Published in final edited form as:

Tob Control. 2015 May ; 24(3): 249–255. doi:10.1136/tobaccocontrol-2013-051276.

Monitoring of Non-cigarette Tobacco Use using Google Trends

Patricia A. Cavazos-Rehg, Ph.D.^a, Melissa J. Krauss, M.P.H.^a, Edward L. Spitznagel, Ph.D.^b, Ashley Lowery^a, Richard A. Grucza, Ph.D.^a, Frank J. Chaloupka, Ph.D.^c, and Laura Jean Bierut, M.D.^a

^aDepartment of Psychiatry, Washington University School of Medicine, St. Louis, MO 63110

^bDepartment of Mathematics, Washington University in St. Louis, St. Louis, MO 63130

^cDepartment of Economics, University of Illinois at Chicago, Chicago, IL 60607

Abstract

Background—Google Trends is an innovative monitoring system with unique potential to monitor and predict important phenomena that may be occurring at a population-level. We sought to validate whether Google Trends can additionally detect regional trends in youth and adult tobacco use.

Methods—We compared 2011 Google Trends relative search volume data for cigars, cigarillos, little cigars and smokeless tobacco with state prevalence of youth (grades 9–12) and adult (age 18 and older) use of these products using data from the 2011 United States state-level Youth Risk Behaviors Surveillance System and the 2010–2011 United States National Survey on Drug Use and Health (NSDUH), respectively. We used the Pearson correlation coefficient to measure the associations.

Results—We found significant positive correlations between state Google Trends cigar relative search volume and prevalence of cigar use among youth ($r=0.39$, $R\text{-square}=0.154$, $p=.018$) and adults ($r=0.49$, $R\text{-square}=0.243$, $p<.001$). Similarly, we found that the correlations between state Google trends smokeless tobacco relative search volume and prevalence of smokeless tobacco use among youth and adults were both positive and significant ($r=0.46$, $R\text{-square}=0.209$, $p=.003$ and $r=0.48$, $R\text{-square}=0.226$, $p<.001$, respectively).

Conclusion—The results of this study validate that Google Trends has the potential to be a valuable monitoring tool for tobacco use. The near real-time monitoring features of Google Trends

Please address correspondence and reprint requests to: Patricia A. Cavazos-Rehg, Ph.D., Campus Box 8134, Department of Psychiatry, 660 South Euclid, St. Louis, MO 63110, Phone: (314) 362-2152, FAX: (314) 362-4247, rehgp@psychiatry.wustl.edu.

Conflicts of Interest

Dr. Bierut is listed as an inventor on Issued U.S. Patent 8,080,371, "Markers for Addiction" covering the use of certain SNPs in determining the diagnosis, prognosis, and treatment of addiction.

Contribution Statement

Patricia Cavazos-Rehg originated and supervised the study and led the writing of the article. Melissa Krauss and Ashley Lowery assisted with data abstraction and management, completed the analyses, and participated in the writing. Edward Spitznagel participated in the writing and assisted with the study and the analyses. Frank Chaloupka assisted with drafting the article, interpreting results, and revising it for important intellectual content. Richard Grucza assisted with interpreting results, supervised the study, and reviewed and commented on the article. Laura Bierut supervised the study, assisted with interpreting results, and participated in the writing. All authors were involved in approving the final, submitted version of the article.

may complement traditional surveillance methods and lead to faster and more convenient monitoring of emerging trends in tobacco use.

What this paper adds—Efforts to promptly track tobacco use at a population-level are an important endeavor that could assist key stakeholders with making informed decisions about prevention plans and policies. Our findings establish the validity of Google Trends as an indicator of cigar and smokeless tobacco use by demonstrating that states with higher relative volumes of Google searches for cigar and smokeless tobacco products have higher prevalence of tobacco use of these products. The near real-time monitoring features of Google Trends could be used to supplement traditional survey methods to obtain deeper insight about youth and adult tobacco use.

