant to E-cigarettes.

RESULTS

Descriptive Statistics

Eighteen posts were excluded from the predicted UA group, and 24 posts were excluded from the predicted OLA group, leaving 57 UA (general vaping: 18, flavor restriction policies: 18, Tobacco 21 policies: 21) and 51

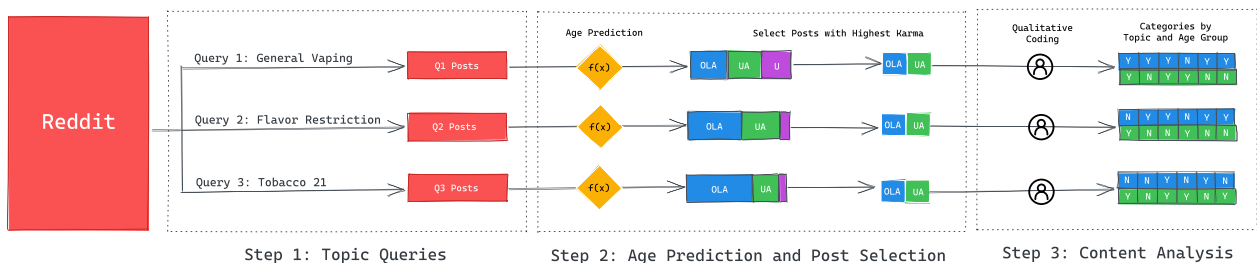


Figure 1. Study design flow chart.

OLA (general vaping: 13, flavor restriction policies: 17, Tobacco 21 policies: 21) posts. For each query, the range of karma scores for coded posts was large, suggesting that most highly engaged posts (i.e., high karma scores) were captured (general vaping: predicted UA group: mean=3,715, min=1,212, maximum [max]=8,327; predicted OLA group: mean=1,071, min=553, max=2,352; flavor restriction policy: predicted UA group: mean=476, min=42, max=5,837; predicted OLA group: mean=438, min=248, max=1,188; Tobacco 21 policy: predicted UA group: mean=62, min=17, max=376; predicted OLA group: mean=185, min=20, max=1,259).

Post Categories

Table 1 reports the frequency and percentages of each post code category and subcategory. Coding categories included flavor restriction policies, access, Tobacco 21 policies, use, motivations for vaping, harm perceptions, products, memes/jokes, coronavirus disease 2019 (COVID-19), barriers to vaping, campaigns by the Center for Tobacco Products, and other. Barriers to vaping and campaigns by the Center for Tobacco Products did not emerge as code categories, even though they were originally in the codebook.

For both UA and OLA groups, the categories of flavor restriction policies and Tobacco 21 policies were the most prominent (>40%). Between the 2 groups, the products and memes/jokes categories were more prominent for UA than for OLA. The categories of use and harm perceptions were more prominent for OLA.

Demonstrating nuances between the groups, subcategory differences continued between predicted age groups. For flavor restriction policies, opposition was a primary subcategory for both predicted age groups, but many flavor restriction posts fell into the other subcategory for the UA group and skepticism for the OLA group. To clarify, *Opposition* was defined as “voicing clear opposition or encouraging work against an ordinance,” and *skepticism* was defined as “doubt about the motives behind or effectiveness of an ordinance.” Posts coded as other were dominated by news stories in both groups. The OLA group had nearly twice as many opposition codes as the UA group, and the second most common codes for UA were links to news stories. A clear distinction between the groups is that the OLA group showed greater opposition and skepticism to flavor restriction policies.

For the Tobacco 21 policy category, a similar pattern emerged for the UA and OLA groups, with opposition, skepticism, and the other subcategories dominating the conversation, although for this topic, the UA group showed greater opposition and skepticism, whereas the OLA group posted mostly other-category news links.

Table 1. Postcategory Prevalence in Both Predicted Underage and Of-Legal-Age Post Authors

Postcategory or subcategory	Underage, n (%)	Of legal age, n (%)
Flavor restriction policies ^a	26 (45.61)	37 (72.54)
Support	0 (0)	1 (2.70)
Oppose	9 (34.62)	17 (45.95)
Skepticism	1 (3.85)	8 (21.62)
Access	4 (15.38)	2 (5.41)
Switching	3 (11.53)	5 (13.51)
Quitting	1 (3.85)	0 (0)
Other	8 (30.77)	4 (10.81)
Tobacco 21 policies ^b	27 (47.37)	21 (41.18)
Support	1 (3.71)	0 (0)
Oppose	8 (29.63)	4 (19.04)
Skepticism	4 (14.81)	1 (4.76)
Legacy Clause	4 (14.81)	0 (0)
Access	2 (7.41)	2 (9.52)
Switching	1 (3.70)	1 (4.77)
Quitting	0 (0)	0 (0)
Other	7 (25.93)	13 (61.91)
Use ^c	11 (19.30)	22 (43.14)
Dual	1 (9.10)	0 (0)
Switching	0 (0)	2 (9.09)
Quitting	2 (18.18)	1 (4.55)
Vape terms	5 (45.45)	11 (50.00)
Other	3 (27.27)	8 (36.36)
Motivations for vaping ^d	9 (15.79)	9 (17.65)
Harm perceptions ^e	2 (3.51)	13 (25.49)
Products ^f	18 (31.58)	8 (15.69)
Memes/jokes ^g	17 (29.82)	5 (9.80)
COVID-19 ^h	1 (1.75)	6 (11.76)
Other ⁱ	1 (1.75)	3 (5.88)

Note: Subcategory percentages are derived from the percent total of the parent code.

