

Applying topic modelling and qualitative content analysis to identify and characterise ENDS product promotion and sales on Instagram

Neal Shah,^{1,2} Matthew Nali,^{2,3} Cortni Bardier,^{2,4} Jiawei Li,² James Maroulis,⁴ Raphael Cuomo ,^{2,3} Tim K Mackey ^{1,2,4}

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/tobaccocontrol-2021-056937>).

¹Department of Healthcare Research and Policy, University of California San Diego, La Jolla, California, USA

²Global Health Policy and Data Institute, San Diego, California, USA

³Department of Anesthesiology, University of California San Diego School of Medicine, La Jolla, California, USA

⁴Global Health Program, Department of Anthropology, University of California San Diego, La Jolla, California, USA

Correspondence to

Professor Tim K Mackey, Global Health Policy and Data Institute, San Diego, USA; tmackey@ucsd.edu

Received 22 July 2021

Accepted 16 November 2021

ABSTRACT

Background Increased public health and regulatory scrutiny concerning the youth vaping epidemic has led to greater attention to promotion and sales of vaping products on social media platforms.

Objectives We used unsupervised machine learning to identify and characterise sale offers of electronic nicotine delivery systems (ENDS) and associated products on Instagram. We examined types of sellers, geographic ENDS location and use of age verification.

Methods Our methodology was composed of three phases: data collection, topic modelling and content analysis. We used data mining approaches to query hashtags related to ENDS product use among young adults to collect Instagram posts. For topic modelling, we applied an unsupervised machine learning approach to thematically categorise and identify topic clusters associated with selling activity. Content analysis was then used to characterise offers for sale of ENDS products.

Results From 70 725 posts, we identified 3331 engaged in sale of ENDS products. Posts originated from 20 different countries and were roughly split between individual (46.3%) and retail sellers (43.4%), with linked online sellers (8.8%) representing a smaller volume. ENDS products most frequently offered for sale were flavoured e-liquids (53.0%) and vaping devices (20.5%). Online sellers offering flavoured e-liquids were less likely to use age verification at point of purchase (29% vs 64%) compared with other products.

Conclusions Instagram is a global venue for unregulated ENDS sales, including flavoured products, and access to websites lacking age verification. Such posts may violate Instagram's policies and US federal and state law, necessitating more robust review and enforcement to prevent ENDS uptake and access.

INTRODUCTION

Given the near ubiquitous use of social media, particularly among youth and young adults, popular platforms such as Facebook, YouTube, TikTok and Instagram have emerged as powerful digital ecosystems to advertise, market and promote various health-related products and services. This includes a growing digital marketplace for tobacco and electronic nicotine delivery systems (ENDS), where various industry actors use social media to directly engage with individual users, including creating promotional product campaigns and using influencers as brand ambassadors to influence appeal and uptake.¹ Instagram, a photo and video-sharing

social media service owned and operated by Facebook, is particularly important given its widespread popularity among priority ENDS populations, with 71% of young adults (aged 18–29 years old) reporting they actively use the platform.²

User interaction on Instagram is facilitated by generation of 'likes' and 'comments' to Instagram posts depicting images or videos on a variety of topics and product promotional areas. Each post can be accompanied with a text description of the image/video and hashtags that help curate content. Specific hashtags can be queried by users, and public posts containing the specific hashtag will be returned as results. This enables users to search for and interact with curated content on ENDS (eg, vaping lifestyle posts or user reviews of ENDS products), both prospectively via a user's account feed and retrospectively through searching for hashtags.³ The format of the platform and the demographics of its user base have made it an extremely popular platform for advertising, including specifically for ENDS products.³

In fact, aggressive ENDS marketing via social media platforms, including on Twitter, Instagram and YouTube, have specifically targeted adolescents and young adults.^{4,5} For example, ENDS products on Instagram have been actively advertised through the use of branded content, where influential accounts have been paid by companies to post an image promoting a vaping product or the general vaping lifestyle along with product placement.⁶ JUUL spent more than \$1 million to market its products through social media campaigns and also hired influencers for product promotion prior to removing their ads and suspending select accounts due to public scrutiny.⁷ A randomised controlled trial demonstrated that youth and young adults gave significant visual attention to influencer posts promoting e-cigarettes even when disclosed with '#sponsored' and '#ad', suggesting that branded content was an effective way of advertising ENDS to younger people.⁶

Crucially, ENDS direct-to-consumer advertising can lead to broad dissemination of pro tobacco/vaping content and generate positive user sentiment, leading to greater sales of these products.^{3,8} This specifically includes exposure to ENDS social media-based marketing that can in turn lead to greater online purchasing and acquisition behaviour, with a recent online national cross-sectional survey of young adult e-cigarette users finding that online purchasing has become the predominant source