Keywords

Smoking/epidemiology; Smoking/prevention & control; Tobacco Industry/economics

INTRODUCTION

Using a web search engine is currently one of the most popular internet activities and Google is, by far, the most used search engine in the world.[1] Google Trends was launched by Google, Inc. in 2006 and is a keyword research tool that provides near real-time search query trend data based on Google searches (i.e., publicly available data are updated daily). In recent years, both health epidemiologists and econometricians have begun to recognize the value of Google Trends data in revealing important information about present behavior and helping to predict future behavior. For instance, studies on search query monitoring of behaviors using data from Google Trends found it to be useful for detecting regional outbreaks of influenza,[2, 3] gastrointestinal[4] and Lyme[5] diseases, as well as suicide.[6, 7] More recently, a study used data from Google Trends to show the rising online popularity of e-cigarettes.[8] In addition, econometricians have shown that Google Trends data can forecast unemployment rates.[9] Taken together, the existing studies suggest that Google Trends is an innovative monitoring system with unique potential to predict important phenomena that may be occurring at a population-level.

In the present study, we examine whether Google Trends can also detect regional trends in use of non-cigarette tobacco products. Cigars, including cigarillos and little cigars, and smokeless tobacco are two non-cigarette tobacco products that are widely consumed, second to cigarettes in terms of tobacco use prevalence.[10] In 2012, approximately 13.4 million people in the United States or 5.2% of persons aged 12 years old were current cigar smokers while 9.0 million (3.5 % of persons aged 12 years old) were current smokeless tobacco product users.[10] When examining tobacco use trends over time, there has been less progress made in reducing the use of these non-cigarette tobacco products relative to cigarette use, especially among underage youth (<18 years of age). For instance, in 2001, ~30% of high-school youth were current cigarette smokers versus 18% in 2011.[11] In contrast, the prevalence of youth using non-cigarette tobacco products has remained static for nearly a decade with ~10–15% of high-school youth reporting current use of various cigar products or smokeless tobacco. These non-cigarette tobacco products can contribute to nicotine dependence and overall tobacco-related morbidity and mortality[12] and are worthy of investigation. As a result, the primary aim of this study is to establish the validity of

Google Trends as an indicator of cigar and smokeless tobacco use by evaluating whether states with higher relative volumes of Google searches for these products will have higher prevalence of tobacco use of these products.

METHODS

Data source and respondents

Youth tobacco use prevalence—We used youth tobacco use prevalence estimates from the state-specific United States Youth Risk Behavior Survey (YRBS), which is a biennial school-based, state-administered survey established by the Centers for Disease Control and Prevention to measure health-risk. The YRBS survey employs a two-stage, cluster sample design to produce representative samples of students in grades 9–12 in the participating state. [13] To protect students' privacy, participants volunteer to participate and completion of survey is completely anonymous. Only states with weighted data contributed to the analysis and for states to be considered weighted, their data had to have a representative sample of students, appropriate documentation, and an overall response rate of 60% or higher. In 2011, 43 states had weighted data and in these states the student sample sizes ranged from 1,147 to 13,201 (median: 2,170). We selected the YRBS to examine youth tobacco prevalence because, across most states, YRBS sample sizes were 2–3 times larger than National Survey on Drug Use and Health (NSDUH, described below) data for youth. Additional strengths of the YRBS are that it is the most comprehensive ongoing study of health-risk behaviors among young persons in the U.S. and participants are queried during adolescence which is the stage of development when exposure to the variables of interest generally occurs. Additional information about the YRBS is available at <http://www.cdc.gov/healthyyouth/yrbs/>.

Prevalence of youth current cigar use and current smokeless tobacco use were downloaded from the Youth Online system at <http://apps.nccd.cdc.gov/youthonline/App/Default.aspx>. Current use of cigar products was assessed by the YRBS question “During the past 30 days, on how many days did you smoke cigars, cigarillos, or little cigars?” Current use of smokeless tobacco was assessed by the YRBS question “During the past 30 days, on how many days did you use chewing tobacco, snuff, or dip, such as Redman, Levi Garrett, Beechnut, Skoal, Skoal Bandits, or Copenhagen?” When participants said they had used the product for one or more days, then we considered them to be current users. Cigar use data were available for 36 states with weighted data, and smokeless tobacco use data were available for 40 states with weighted data.

