Estimating E-Cigarette Use Prevalence among US Adolescents Using Vaping-Related Online Search Trends

Floe Foxon
Pinney Associates, Inc., Pittsburgh, Pennsylvania, USA

ABSTRACT

Introduction: Adolescent e-cigarette use is a developing phenomenon. Greater surveillance of under age use is necessary to inform e-cigarette policy and mitigate adolescent e-cigarette use. Accurate prevalence estimates for adolescent e-cigarette use are provided by large national surveys. However, these surveys are costly and provide only annual estimates. To obtain more affordable estimates faster and more frequently, novel methods are required. Methods: Online search term popularity data were taken from Google Trends. Interest in vaping-related search terms were followed monthly from January 2011 to November 2020. Time-lagged zero-normalized cross-correlations were performed between the Google data and current (past 30 day) high-school e-cigarette use prevalence estimates from the National Youth Tobacco Survey (NYTS). The search interest data were then calibrated to the NYTS data to estimate adolescent e-cigarette use prevalence using online searches. Results: Maximum correlation coefficients of 0.979 for “vapes” and 0.938 for “vape” were obtained when search interest lagged use prevalence by one month, and 0.970 for “vape pen” when the lag was two months (p < 0.001 for all). Calibrating the search term data to NYTS provided a high-school current e-cigarette use prevalence estimate of 12.1–18.4% for November 2020, suggesting adolescent use of e-cigarettes has continued to decline since the NYTS estimate of 19.6% for January–March 2020. Conclusions: Online search trend data may provide reasonably reliable and more frequent estimates of adolescent e-cigarette use prevalence at substantially lower costs than traditional surveys. Such additional data may help to assess immediate impacts of policies and events.

Introduction

E-cigarettes have emerged as an alternative nicotine source and possible future smoking-cessation device for adult smokers (American College of Cardiology, 2020; Hajek et al., 2019), but adolescent use of e-cigarettes (Gentzke et al., 2019) presents a public health challenge. Tobacco control policies are shaped by available data on prevalence of product use. Existing data sources for adolescent e-cigarette use prevalence include Monitoring the Future (University of Michigan, 2021), the Youth Risk Behavior Survey (YRBS) (US Centers for Disease Control and Prevention, 2020) and the National Youth Tobacco Survey (NYTS) (US Centers for Disease Control and Prevention, 2019a), with YRBS and NYTS administered by the US Centers for Disease Control and Prevention, as well as various state-wide surveys such as the state Youth Tobacco Surveys (YTS) (US Centers for Disease Control and Prevention, 2019b) and Healthy Kids surveys (California Department of Education, 2020; Colorado Department of Public Health & Environment, 2019).

These surveys utilize large sample sizes and complex sample designs which provide the best existing estimates of adolescent e-cigarette use prevalence. However, administering state- and nation-wide surveys is a costly endeavor in terms of the time, money, and labor involved. Consequently, these surveys are administered at most annually or biannually. Because adolescent e-cigarette use is evolving, data with higher temporal resolution (i.e. less time between point estimates) and which require fewer resources to collect are necessary; waiting an entire year to determine the impact of a policy may risk irreversible consequences if that policy fails or backfires.

Adolescent e-cigarette use intersects with the online world. Half of US adolescents may be exposed to tobacco or e-cigarette related social media (Hebert et al., 2017), and about 1 in 10 adolescent e-cigarette users source the e-cigarettes they use from the Internet (Merianos et al., 2019). E-cigarette related discussion forums have also been identified as a rich source for information on adolescent e-cigarette use behavior (Zhan et al., 2019). Given then that adolescent e-cigarette use is associated with online content, this raises the question of whether vaping-related internet use predicts real-world e-cigarette use.

Google Trends (Google, 2020), an online, search-term-popularity data generator, has been demonstrated to have modest reliability in reflecting geographical and temporal patterns in epidemiological settings (Cervellin et al., 2017). Extant
research has applied temporal Google Trends data to predict outbreaks of influenza (Carneiro & Mylonakis, 2009) and COVID-19 (Ahmad et al., 2020), the latter with correlation as high as 0.998. In nicotine and tobacco science, relative Google search volumes have been compared for heat-not-burn products and e-cigarettes to investigate heat-not-burn market growth (Caputi et al., 2017). Most relevant to the present study, statistically significant but weak-to-moderate strength geographical correlation was demonstrated between Google Trend data and state use prevalence for cigars in 2011 (Cavazos-Rehg et al., 2015).

This study aims to determine whether Google Trends data correlate temporally with adolescent e-cigarette use prevalence in the US using NYTS prevalence estimates from 2011–2020. This presents novel research which may provide faster and higher temporal resolution estimates of adolescent e-cigarette use prevalence.