© Author(s) (or their employer(s)) 2021. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Shah N, Nali M, Bardier C, et al. *Tob Control* Epub ahead of print: [please include Day Month Year]. doi:10.1136/tobaccocontrol-2021-056937

of self-reported ENDS point-of-purchase behaviour.⁹ Online purchasing of ENDS represents its own unique risk environment, particularly for youth and adolescents due to lack of sufficient regulation. For example, findings from qualitative interviews with young adults have found that ENDS online sales rarely involved age verification.¹⁰

Federal and state regulators have taken notice of the growing popularity of ENDS among youth and young adult populations, giving rise to concerns about short-term and long-term health impacts.¹¹ The US Food and Drug Administration (FDA) requires preapproval for manufacturing and sale of vaping products appealing to youth, prioritising enforcement against flavoured e-liquids.¹² States and tribes have also taken strong steps to address the ‘vaping epidemic’, including enacting their own sales bans of certain ENDS products.¹¹ In response to regulatory action, activist pressure and the growing number of youth users, Facebook and Instagram enacted restrictive marketing measures in their advertising policy in December 2019, prohibiting content aimed at selling tobacco products in sponsored ads and through influencers. Specifically, Facebook and Instagram ads ‘must not promote the sale or use of tobacco products and related paraphernalia’.¹² This policy focuses on restricting ads from retailers and also use of certain images but does not address user-generated content that are not sponsored ads.¹³

Previous research examining ENDS products promoted on Instagram have examined changes in content characteristics of ENDS-related posts, classification of advertisement-themed vaping images and image classification of vaping posts using deep learning.^{14–17} However, no studies to our knowledge have specifically assessed the characteristics of ENDS online sales activity on Instagram and whether it involves regulated and unregulated sellers in potential violation of federal, state and platform specific policies. In this study, we use an unsupervised machine learning and content analysis approach to identify and characterise offers for sale of ENDS on Instagram including types of sellers, types of products, geographic location and use of age-verification processes.

METHODS

The study was carried out in three distinct phases using approaches in data mining and processing, machine learning and content analysis. The first stage involved collection of publicly available ENDS product relevant posts on Instagram. The second phase used unsupervised machine learning to thematically identify and categorise Instagram posts and comments that were associated with ENDS sales. The third and final phase conducted content analysis to assess ENDS selling characteristics of interest.

Data collection and processing

The data collected in this study are publicly available posts collected from the Instagram platform between August 2019 and August 2020. On Instagram, each post may contain an image, text description and specific hashtags associated with the image/video. Through Instagram’s internal search engine, querying specific hashtags will display publicly available posts with the corresponding hashtag. We searched for posts using hashtag queries specifically related to youth vaping based on our own manual searches and hashtags used in previous studies (full list of hashtags queried are available in the online supplemental file 1.¹⁸) Posts were then automatically collected using a web scraper created in the computer programming language Python using the Beautiful Soup library package. This web scraper collected publicly available Instagram post text descriptions and the link to the original post, along with related metadata such as the account name and date from the posts returned by our hashtag queries. Posts were then stored in a JavaScript Object Notation standard file format, which contains the text of the post, username of the account, post creation time and comments of the post. For the purposes of this study, only the primary (parent) post was analysed (ie, user comments to posts were excluded from analysis), and metadata of accounts was used to assess select selling characteristics (see table 1).

Data analysis using unsupervised machine learning

In order to process large volumes of unstructured data, we used the Biterm Topic Model (BTM) to thematically group and

Table 1 ENDS seller characteristics and classifications

Classifications	Categories	Description
Type of seller	Individual sellers, online sellers and retail sellers	Individual sellers: accounts providing information about how to contact them through closed/encrypted communication applications (eg, direct messaging via the Instagram platform, WhatsApp, Telegram, Wickr and Kik) to enter into a transaction. Online sellers: accounts offering the sale of vaping products through external hyperlinked websites or hyperlinked e-commerce platforms Retail sellers: accounts offering the sale of vaping products associated with physical store location.
ENDS product categories	Flavouring products, vaping device products, vaping accessories and other associated products	Flavouring products: ENDS products that provide flavours/aromas for vaping including flavoured e-liquids and nicotine salts. ENDS devices: devices used to vape including those identified as e-cigarettes, vaping pens, disposable vapes such as flavoured pods and larger vaping devices such as vape mods or boxes. ENDS-related accessories: products required for vaping that are not flavours or the devices themselves.
Source and origin of sellers	Location of seller post WHOIS data on location of domain of external website	Location of post: purported location of a seller was determined based on the physical location or address provided and self-reported on Instagram user accounts, the language of the post, or hashtags and account information providing location information. Accounts with location information were reported by country, and those accounts without location information were categorised as ‘unknown’. WHOIS data: For Instagram posts that included online sellers, we identified publicly available information on domain registration locations of sellers based on querying the ICANN WHOIS database (a publicly accessible database of the registration information of internet domain names).
Age verification	Point-of-entry verification Point-of-purchase verification	Point-of-Entry (PoE) verification: identified if website asked for confirmation of age at point-of-entry of website (eg, home page) Point-of-Purchase (PoP) verification: identified if website asked for proof of age documentation on purchase (eg, in e-commerce shopping cart)