Adult tobacco use prevalence—Adult tobacco use prevalence by state was obtained from the 2010–2011 United States National Survey on Drug Use and Health (NSDUH). The NSDUH is an annual nationwide cross-sectional survey that provides population estimates of substance use and health status of the noninstitutionalized U.S. civilian population aged 12 or older. Independent, multistage area probability samples are collected for each of the 50 states and the District of Columbia. To ensure confidentiality of participants' responses and to increase the validity of respondents' reports of drug use behaviors, data are collected via a combination of computer-assisted personal interviewing for demographic items and audio

computer-assisted self-interviewing to collect sensitive information; i.e., responses to substance use modules were not disclosed to the interviewer. Informed consent was obtained from all participants. Young adults (ages 18–25) were oversampled in the NSDUH. NSDUH employed a 50-State design with an independent, multistage area probability sample for each state. In the 2010–2011 data, all 50 states and the District of Columbia contributed data with sample sizes ranging from 7,684 to 1,773.

A number of key strengths prompted our use of the NSDUH to measure adult tobacco use prevalence including its query on both non-cigarette tobacco products of interest and its large, representative, and randomly selected sample and high response rate.[14] Prevalence of adult current cigar use and current smokeless tobacco use were downloaded from the NSDUH using the Restricted-use Data Analysis System (R-DAS). The R-DAS is an online analysis system that allows researchers to produce tables using restricted-use NSDUH data files (geographic identifiers are not included in publicly available NSDUH files) <http://www.icpsr.umich.edu/icpsrweb/SAMHDA/studies/34482/datasets/1/sda/table>. Current use of cigar products was assessed by the NSDUH question “By cigars we mean any kind, including big cigars, cigarillos, and even little cigars that look like cigarettes. During the past 30 days, that is since (interviewer fills in date), on how many days did you smoke part or all of a cigar?” The responses for two items (chewing tobacco and snuff) were combined to assess current use of smokeless tobacco. First, chewing tobacco was assessed by the NSDUH question “During the past 30 days, that is since (interviewer fills in date), on how many days did you use chewing tobacco?” Second, snuff was assessed by the NSDUH question, “during the past 30 days, that is since (interviewer fills in date), on how many days did you use snuff?” When participants said they had used the product for one or more days, then we considered them to be current users. Data were available for all states plus the District of Columbia.

Google Trends data—Relative search volume for tobacco products by state were represented by 2011 Google Trends data which is an online log that indicates how often a particular search-term is entered in the Google search engine relative to the total search-volume across various regions/time. This analysis indicates the likelihood of a random user to search for a particular term from a certain location at a certain time. The data are displayed on a scale of 0 to 100. The data are first normalized in accordance with the total search volume in that region in the given time period. If the data were not normalized, regions with the most search volume would always have the highest ranking. After normalization, each point on the graph is divided by the highest value and multiplied by 100. Search queries which have extremely low values in comparison to the normalized value of 100 are assigned a 0 volume.[15]

Using Google Trends data, we examined the relative search volumes for cigar and smokeless tobacco products. Google Trends compiles data for any of several search terms, up to 30 words. We compiled lists of search terms that reflected the most commonly used search terms for the product of interest. First, a list of potential search terms was composed including general terms for the tobacco products (e.g., cigars, cigarillos, little cigars) as well as popular name brands of the products (e.g., Black and Mild, Swisher Sweets), including those brands with highest market shares.[16] We included singular and plural forms of the

items and alternate phrases. To continue building the list, we added popular “related terms” as indicated by our searches in Google Trends. When the list exceeded 30 words (the Google Trends limit), we compared relative search volume for the individual terms and removed those with the lowest relative search volume. Words or phrases that might describe the tobacco products of interest but would also generate irrelevant (non-tobacco) results were excluded. For example, “dip” is commonly used to describe smokeless tobacco but it was excluded from analysis because of the high volume of unrelated tobacco searches that are also generated by this term. Results were limited to the United States and to the year 2011 because it coincided with the most recent YRBS and NSDUH available data. For cigars, relative search volume for any of the following terms were used: *cigars, cigar, little cigars, cigarillo, cigarillos, flavored cigars, Black and Mild, Black and Milds, Phillies cigars, Swisher Sweet, Swisher Sweets, sweet cigars, JR cigars, Dutch Master cigars, White Owl cigars*. Search terms/phrases to examine smokeless tobacco products included: *chewing tobacco, smokeless tobacco, chew tobacco, Skoal, Copenhagen tobacco, US tobacco, Copenhagen snuff, snuff tobacco, snus, Camel snus, Skoal tobacco, Skoal dip, dip tobacco, tobacco snuff, Grizzly tobacco, Marlboro snus*.