^aAny post referencing flavor restriction policies.

^bAny post referencing general mention of Tobacco 21 policies.

^cAny mention of other vaping use behaviors, including mentions of using vaping to quit cigarettes or other tobacco products.

^dAny mention or discussion of why someone vapes (e.g., makes them feel relaxed, to escape, for fun). Includes noting the motives of other users.

^eAny noted harms or perceived harms associated with vaping.

^fDescriptions, reviews, or questions about a vaping product.

^gVaping related photo/GIF memes or jokes

^hAny general mention of COVID-19 related to vaping.

ⁱAny other conversations related to vaping tobacco not covered in the other codes.

GIF, graphics interchange format.

For the UA group, a subcategory code emerged that detailed the desire to allow ENDS users aged 18–20 years, who were previously able to use ENDS products, to continue having the ability to purchase ENDS products (*legacy clause*, sometimes referred to by posters as grandfather clause). Within the subcategories of use, the vape terms subcategory was the most prominent for

both groups. These terms consisted of vapes, vaping, vape master, ripping, and Juuling for the predicted UA posts and fire up your rig, e-liquid, ejuice, nic juice, coils, and pod system for OLA posts.

For motivations for vaping, the primary motivation mentioned for both UA and OLA posts was the desire to avoid cigarettes. Harm perception posts were primarily identified for the OLA group and ranged in topic from vaping-related illnesses to feeling better after quitting. The product category was primarily made up of brand names. For the UA group, this brand was exclusively JUUL, but the OLA group included others such as Lava Pods. Memes/jokes emerged predominantly among UA posts and included visual jokes for various sorts of media and contained jokes mocking vaping. COVID-19 information, in the form of news articles, was discussed mostly within OLA posts. The other-category posts were more prominent for OLA posts and consisted of an individual's personal relationship with vaping, usually with a form of judgment.

DISCUSSION

This mixed methods analysis of Reddit posts provided insight into ENDS online conversations by differentiating conversations by 2 predicted age groups (13–21 and 21–54 years). Differences between predicted age groups emerged for both frequency of code categories and more specific content within categories. Posts were coded into the categories of flavor restriction policies, Tobacco 21 policies, use, motivations for vaping, harm perceptions, products, memes/jokes, COVID-19, and other. Looking at the subcategories, a more nuanced story emerged such that most posts for the UA group fell into the other category, and skepticism posts were most prevalent for the OLA group. A similar pattern emerged for the Tobacco 21 policy category. One differentiating subcategory for the Tobacco 21 policy category was the legacy clause code for UA posts. This study aligns with previous research, which found age restriction opposition by UA Reddit users, using E-cigarettes to avoid cigarettes,⁹ and significant discussion about JUUL¹⁴ and flavor access.^{8,15} Findings are in line with recent ENDS studies that show Twitter users' positive sentiment toward flavors.¹⁵

Future Directions

This study has implications for future research and for public health surveillance. The mixed methodologies (i.e., data science models and qualitative coding) used in this study can be applied to a vast number of public health topics. In addition, age algorithms have been applied to other platforms in the past and there can be an expansion

of the platforms that are analyzed.¹⁶ Finally, automated data science methodologies (e.g., topic modeling) could provide a way to autocategorize posts, making it easier to provide thematic analyses for large amounts of data and provide a more rapid form of surveillance.

Limitations

This study had several limitations. The keywords used in the query did not reflect all the relevant keywords that could differentiate between posts written by UA and OLA users of Reddit. Relevant posts could have been missed. Although a sample of the top 25 most engaged posts was used, the sample size is still fairly small. This may limit the generalizability of the results. Because the sample was small, it was inappropriate to conduct any statistical analyses.

CONCLUSIONS

Reddit posts provide a robust public access data source that can be used by researchers.^{8,10,11} This study used a combination of methodologies to paint a picture of the current ENDS landscape on Reddit. Differences were found across all the 3 queries (i.e., general vaping, flavor restriction policies, and Tobacco 21 policies). These differences highlight the importance of using a combination of classification tools and qualitative coding that allows researchers and public health professionals to better understand perceptions and knowledge, attitudes, and beliefs about a product to develop more targeted messaging.

ACKNOWLEDGMENTS

This publication represents the views of the author(s) and does not represent Food and Drug Administration Center for Tobacco Products position or policy.

This work was funded by contract with the Center for Tobacco Products, U.S. Food and Drug Administration, U.S. Department of Health and Human Services (number [HHSF223201510002B-Order #75F40119F19020](#)).