ENDS, electronic nicotine delivery systems.

characterise Instagram posts based on textual content. BTM is an unsupervised machine learning approach based on natural language processing that is used to extract themes from short-text content.¹⁹ Unsupervised machine learning using natural language processing and topic modelling do not require a pretrained or prelabelled (human coded) training dataset, while supervised machine learning algorithms involve use of a subset of data for training that is used to further classify similar content. BTM has previously been used in academic research as a means of unsupervised topic exploration, particularly effective when existing training data to conduct supervised machine learning for classification tasks are unavailable or a particular topic is unexplored. It can also generate broader insights into the content of the full corpus of data in comparison with keyword searches or filtering datasets for certain terms, particularly in the context of helping to filter out irrelevant posts that have similar keywords but are not relevant to a certain research topic or area of exploration. BTM has successfully been used to conduct content analysis on a variety of underexplored public health topics on social media platforms, including opioid use disorder, drug diversion, COVID-19 and exploring tobacco and vaping user self-reported behaviour.^{18 20–22}

We used a bag-of-words model to characterise processed text data into a specific vector, building a dictionary of all seen words in our corpus of Instagram posts. With this vectorised short text dictionary, BTM places a discrete probability distribution for each word into a predetermined set of topics, placing a greater weight on words that are most representative of an outputted topic. Therefore, the entire corpus of ENDS filtered Instagram posts were transformed into (k) distinct topic clusters consisting of posts representing similar themes. Coders then selected three specific BTM topic clusters that included high concentrations of word groupings that were highly relevant to our study aims. Clusters chosen for analysis specifically contained words related to the advertising, buying and selling of ENDS products (eg, common selling arguments and queues similar to those used in other studies, see online supplemental file for details on content coding and topic model evaluation).²¹

We also further filtered the dataset for keywords related to contact and purchasing information in order to specifically identify posts marketing and selling ENDS products with specific transaction details. Transaction keywords primarily referred to encrypted messaging applications, direct messaging on platforms and forms of payment for products bought and sold through and between social media users.

Content analysis

After BTM, all selected study posts were exported into a Microsoft Excel spreadsheet for purposes of content analysis. Human annotation was used for content analysis and to filter out posts unrelated to our study aims. Instagram posts not directly related to the marketing, buying and selling of vaping products were removed from the final dataset and were not content coded. Irrelevant posts included those containing vape-related hashtags but that did not actively offer to sell or buy ENDS products. These posts most frequently included vaping enthusiasts selling apparel, advertising of food, drinks and entertainment offered at vaping stores or lounges and vaping stores updating customers of hours or store operations unrelated to the sale of ENDS products.

Authors developed a codebook for qualitative content analysis and categorisation of Instagram ENDS selling posts. Coders (first, second and third authors) have experience in conducting

Table 2 Results of bivariate logistic regression models predicting age verification, with predictors organised by store physical address and type of product sold

Predictor	n	OR	2.5% CI	97.5% CI	P value*
Indonesia	156	0.32	0.19	0.53	<0.01
Lebanon	6	1.48	0.27	8.11	0.64
UAE	25	44.45	9.16	801.37	<0.01
UK	33	2.53	1.21	5.44	0.01
USA	23	1.14	0.47	2.79	0.77
E-liquid	187	0.23	0.14	0.40	<0.01
Equipment	39	5.38	2.58	12.12	<0.01
Vape	35	2.47	1.21	5.20	0.01

Column n denotes posts in which stores had the specified predictor, out of total sample size n=276 websites. Characteristics attributed to under five websites were excluded.

*Red text denotes statistically significant association.

manual annotation of tobacco and ENDS-related themes using social media data as published in prior studies.^{18 22} Authors NS, MN and CB annotated posts independently and achieved high intercoder reliability (kappa=0.97) (see details in online supplemental file). Coders denoted positive or negative attribution in separate Excel spreadsheets, which were then merged, with disagreements resolved by senior investigators RC and TKM to produce the final dataset. For the development of the codebook, study aims included conducting content analysis to characterise the type of sellers, ENDS products offered for sale, source and origin of purported sale/distribution and conducting analysis of age verification processes for sites that linked to Instagram posts to enable sales transactions (see table 1). For each Instagram post, the text and image data were classified for purposes of content analysis. A first pass of content analysis was used on the text description, and then the image content was subsequently analysed for additional context for thematic analysis. If text did not include marketing or selling characteristics, associated images for that post (eg, ENDS product images, etc) were not deemed relevant.

Bivariate logistic regression models were computed to test the association between age verification and two sets of binary characteristics: country affiliation and type of product marketed (table 2). Specifically, logistic regression using the Wald χ^2 test was used to determine whether the rate of age verification for posts with retailers attributed to a given country or product type was significantly different from the rate of age verification for posts with retailers from other countries or product types. Results from logistic regressions were used to produce odds ratios (ORs) describing the likelihood of age verification for transactions from a post in a given category (ie, country or product type) divided by the likelihood of age verification for transactions from a post in all other categories (ie, all other countries or all other product types). Furthermore, frequencies and proportions of posts leading to transactions with stores in given countries or product types were tabulated and presented as table 3. Posts leading to sales in unknown countries or in unclear product types were included in this tabulation.