Statistical Analysis

To correlate the state Google Trends relative search volume for cigars and smokeless tobacco in 2011 with the 2011 state prevalence of youth and adult cigar and smokeless tobacco use, we used the Pearson correlation coefficient, as all variables were normally distributed (all Kolmogorov-Smirnov tests $p > .05$). R-square values are also reported to provide the proportion of variance in state estimates of tobacco product use that can be explained by state Google Trends relative search volume. $P < .05$ was considered statistically significant. We constructed scatterplots and choropleth maps to visualize the associations and regional trends in tobacco use and Google Trends relative search volume. SAS version 9.3 for Windows (SAS Institute, Inc., Cary, NC) was used for these analyses.

We additionally examined the correlations weighted by 2011 state population estimates for youth (ages 14–17 years) or adults (age 18 years) as appropriate (population estimates were obtained from the U.S. Census). We used Stata version 10.0 (StataCorp, College Station, TX) for weighted regression analyses, and used the “robust” option to estimate standard errors using the Huber-White sandwich estimators. Unweighted estimates give equal importance to each state in analysis, regardless of population size, which is important for interpretation given tobacco control efforts are primarily organized at the state level. The weighted estimates provide a national perspective, as states with larger populations are given more weight and states with small populations are given less weight.

RESULTS

Cigars, including cigarillos and little cigars

Figure 1 presents prevalence of youth (Figure 1a) and adult (Figure 1b) cigar use across states as well as cigar relative search volume in Google for 2011 (Figure 1c). For youth cigar use, states with the highest rates were South Carolina (18.3%), Georgia (17.8%), Kentucky (17.5%), and Louisiana (17%) and states with the lowest rates were Utah (5%), Hawaii

(6.8%), Nebraska (9.6%), and Alaska (10.3%). For adult cigar use, states with the highest rates were West Virginia (7.3%), Connecticut (7.2%), District of Columbia (6.8%), and Missouri (6.7%) and states with the lowest rates were Utah (2.4%), North Dakota (3.5%), California and South Dakota (both 3.7%). States with the highest relative volume of cigar searches (Figure 1c) included Florida (100), Rhode Island (98), Pennsylvania (91), and New York and District of Columbia (both 90) and states with the lowest relative volume of cigar searches were Utah (34), Idaho (41), Iowa (49), and North Dakota (51).

Scatterplots of state Google cigar relative search volume by youth and adult prevalence of current cigar use are presented in Figure 2. For youth cigar use (Figure 2a), we found a moderate positive correlation ($r=0.39$, $R\text{-square}=0.154$, $p=.018$). Similarly, for adult cigar use (Figure 2b), we also found a moderate positive correlation ($r=0.49$, $R\text{-square}=0.243$, $p<.001$). When removing influential states for youth, we observed that the correlation coefficient was a bit lower and did not reach statistical significance (excluding Utah and Rhode Island $r=0.26$, $R\text{-square}=0.070$, $p=.130$). For adults, no states were very influential. The states with the largest Cook's distance values were New York and Utah, but when excluded from the analysis results were essentially unchanged ($r=0.48$, $R\text{-square}=0.230$, $p=.001$).

When examining the weighted results that account for state population size, we found a moderate positive correlation (weighted $r=0.47$, $R\text{-square}=0.221$, $p=.022$) for state Google cigar relative search volume by youth prevalence of current cigar use. When excluding Utah (which appeared to be influential), the correlation coefficient was a bit lower and did not reach statistical significance (excluding Utah weighted $r=0.25$, $R\text{-square}=0.062$, $p=.139$). The correlation for adults in weighted analysis was lower and not statistically significant (weighted $r=0.20$, $R\text{-square}=0.039$, $p=.254$). However, the state with the largest Cook's distance value was New York; when excluding New York, the correlation coefficient was higher and statistically significant (weighted $r=0.35$, $R\text{-square}=0.119$, $p=.001$).