Declarations of interest: none.

CREDIT AUTHOR STATEMENT

Mario A. Navarro: Conceptualization, Methodology, Supervision, Writing – original draft, Writing – review and editing. Andrea Malterud: Methodology, Writing – original draft, Writing – review and editing. Zachary P. Cahn: Writing – original draft, Writing – review and editing. Laura Baum: Formal analysis, Project administration, Methodology, Writing – original draft, Writing – review and editing. Thomas Bukowski: Data curation, Formal analysis, Software, Writing – original draft, Writing – review and editing. Caroline Kery: Data curation, Formal analysis, Writing – review and editing. Robert F. Chew: Data curation, Formal analysis, Software, Writing – original draft, Writing – review and editing. Annice E. Kim: Conceptualization, Supervision, Writing – review and editing.

REFERENCES

- Gentzke AS, Wang TW, Cornelius M, et al. Tobacco product use and associated factors among middle and high school students – National Youth Tobacco Survey, United States, 2021. *MMWR Surveill Summ*. 2022;71(5):1–29. <https://doi.org/10.15585/mmwr.ss7105a1>.
- Groom AL, Vu TT, Kesh A, et al. Correlates of youth vaping flavor preferences. *Prev Med Rep*. 2020;18:101094. <https://doi.org/10.1016/j.pmedr.2020.101094>.
- Sun R, Mendez D, Warner KE. Trends in nicotine product use among U.S. adolescents, 1999–2020. *JAMA Netw Open*. 2021;4(8):e2118788. <https://doi.org/10.1001/jamanetworkopen.2021.18788>.
- Domenico L, DeRemer CE, Nichols KL, et al. Combatting the epidemic of E-cigarette use among students and transitional-age youth. *CPSP*. 2021;10(1):5–16. <https://doi.org/10.2174/221155600999200613224100>.
- Singh S, Windle SB, Filion KB, et al. E-cigarettes and youth: patterns of use, potential harms, and recommendations. *Prev Med*. 2020;133:106009. <https://doi.org/10.1016/j.ypmed.2020.106009>.
- Cahn Z, Drope J, Douglas CE, et al. Applying the population health standard to the regulation of electronic nicotine delivery systems. *Nicotine Tob Res*. 2021;23(5):780–789. <https://doi.org/10.1093/ntr/ntaa190>.
- Owotomo O, Walley S. The youth E-cigarette epidemic: updates and review of devices, epidemiology and regulation. *Curr Probl Pediatr Adolesc Health Care*. 2022;52(6):101200. <https://doi.org/10.1016/j.cppeds.2022.101200>.
- Wang L, Zhan Y, Li Q, Zeng DD, Leischow SJ, Okamoto J. An examination of electronic cigarette content on social media: analysis of E-cigarette flavor content on Reddit. *Int J Environ Res Public Health*. 2015;12(11):14916–14935. <https://doi.org/10.3390/ijerph121114916>.
- Brett EI, Stevens EM, Wagener TL, et al. A content analysis of JUUL discussions on social media: using Reddit to understand patterns and perceptions of JUUL use. *Drug Alcohol Depend*. 2019;194:358–362. <https://doi.org/10.1016/j.drugalcdep.2018.10.014>.
- Sharma R, Wigginton B, Meurk C, Ford P, Gartner CE. Motivations and limitations associated with vaping among people with mental illness: a qualitative analysis of Reddit discussions. *Int J Environ Res Public Health*. 2016;14(1):7. <https://doi.org/10.3390/ijerph14010007>.
- Chew R, Kery C, Baum L, Bukowski T, Kim A, Navarro M. Predicting age groups of Reddit users based on posting behavior and meta-data: classification model development and validation [published correction appears in *JMIR Public Health Surveill*. 2021;7(4):e30017]. *JMIR Public Health Surveill*. 2021;7(4):e25807. <https://doi.org/10.2196/25807>.
- Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Statist*. 2001;29(5):1189–1232. <https://doi.org/10.1214/aos/1013203451>.
- Herbei R, Wegkamp MH. Classification with reject option. *Can J Statistics*. 2009;34(4):709–721. <https://doi.org/10.1002/cjs.5550340410>.
- Zhan Y, Zhang Z, Okamoto JM, Zeng DD, Leischow SJ. Underage JUUL use patterns: content analysis of Reddit messages. *J Med Internet Res*. 2019;21(9):e13038. <https://doi.org/10.2196/13038>.
- Lu X, Chen L, Yuan J, et al. User perceptions of different electronic cigarette flavors on social media: observational study. *J Med Internet Res*. 2020;22(6):e17280. <https://doi.org/10.2196/17280>.
- Kim AE, Chew R, Wenger M, et al. Estimated ages of JUUL Twitter followers [published correction appears in *JAMA Pediatr*. 2019;173(7):704]. *JAMA Pediatr*. 2019;173(7):690–692. <https://doi.org/10.1001/jamapediatrics.2019.0922>.