RESULTS

We obtained a total corpus of 70 725 Instagram posts. After using BTM to identify specific topic clusters of suspected ENDS selling activity, we reduced the total number of posts to 8821 posts, of which 3331 unique Instagram posts composed of 298 unique sellers were manually annotated and confirmed as associated

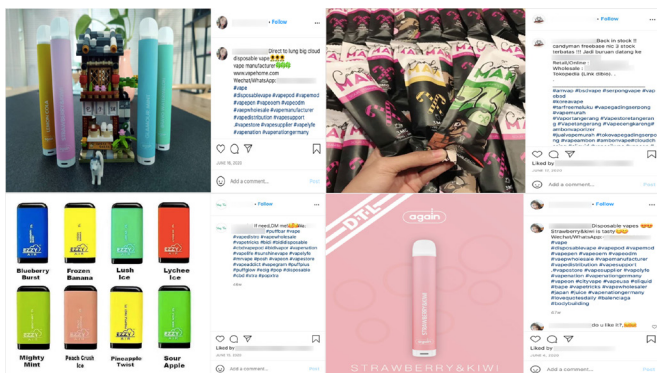
Table 3 Frequency and proportion of posts denoting stores from given countries or selling given types of vaping products

Category	n	%
Canada	3	1.1
China	4	1.4
Indonesia	156	56.5
Italy	1	0.4
Kuwait	3	1.1
Lebanon	6	2.2
Philippines	1	0.4
Poland	5	1.8
Slovakia	1	0.4
UAE	25	9.1
UK	33	12.0
USA	23	8.3
Unknown location	15	5.4
E-liquid	187	67.8
Equipment	39	14.1
Unclear product type	15	5.4
Vape	35	12.7

with ENDS selling activity (see online supplemental file for flow chart). Posts filtered out and deemed unrelated to study aims in topic clusters selected primarily included advertising of ENDS products by retailers that did not involve online sales activity and advertisement and selling of non-ENDS products (eg, vape store conducting promotion on apparel).

Content analysis

The text content of ENDS selling posts exhibited some variation; however, the vast majority of posts were thematically similar, offering brief descriptions of the ENDS product offered, any sales promotions and contact information to purchase followed by a large array of vaping-related hashtags. Generally, posts would come from an account with a username identifying themselves as a business and the image in a post consisting of actual ENDS products being sold. Images of different ENDS products in posts comprised of either physical pictures of the products in stores or stock images posted by purported manufacturers/sellers. Text content of posts primarily consisted of specific ENDS products advertised, pricing information and multiple hashtags relating to vaping behaviour, types of ENDS products, and locations (eg, #vapelyfe, #vapedubai, etc). Examples of deidentified Instagram selling posts detected in this study are included in [figure 1](#).

**Figure 1** Examples of Instagram ENDS selling posts detected.

In relation to characteristics of ENDS sellers, the most common seller category detected were retail accounts (ie, those purporting to operate out of a physical brick-and-mortar location and posting via their Instagram vape store account), which made up 46.9% (n=1561) of all sellers. This was followed by individual seller accounts (ie, those using contact information to sell product but who appear to be individuals and not retail outlets), which accounted for 43.4% (n=1446). A much smaller volume were identified as online sellers (ie, those operating external websites or e-commerce stores engaged in the sale of ENDS who also use Instagram accounts), which accounted for 8.8% (n=292) of all sellers identified.

Retail sellers offered ENDS via different transaction formats, including directly through encrypted messaging platforms or through in-store purchase and always provided an address for a brick-and-mortar store location. Individual sellers most frequently used encrypted messaging platforms as contact information, namely WhatsApp, to conduct sales privately, or through Instagram's direct messaging function. As opposed to retail sellers, individual sellers did not provide specific business addresses and emphasised purchase directly through online communication means. In contrast, online sellers conducted direct-to-consumer sales through external websites that they owned and operated linked in Instagram posts. Between retail sellers, individual sellers and online sellers, images posted were similar, mostly involving specific ENDS products offered.

In relation to ENDS product categories, the majority of products offered by Instagram selling accounts focused on different types of flavoured products. Over half (53.0%, n=1767) of detected posts marketed and sold specific nicotine/vaping flavours, which ranged from menthol to a host of other flavours that traditionally appeal to youth and younger adults. Other flavours included fruits (eg, 'grape currant ice', 'banana', 'strawberries', 'pineapple mango' and 'fuji apple') and dessert flavoured products (eg, 'brown sugar cookie', 'oat milk', 'strawberry cheesecake caramel', 'peanut butter and jelly', 'cereal and milk' and 'melon dew soda milk'). Vaping devices were also detected (20.5%, n=684) and consisted of vape pens, larger box mod vapes, disposable single use vapes and flavoured pods. Finally, only 5.0% (n=166) of posts detected sold vaping related accessories (eg, coils, batteries and different modifications of vaping devices and maintenance equipment for purchase). An additional 21.4% (n=714) of posts reviewed included general ENDS promotion and selling queues, but we were unable to identify the specific ENDS product being offered, either due to non-functioning hyperlinks, no product image included for the post or lack of specific descriptions of offered products. The full categorisation of identified products sold and detected in this study is available in the online supplemental file.