Smokeless tobacco

Figure 3 presents prevalence of youth (Figure 3a) and adult (Figure 3b) smokeless tobacco use across states as well as smokeless tobacco relative search volume in Google for 2011 (Figure 3c). Youth smokeless tobacco use rates were highest in Kentucky (16.9%), Wyoming (15.1%), South Dakota (14.7%), and West Virginia (14.4%); youth smokeless tobacco use rates were lowest in Hawaii (3.5%), Utah (3.7%), Rhode Island (5.7%), and Illinois (5.8%). Adult smokeless tobacco use rates were highest in Wyoming (9.9%), Montana (9.4%), West Virginia (9.1%), and South Dakota (8.4%); adult smokeless tobacco use rates were lowest in New Jersey (0.8%), Hawaii (0.9%), Rhode Island (1.0%), and District of Columbia (1.2%). Relative search volume for smokeless tobacco (Figure 3c) was highest in North Dakota (100), West Virginia (99), Iowa (81), and Montana (77).

Figure 4 is a scatterplot of state Google smokeless tobacco relative search volume by youth and adult prevalence of current smokeless tobacco use in 2011. Again, we found a moderate positive correlation for youth ($r=0.46$, $R\text{-square}=0.209$, $p=.003$) (Figure 4a). Correspondingly, for adult smokeless tobacco use prevalence (Figure 4b), we also found a moderate positive correlation ($r=0.48$, $R\text{-square}=0.226$, $p<.001$). When excluding Wyoming,

the most influential state, the correlation coefficients were much higher (youth $r=0.66$, $R\text{-square}=0.440$, $p<.001$; adult $r=0.70$, $R\text{-square}=0.495$, $p<.001$). Notably, Wyoming had zero relative search volume for smokeless tobacco but high prevalence of use.

When examining the weighted results that account for state population size, we found a strong positive correlation between state Google smokeless tobacco relative search volume and youth prevalence of current smokeless tobacco use (weighted $r=0.66$, $R\text{-square}=0.432$, $p<.001$), similar to unweighted analysis when excluding Wyoming. Even when excluding the most influential states (Kentucky and Texas) the correlation was still quite strong (weighted $r=0.59$, $R\text{-square}=0.354$, $p<.001$). Correspondingly, for adult smokeless tobacco prevalence, we also found a strong positive correlation (weighted $r=0.67$, $R\text{-square}=0.454$, $p<.001$). Again, when excluding the most influential states (California and New York), the correlation coefficient was still relatively strong (weighted $r=0.58$, $R\text{-square}=0.331$, $p<.001$).

DISCUSSION

The relative volume of Google searches for non-cigarette tobacco products differed across states and illustrate how geographically varied the public is in their relative search volume trends of these products. For example, Texas and Georgia had high Google relative search volumes for cigar products but were relatively low in their Google searches for smokeless tobacco products. In contrast, North Dakota and West Virginia had relatively high Google relative search volumes for smokeless tobacco products but were comparatively low in their Google searches for cigars. Correspondingly, youth use of these tobacco products in these states matched with the state Google relative search volume trends. When examining these associations across states and among youth and adults, we found a moderate positive correlation between relative volume of Google searches for non-cigarette tobacco products and prevalence of youth and adult use of cigars and smokeless tobacco products. Because tobacco control efforts are often regulated at the state-level, these primary results are derived from unweighted analyses where data from all states are considered equal, irrespective of the state's population size. Nevertheless, associations remained relatively consistent even when accounting for population size of the states with the exception of adult tobacco use of cigars, which reached significance after New York, the most influential state in this analysis, was excluded. Thus, while internet relative search volumes have been shown in past studies to detect regional outbreaks of influenza [2, 3] and suicide,[6, 7] our findings validate that this methodology may also be a useful indicator for tracking tobacco use among youth and adults.