Materials & methods

Sample

National estimates for the prevalence of past 30-day (current) e-cigarette use among US high-schoolers (here “adolescents”) from 2011–2020 were taken from resources published by the US Department of Health and Human Services (US Food and Drug Administration, 2019b; Wang et al., 2020). These prevalence estimates are derived from ten waves of NYTS, which is an annual, nationally representative survey of US adolescents with a multi-stage probabilistic cluster sample design. Approximately 17,000–25,000 adolescents were surveyed at each wave. The exact dates through which these surveys were administered were taken from their annual methodology reports (US Centers for Disease Control and Prevention, 2019a).

Online search popularity data were taken from Google Trends. Google Trends provides normalized “search interest” values which represent the popularity of a given search term through Google web searches on a scale from zero to 100. A value of 100 represents the peak popularity for that term, which falls on some month. A value of 50 on another month means the search term was half as popular on that month than it was in the month when the value was 100. Monthly data for the vaping-related search terms “e-cigarettes”, “e-cigs”, “mods”, “e-hookahs”, “vapes”, “vape”, and “vape pen” were collected from January 2011 to November 2020. The above search terms were selected because they are listed as alternative names for electronic cigarettes in the 2019 NYTS questionnaire definition (which precedes the e-cigarette-related questions in this survey) (US Centers for Disease Control and Prevention, 2019a).

The range of dates was selected to coincide with NYTS data collection, and the data were restricted geographically to ensure only searches from within the United States were included for analysis.

Analyses

The NYTS prevalence estimates were normalized, and for visualization purposes the Google web search interest data were re-normalized to the same scale by dividing each popularity score in each set of search term data by the maximum popularity score for that search term such that for all search terms, a value of one represented the maximum popularity score for that search term. These data were then plotted to observe the similarities and differences in their trends over time.

The midpoint dates between the beginning and end of survey administration for each NYTS wave were found, and the months these dates fell on were taken to be the months which correspond to the prevalence estimate for that year. The NYTS data were then paired with the Google web search data for the same months (zero lag). Treating the data as signals over time, zero-normalized cross-correlations were performed between NYTS and each of the Google web searches with these month-pairs, following the methods of Ahmad et al. (2020). Conceptually, it is plausible that a change in search popularity may precede a corresponding change in current use prevalence and vice versa. For example, an adolescent who has not used an e-cigarette in the past 30 days may first search for online vaping-related content out of curiosity, become influenced by that content over time, and then initiate e-cigarette use at some later date when an opportunity arises or when their interest reaches a threshold to do so. Conversely, an adolescent who is already a current user may after some time using e-cigarettes seek information on how to quit. In both cases there will occur some time between online searching and meeting the definition of a current user. Since there is no published literature on the duration of this lag, cross-correlation as described above were repeated with exploratory time lags of months up to three months in either direction, for example a one-month lag was implemented by cross-correlating the NYTS data with Google Trends data which correspond to the NYTS months minus one month, and a ‘-1’ month lag was implemented by cross-correlating the NYTS data with Google Trends data which correspond to the NYTS months plus one month. The use of lags accounts for the time taken for online interest to develop into real-world use and vice versa. Corresponding p-values were calculated for the cross-correlation coefficients with statistical significance at alpha = 0.05.

Finally, the peaks of the search term popularity trends were calibrated to the peak NYTS e-cigarette use prevalence (27.5%) such that e-cigarette use prevalence estimates could be obtained from the search term data for any given month. In this way, prevalence estimates were obtained for November 2020.

All analyses were conducted in Python version 3.7.6 with the packages NumPy version 1.18.5, Matplotlib version 3.2.2, and Pandas version 1.0.5. Code will be made available upon reasonable request.

Results

Figure 1 shows the normalized trends in high school e-cigarette use prevalence and Google web search popularity for the search terms “vapes”, “vape”, and “vape pen” over time. It can be readily seen than all three trends display very similar shapes.
Table 1 shows the zero-normalized cross-correlations between the NYTS prevalence estimates and corresponding Google web search interest for a range of time lags. The maximum cross-correlation coefficients were 0.979 for “vapes” and 0.938 for “vape” with search interest lagging NYTS prevalence by one month, and 0.970 for “vape pen” with search interest lagging NYTS prevalence by two months. Across the entire range of lags, all cross-correlations for all terms were statistically significant. In contrast, correlations were not significant for the search terms “e-cigarettes”, “mods”, and “e-hookahs”. This may suggest that the prior terms are more commonly used by adolescents to describe vaping than the latter terms. Correlations were strongest with 1–2 month lags, suggesting changes in vaping-related search trends are shortly followed by changes in adolescent e-cigarette use trends, however whether this relationship is causal is not determined.

Calibrating the search term popularities to NYTS prevalence data for January–March 2020 gave current e-cigarette use prevalence estimates of 19.3% (“vape” search term with one-month lag), 17.3% (“vapes” search term with one-month lag), and 12.7% (“vape pen” search term with two-month lag). Since the actual prevalence estimate for January–March 2020 from NYTS was 19.6%, this suggests good agreement between NYTS and the “vape” search term with one-month lag model, and less agreement with NYTS for the other two models.