Location information

Based on the self-reported information on location in the seller's Instagram profile, text, metadata or other geographic specific information in the post, we identified sellers in over 20 countries including from regions in North America, Europe and Asia. The specific countries most commonly detected in our analysis were sellers from Indonesia (53.6%, n=1785), the United Arab Emirates (20.2%, n=673), the UK (3.3%, n=111) and the USA (3.2%, n=107). The total list of countries detected is available in [figure 2](#). For example, many online sellers located in Indonesia sold ENDS products via an online Indonesian e-commerce platform called Tokopedia, where links embedded in the Instagram

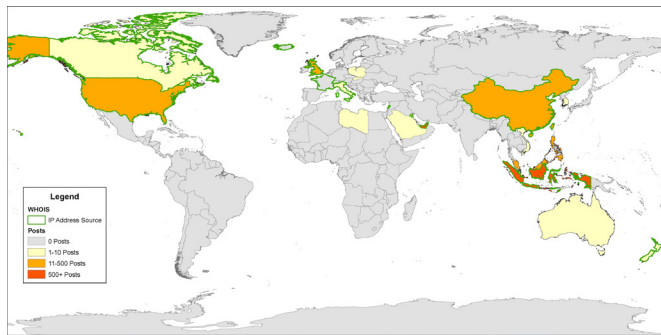


Figure 2 Geographic location and IP address of Instagram ENDS seller posts. ENDS, electronic nicotine delivery systems.

post would lead to specific products listings on the Tokopedia platform.

We also used the ICANN WHOIS database to assess the domain registration addresses of online sellers that included a link to an external website. Of the online sellers reviewed, we identified 56 unique domain names through which online sellers offered ENDS products for sale. We then identified the domain registrant addresses in order to assess if it matched self-reported location information found via information posted on Instagram. Notably, we found 2 (3.6%) websites that were registered in countries outside of those where the Instagram posts reported their location. One post advertised ENDS and provided information to users that they were located in the United Arab Emirates, including a domain name with the name of a city in the UAE ('dubaimrvape'), yet the domain was registered from an address in Auckland, New Zealand. An additional website was being advertised as an online seller based in Indonesia but had the domain registered in the USA.

Age verification requirements

For Instagram posts categorised as online sellers, we conducted separate analysis of age verification processes used on their externally linked websites. Using simulated online purchase methods (ie, visiting a site, putting ENDS product into an e-commerce shopping cart and proceeding to purchase process without providing payment information), we found that the majority of websites had no form of age verification on entry (56.2%, $n=164$). For those that did use a form of age assessment, 38.4% ($n=112$) simply included a 'yes/no' question upon visiting a website (varying from 18 to 21 years depending on country.) At online point of purchase, age verification dropped significantly, with only 1.4% ($n=4$) of sites requesting a form of verification, such as providing appropriate proof of age with valid identification credentials.

We observed that age verification was significantly lower for Indonesian websites (29% vs 56% elsewhere) and higher in websites from the United Arab Emirates (96% vs 35% elsewhere). Overall, Indonesia had the lowest number of posts linked to a site with age verification (29%), with the UK having 61% and the United Arab Emirates having 96% in comparison. Furthermore, age verification was lower for those selling flavoured products (29% vs 64%), higher for those selling vaping-associated products, such as equipment (74% vs 35%), and higher for those selling vapes (60% vs 38%). All differences were statistically significant, at $p<0.05$.

Limitations

Our study has certain limitations. The study was limited to a range of posts collected within a year long period from August 2019 to August 2020 using specific hashtags. This includes the age verification assessment, which was only taken at a point of time, though changes in website age verification processes may occur. Such parameters may exclude a much larger dataset of users posting over a longer time period, where distributions in types of sellers, products and regions could vary. Additionally, this study's focus on hashtags/keywords related to both ENDS product promotion and youth and young adult vaping behaviour could limit the generalisability of study results, particular in the context of more general hashtags/terms related to ENDS product marketing, selling and distribution. We also only examined three BTM clusters that had high frequency of selling argument-related keywords, though selling activity in other BTM clusters may also be present. This study also employed a combination of text and image content analysis to characterise ENDS selling posts, but those Instagram posts that only contained images and no hashtag/text associated with marketing and/or selling activity may have been excluded. Furthermore, the time difference between the data collection, BTM processing and human annotation may have resulted in certain posts being taken down by the platform and third-party websites to be removed. Hence, study results are not generalisable to the activity of all Instagram ENDS-related marketing and selling posts, and other sales activity in this dataset may not have been captured. Future studies should explore the use of supervised machine learning approaches informed by selling and account characteristics and features detected in this study to enable higher volume and more automated detection of ENDS selling activity. Authors also did not purchase any ENDS products from individual, retail or online sellers from vendors on Instagram, so we cannot confirm if ENDS products advertised were actually available or could be shipped to a specific jurisdiction. Specifically, conducting 'secret shopper' test buys of products and shipping them to a US-based address could be illegal and likely require a state waiver or other Federal exception to conduct research, which was beyond the scope of this study.