The strength of the associations in our study were similar, but slightly lower for cigars, compared with the measures of association found in the past studies that examined the utility of Google Trends as a public health monitoring tool. For example, the proportion of variance in youth suicide behaviors explained by suicide-related internet searches ranged from 0.25 to 0.49.[6] In addition, the $R\text{-square}$ values for the association between kidney stone Google searches and kidney stone disease ranged from 0.25 to 0.37.[22] In Canada, the $R\text{-square}$ value for the association between internet searches of the term "listeriosis" with an early detection of outbreaks was 0.38.[23] In our study, associations were comparable for smokeless tobacco with $R\text{-square}$ values of 0.21 to 0.49, and lower for cigars

with R-square values of 0.07 to 0.24. The proportion of variance in use of these tobacco products explained by Google Trends searches is relatively modest. Based on these findings, we would not recommend that Google Trends data be used as a substitute for traditional surveillance methods, but rather to complement such methods in order to provide an idea of real-time patterns in tobacco product use and interest which could help state tobacco control programs identify prevention needs. We acknowledge the novelty of using Google Trends as a monitoring tool that could undoubtedly be improved upon with further evaluation and refinement.

We found that the correlations between Google Trends relative search volume and prevalence of smokeless tobacco use were somewhat stronger than the correlations between Google Trends relative search volume and prevalence of cigar use. This possibly reflects the distinct demographic and regional variations that exist among smokeless tobacco consumers. For instance, smokeless tobacco use is more common and socially acceptable in rural areas of the United States and across Southern and Midwestern states.[17–20] Thus, it may be that Google Trends data are more sensitive to tobacco use prevalence when distinct geographic patterns of use exist.

To a relatively small degree, Google Trends data better predicted cigar consumption patterns for adults versus youth in unweighted analysis. It may be that the relative volume of Google searches for cigars and smokeless tobacco products is driven in large part by searches from adults who are more likely to consume cigars versus youth.[10] This could also be because of the growing popularity of cigars, in general, but especially among young adults who are 18–25 years old.[21] Irrespective of the cause, our findings provide evidence that Google Trends has the potential to be a valuable surveillance tool for youth and adult tobacco use.

The near real-time monitoring features of Google Trends in conjunction with traditional survey methods could improve researchers and public health officials' ability to stay abreast of emerging trends in tobacco use. Google Trends data are immediate, free, and easy to access for all users of the World Wide Web. In the present study, we document the utility of these data in assisting key stakeholders with tracking tobacco use patterns with reduced cost and in near real-time. Google Trends data might also be useful when developing and/or updating survey questions that query youth on their tobacco product use. For example, the YRBS smokeless tobacco question lists specific brands of this product to more thoroughly question youth about their use. However, brands listed by YRBS included “Redman tobacco” and “Levi Garrett” which yielded lower Google relative search volumes than other smokeless tobacco brands/search terms (e.g., Skoal tobacco, Grizzly tobacco). We therefore speculate if survey developers can similarly use Google Trends data to delineate more popular and contemporary tobacco brands/products and to update questionnaires to improve the measurement of tobacco use.

A limitation of Google Trends is that both adult and youth searches are captured in the data and Google search differences across age groups cannot be considered. In addition, not all online search activities are represented in this database. Other web search engines could have been used (e.g. Bing) or bypassed all together by individuals going directly to certain websites. In addition, we examined only search terms in the English-language and internet

search queries in the United States; thus, generalization of the results to other countries requires caution. Moreover, Google Trends designates a certain threshold of traffic for search terms, so that states with low relative search volume are assigned a zero. Also, the available data are derived from a sample of Google web searches, which could possibly cause a non-representative sampling bias.[15] Only high school youth (primarily attending public schools) were assessed in the YRBS and more vulnerable youth who are not attending school are not assessed. Additionally, youth and adult use of cigars and smokeless tobacco products were the only non-cigarette tobacco products that were queried; therefore, use of e-cigarettes, hookah, and other alternative tobacco products were not assessed and consequently unavailable for analysis. Lastly, validating the use of Google Trends to monitor temporal trends is beyond the scope of this study. Nevertheless, based on data showing increased search volume for electronic cigarettes over time, it has been suggested that surveillance of search queries could be used to monitor the diffusion of new health products.[8] Google Trends data could be useful for monitoring the trends in Google searches in newest tobacco products that are not yet assessed in standard monitoring systems as potential consumers seek out more information.