Using these models, prevalence estimates for November 2020 of 18.4% (“vape” search term with one-month lag), 16.8% (“vapes” search term with one-month lag), and 12.1% (“vape pen” search term with two-month lag) were obtained.

### Discussion

This research aimed to determine whether online vaping-related search data from Google Trends correlate with adolescent e-cigarette use prevalence in the US. The results presented here show very high and statistically significant correlation between NYTS prevalence estimates and interest in the search terms “vapes”, “vape”, and “vape pen”, but no statistically significant correlations were observed for the search terms “e-cigarettes”, “e-cigs”, “mods”, and “e-hookahs”. This may suggest that the prior terms are more commonly used by adolescents to describe vaping than the latter terms. Correlations were strongest with 1–2 month lags, suggesting changes in vaping-related search trends are shortly followed by changes in adolescent e-cigarette use trends, however whether this relationship is causal is not determined.

Using the search trend data to predict high school current e-cigarette use prevalence provided a prevalence estimate for November 2020 of 12.1–18.4%, which is lower than the 19.6%
Acknowledgements

The author would like to acknowledge Dr Arielle Selya for proof-reading this article, and Joe Gitchell for providing insightful comments and suggestions.

Declaration of interest

In 2020, Floe Foxon became a consultant to PinneyAssociates, Inc. PinneyAssociates provides consulting services on tobacco harm reduction on an exclusive basis to Juul Labs, Inc. In recent years, PinneyAssociates has consulted for British American Tobacco and Reynolds American Inc and subsidiaries on tobacco harm reduction. Juul Labs, Inc. did not sponsor this article or participate in the design, study execution, data analysis, writing, or publication.

Funding

This work was not supported by any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ORCID

Floe Foxon https://orcid.org/0000-0002-4893-9178

References

Ahmad, I., Flanagan, R., & Staller, K. (2020). Increased internet search interest for GI symptoms may predict COVID-19 cases in U.S. hotspots. *Clinical Gastroenterology and Hepatology*, 18(12), P2833–2834. https://doi.org/10.1016/j.cgh.2020.06.058

American College of Cardiology. (2020). E3 trial: E-Cigarettes plus counseling more effective than counseling alone in smoking cessation. https://www.acc.org/latest-in-cardiology/articles/2020/03/24/16/41/mon-1045-e3-efficacy-safety-e-cigarettes-acc-2020

California Department of Education. (2020). California *Healthy Kids* Survey, Retrieved October 4 from https://www.cde.ca.gov/ls/be/at/chks.asp

Caputi, T. L., Leas, E., Dredze, M., Cohen, J. E., & Ayers, J. W. (2017). They’re heating up: Internet search query trends reveal significant public interest in heat-not-burn tobacco products. *PloS One*, 12(10), e0185735. https://doi.org/10.1371/journal.pone.0185735

Carneiro, H. A., & Mylonakis, E. (2009). Google trends: A web-based tool for real-time surveillance of disease outbreaks. *Clinical Infectious Diseases*, 49(10), 1557–1564. https://doi.org/10.1086/630200

Cavazos-Rehg, P. A., Krauss, M. J., Spitznagel, E. L., Lowery, A., Grucza, R. A., Chaloupka, F. J., & Bierut, L. J. (2015). Monitoring of non-cigarette tobacco use using Google Trends. *Tobacco Control*, 24(3), 249–255. https://doi.org/10.1136/tobaccocontrol-2013-051276

Cervellin, G., Comelli, I., & Lippi, G. (2017). Is Google trends a reliable tool for digital epidemiology? Insights from different clinical settings. *The Journal of Epidemiology and Global Health*, 7(3), 185–189. https://doi.org/10.1016/j.jegh.2017.06.001

Colorado Department of Public Health & Environment. (2019). *Healthy kids Colorado* survey. Retrieved October 4 from https://www.colorado.gov/cdphe/hkcs

Dai, H., & Leventhal, A. M. (2019). Prevalence of e-Cigarette use among adults in the United States, 2014-2018. *JAMA*, 322(18), 1824. https://doi.org/10.1001/jama.2019.15331

Gentzk, A. S., Creamer, M., Cullen, K. A., Ambrose, B. K., Willis, G., Jamal, A., & King, B. A. (2019). Vital signs: Tobacco product use among middle and high school students - United States, 2011-2018. *MMWR Morbidity and Mortality Weekly Report*, 68(6), 157–164. https://doi.org/10.15585/mmwr.mm6806e1

Google. (2020). *Google Trends*. Retrieved October 3 from https://trends.google.com/trends?geo=US

Hajek, P., Phillips-Waller, A., Przulj, D., Pesola, F., Myers Smith, K., Bisal, N., Li, J., Parrott, S., Sasieni, P., Dawkins, L., Ross, L., Goniewicz, M., Wu, Q., & McBride, H. J. (2019). A randomized trial of E-cigarettes versus nicotine-replacement therapy. *New