DISCUSSION

Our results demonstrated that retail and individual accounts were the most commonly detected Instagram ENDS seller types (46.3% and 43.4%, respectively), followed by online sellers (8.8%). Results also indicate that Instagram has global coverage, with over 20 countries claimed as places of business or locations of ENDS sellers. Importantly, the three most prevalent countries (Indonesia, UAE and UK) are outside of the USA, yet users in the USA and other countries that prohibit the sale of certain flavoured vaping products or restrict import of ENDS from other countries (eg, flavoured cartridge-based ENDS products that are not menthol are prohibited based on FDA guidance at the Federal level and other flavoured products may be banned or restricted based on state or local policies), may nevertheless be exposed to and have access to these products. A variety of flavoured products, many of which appeal to youth and young adult populations (fruit and dessert flavoured), were the predominant ENDS product type offered. For online sellers offering the sale of ENDS via external websites linked in Instagram posts, the majority (56.2%) contained no age verification at point of entry, and almost all websites had no age verification at point of purchase (98.4%). Notably, we found that age verification was significantly lower for websites offering the sale of flavoured products.

Additionally, results from bivariate logistic regression models indicate that some countries and products significantly deviate from the overall norms vis-à-vis age verification for specific vaping products (see table 2). Namely, websites for Indonesian retailers were less likely to use age verification when compared with accounts from other countries sampled, indicating potential disregard for online measures designed to protect adolescents from the harms associated with vaping products or possibly reflective of lack of domestic regulation and enforcement of restricting underage sales online. Furthermore, all retailers selling flavoured vaping products online were less likely to have an age verification component when compared with other products sold. These retailers may already disregard or attempt to circumvent prohibitions on underage selling, as the intention of selling many flavoured vaping products may itself be to encourage appeal and uptake among younger consumers.

These data suggest that Instagram acts as a hub for the sale of potentially unregulated or restricted ENDS and flavoured e-liquids, despite attempts to curb ENDS marketing, appeal and access through tobacco product regulation. Instead, there remain numerous avenues for user to purchase potentially unregulated, illegal and unapproved flavoured vaping products from a variety of international sellers with little to no meaningful age verification. This availability may be contributing to the youth vaping epidemic in the USA, where in September 2020, the US Centers for Disease Control and Prevention reported that the use of flavoured disposable vaping devices, such as fruit, mint, menthol and dessert flavour pods, increased 1000% among high school students and 400% among middle school students.²³

In response to these concerning youth flavoured vaping trends, the FDA has sought to ban the sale of flavoured ENDS products without premarket approval and has made it an enforcement priority to seek out those who may be violating this ban.¹⁰ However, comprehensively addressing the unique challenges of online outlets that enable ENDS marketing, sales and access requires more comprehensive surveillance, regulation and enforcement at the local, state, Federal *and* digital platform level. For example, existing tobacco regulatory efforts should leverage new approaches in data mining and machine learning (including both unsupervised and supervised approaches) to ensure comprehensive surveillance of violators, limit illegal importation and strengthen implementation and enforcement of tobacco control laws meant to curb youth access and ENDS product appeal whether originating from home or abroad.

Beyond Instagram, other social media platforms popular among youth and adolescents, such as YouTube and TikTok, should also be assessed for how they may portray and promote provaping marketing content and potential illegal sales activity that could disproportionately impact young users given platform popularity among this demographic.^{24–26} This includes addressing access to ENDS facilitated by social media sites that connect users to domestic and international vendors that do not use appropriate age verification. Importantly, the operation of these vendors runs contrary to the passage of the US Preventing Online Sales of E-Cigarettes to Children Act in December of 2020, which made it explicit that proper age verification is required for ENDS purchasing regardless of whether the seller originates from the USA or internationally.²⁷

However, there remain challenges for effective regulation in a borderless digital environment, as our results indicate that domestic and foreign-based ENDS sales appear to be facilitated through the Instagram platform. For instance, the rising popularity of ENDS in low-income and middle-income countries coupled with inadequate product and marketing regulation in

these international markets can lead to unfettered access on social media platforms populated with large numbers of US users.²⁸ Indonesia, for example, has seen a rapid rise in popularity of ENDS products, and the country reportedly has the second largest global share of vaping-related Instagram posts.²⁹ Our study found a similarly high share of Indonesia-based flavoured ENDS selling posts and linked websites that did not require age verification, the presence of which could impact efforts to regulate and restrict access in the USA and in markets that expressly prohibit access and underage sales.

Importantly, in order to establish jurisdiction authority over these online sellers, US regulators must first distinguish if either international online sellers are deliberately selling to US consumers or are inadvertently transacting with US consumers via their Instagram accounts/websites. Furthermore, individual sellers, who leave limited contact information and often conduct sales on encrypted messaging platforms, may be hard to prosecute due to lack of sufficient evidence, due to the inability to confirm details and completion of ENDS transactions. Faced with these challenges, congressional and public pressure on Instagram to take a more broad-based approach against the creation and use of accounts that expressly use platform features to directly facilitate ENDS product sales and illegal importation should be pursued.

In addition to content moderation, Facebook and Instagram policies and terms and conditions regarding ENDS-related content and product promotion must be clarified and strengthened with the intent of reducing opportunities for users to misinterpret or circumvent existing international or domestic tobacco control laws aimed at reducing the burden of the youth vaping epidemic. Specifically, Facebook and Instagram's ENDS/tobacco policy only explicitly bans paid advertisements and influencers, though allows promotion from legitimate brick-and-mortar entities (including retailers, websites, brands and private individuals representing these entities). This policy creates sufficient loopholes and policy gaps that allow various forms of user-generated content to engage in the promotion and sale of ENDS. Instead,

What this paper adds

What is already known on this subject

- ▶ Exploratory analyses have examined types of vaping and electronic nicotine delivery systems (ENDS) related posts on social media, focusing on industry advertising, image analysis and thematic analysis regarding the messaging and influence of vaping related posts on Instagram. However, no studies to our knowledge have analysed the promotion and sale of ENDS products aimed at youth and young adults on Instagram, nor have they conducted age verification analysis for Instagram linked websites at the point of purchase.

What this paper adds

- ▶ This paper demonstrates the wide geographic scope of Instagram sellers and provides initial quantification of the specific types of products most frequently being sold: flavoured e-liquids and vaping devices. Importantly, our results demonstrate that minimal age verification is used by online sellers to promote and sell specific products implicated in the youth vaping epidemic. Our results demonstrate gaps in tobacco regulatory science and the need for greater oversight at state, federal and the private sector level to address the global unregulated ENDS marketplace on Instagram.

current ENDS/tobacco promotional and sales content restrictions should extend to all non-sponsored and organic user-generated posts that facilitate direct-to-consumer sales, regardless of if they originate from an individual user, physical vaping retailer or international seller. Current Instagram policy guidelines lack this specificity, and possibly as a result, an ENDS global marketplace remains active on the platform, potentially enticing millions of young online users to vape.¹²

Twitter Tim K Mackey @tkmackey

Contributors JL and RC collected data. NS, MN, CB and TKM conducted manual annotation. NS, MC, CB, JL, RC and TKM conducted data analyses. All authors contributed to the design, formulation, drafting, completion and approval of the final manuscript. TKM acquired funding for the study. The guarantor, TKM, accepts full responsibility for the work and/or the conduct of the study, had access to the data, and controlled the decision to publish.

Funding This study was funded by the University of California Tobacco-related Disease Research Program award no. T29IP0384 and T31IP1928.

Disclaimer The opinions expressed are those of the authors alone.

Map disclaimer The inclusion of any map (including the depiction of any boundaries therein), or of any geographic or locational reference, does not imply the expression of any opinion whatsoever on the part of BMJ concerning the legal status of any country, territory, jurisdiction or area or of its authorities. Any such expression remains solely that of the relevant source and is not endorsed by BMJ. Maps are provided without any warranty of any kind, either express or implied.

Competing interests MN, CB, JL, and TKM are employees of the startup company S-3 Research LLC. S-3 Research is a startup funded and currently supported by the National Institutes of Health – National Institute of Drug Abuse through a Small Business Innovation and Research contract for opioid-related social media research and technology commercialisation. Author reports no other conflict of interest associated with this manuscript.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Data are available on reasonable request to authors subject to appropriate deidentification.

ORCID iDs

Raphael Cuomo <http://orcid.org/0000-0002-8179-0619>

Tim K Mackey <http://orcid.org/0000-0002-2191-7833>

REFERENCES

- O'Brien EK, Hoffman L, Navarro MA, et al. Social media use by leading US e-cigarette, cigarette, smokeless tobacco, cigar and hookah brands. *Tob Control* 2020;29:tobaccocontrol-2019-055406–97.
- Pew Research Center. Social media use in 2021. Available: <https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/> [Accessed 17 Jul 2021].
- Ketonen V, Malik A. Characterizing vaping posts on Instagram by using unsupervised machine learning. *Int J Med Inform* 2020;141:104223.
- Pokhrel P, Fagan P, Herzog TA, et al. Social media e-cigarette exposure and e-cigarette expectancies and use among young adults. *Addict Behav* 2018;78:51–8.
- Kong G, LaVallee H, Rams A, et al. Promotion of Vape tricks on YouTube: content analysis. *J Med Internet Res* 2019;21:e12709.
- Klein EG, Czaplicki L, Berman M, et al. Visual Attention to the Use of #ad versus #sponsored on e-Cigarette Influencer Posts on Social Media: A Randomized Experiment. *J Health Commun* 2020;25:925–30.
- Czaplicki L, Tulsiani S, Kostygina G, et al. #toolittleoolate: JUUL-related content on Instagram before and after self-regulatory action. *PLoS One* 2020;15:e0233419.
- McCausland K, Maycock B, Leaver T, et al. The messages presented in electronic Cigarette-Related social media promotions and discussion: Scoping review. *J Med Internet Res* 2019;21:e11953.
- Mackey TK, Miner A, Cuomo RE. Exploring the e-cigarette e-commerce marketplace: identifying Internet e-cigarette marketing characteristics and regulatory gaps. *Drug Alcohol Depend* 2015;156:97–103.
- Schiff SJ, Kechter A, Simpson KA. Accessing Vaping products when Underage: a qualitative study of young adults in southern California. *N&TR* 2021;23:836–41.
- Chadi N, Hadland SE, Harris SK. Understanding the implications of the "vaping epidemic" among adolescents and young adults: A call for action. *Subst Abuse* 2019;40:7–10.
- Fda finalizes enforcement policy on unauthorized flavored cartridge-based e-cigarettes that appeal to children, including fruit and MINT | FDA. Available: <https://www.fda.gov/news-events/press-announcements/fda-finalizes-enforcement-policy-unauthorized-flavored-cartridge-based-e-cigarettes-appeal-children> [Accessed 18 Jul 2021].
- Public Health Law Center. States and tribes stepping in to protect communities from the dangers of e-cigarettes: actions and options, 2020]. Available: <https://www.publ.ichealthlawcenter.org/resources/states-and-tribes-stepping-protect-communities-dangers-e-cigarettes-actions-and-options> [Accessed 17 Jul 2021].
- Instagram Help Center, Facebook. Community guidelines, 2021. Available: <https://www.facebook.com/help/instagram/477434105621119> [Accessed 18 Jul 2021].
- Facebook AP, 2021. Available: <https://business.facebook.com/policies/ads/> [Accessed 18 Jul 2021].
- Vassey J, Metayer C, Kennedy CJ. #Vape: Measuring E-Cigarette Influence on Instagram With Deep Learning and Text Analysis. *Front Commun* 2020:75.
- Laestadius LI, Wahl MM, Cho YI. #Vapelife: An Exploratory Study of Electronic Cigarette Use and Promotion on Instagram. *Subst Use Misuse* 2016;51:1669–73.
- Gao Y, Xie Z, Sun L, et al. Electronic Cigarette-Related contents on Instagram: observational study and exploratory analysis. *JMIR Public Health Surveill* 2020;6:e21963.
- Chu K-H, Allem J-P, Cruz TB, et al. Vaping on Instagram: cloud chasing, hand checks and product placement. *Tob Control* 2017;26:575–8.
- Cuomo RE, Purushothaman VL, Li J, et al. Characterizing self-reported tobacco, Vaping, and Marijuana-Related Tweets Geolocated for California College Campuses. *Front Public Health* 2021;9:343.
- Yan X, Guo J, Lan Y. A bitern topic model for short texts. WWW 2013. *Proceedings of the 22nd International Conference on World Wide Web* 2013:1445–55.
- Cai M, Shah N, Li J, et al. Identification and characterization of tweets related to the 2015 Indiana HIV outbreak: a retrospective infoveillance study. *PLoS One* 2020;15:e0235150.
- Li J, Xu Q, Shah N, et al. A machine learning approach for the detection and characterization of illicit drug Dealers on Instagram: model evaluation study. *J Med Internet Res* 2019;21:e13803.
- Sun T, CCW L, Chung J. Vaping on TikTok: a systematic thematic analysis. *Tob Control*.
- ASL T, Weinreich E. #PuffBar: how do top videos on TikTok portray Puff Bars? *Tob Control* 2020:tobaccocontrol-2020-055970.
- Sears CG, Walker KL, Hart JL, et al. Clean, cheap, convenient: promotion of electronic cigarettes on YouTube. *Tob Prev Cessat* 2017;3. doi:10.18332/tpc/69393. [Epub ahead of print: 07 04 2017].
- Feinstein DS. 1253 - 116th Congress (2019-2020): Preventing Online Sales of E-Cigarettes to Children Act, 2020. Available: <https://www.congress.gov/bills/116th-congress/senate-bill/1253> [Accessed 17 Jul 2021].
- Newman J, Nurfaiza MW. Policy design, non-design, and anti-design: the regulation of e-cigarettes in Indonesia. *Policy Stud* 2020;2:1–18.
- Orlan EN, Parascandola M, Grana R. JUUL from the USA to Indonesia: implications for expansion to LMICs. *Tob Control* 2020;29:e155–6.