Despite intensified control policies, tobacco use remains the most preventable cause of premature death and disability in the United States.[24, 25] Therefore, efforts to track tobacco use in near real-time are an important endeavor that could assist key stakeholders with making informed decisions about prevention plans and policies. This is especially relevant in consideration of the new and emerging tobacco products that are increasing in promotion and consumption. [26, 27] In the present study, we leverage free data from Google Trends to study relative volume of searches for tobacco-related terms and obtain deeper insight about youth and adult tobacco use. We do not advise that Google Trends be used as a substitute for standard surveillance tools that have been extensively validated as reliable resources for measuring tobacco use at a population-level; yet, Google Trends data can be used to complement these standard monitoring systems that are typically published with a 1 to 2 year delay. Thus, our findings validate that this accessible methodology may promptly track tobacco use at a population-level.

Acknowledgments

Role of funding source

This publication was made possible by Grant Numbers UL1 RR024992 and KL2 RR024994 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. Other support includes an NIH Career Development Award awarded to Dr. Cavazos-Rehg (NIDA, K01DA025733) and NIDA, R01 DA032843, an NIH Midcareer Investigator Award awarded to Dr. Bierut (K02 DA021237), R01 DA031288 awarded to Dr. Grucza.

References

1. Purcell, K.; Brenner, J.; Rainie, L. Search Engine Use 2012. Washington: Pew Research Center; 2012. <http://www.pewinternet.org/Press-Releases/2012/Search-Engine-Use-2012.aspx> [Accessed 6/6/2013]
2. Carneiro HA, Mylonakis E. Google Trends: A web-based tool for real-time surveillance of disease outbreaks. Clin Infect Dis. 2009; 49:1557–64. [PubMed: 19845471]

3. Ginsberg J, Mohebbi MH, Patel R, et al. Detecting influenza epidemics using search engine query data. *Nature*. 2009; 457:1012–U4. [PubMed: 19020500]
4. Wilson K, Brownstein JS. Early detection of disease outbreaks using the Internet. *CMAJ*. 2009; 180:829–31. [PubMed: 19364791]
5. Seifter A, Schwarzwald A, Geis K, et al. The utility of “Google Trends” for epidemiological research: Lyme disease as an example. *Geospat Health*. 2010; 4:135–7. [PubMed: 20503183]
6. McCarthy MJ. Internet monitoring of suicide risk in the population. *J Affect Disord*. 2010; 122:277–79. [PubMed: 19748681]
7. Gunn JF III, Lester D. Using google searches on the internet to monitor suicidal behavior. *J Affect Disord*. 2013; 148(2–3):411–2. [PubMed: 23182592]
8. Ayers JW, Ribisl KM, Brownstein JS. Tracking the rise in popularity of electronic nicotine delivery systems (electronic cigarettes) using search query surveillance. *Am J Prev Med*. 2011; 40(4):448–53. [PubMed: 21406279]
9. Askatas N, Zimmermann KF. Google econometrics and unemployment forecasting. *Applied Economics Quarterly*. 2009; 55(2):107–20.
10. Substance Abuse and Mental Health Services Administration. Results from the 2012 National Survey on Drug Use and Health: Summary of National Findings. Rockville, MD: Substance Abuse and Mental Health Services Administration; 2013. NSDUH Series H-46, HHS Publication No. (SMA) 13-4795
11. Centers for Disease Control and Prevention. Current tobacco use among middle and high school students—United States, 2011. *MMWR Morb Mortal Wkly Rep*. 2012; 61(31):581–5. [PubMed: 22874835]
12. U.S. Department of Health and Human Services. How tobacco smoke causes disease: the biology and behavioral basis for smoking-attributable disease: a report of the surgeon general. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2010.
13. Centers for Disease Control and Prevention. Methodology of the youth risk behavior surveillance system, 2013. *MMWR Morb Mortal Wkly Rep*. 2013; 62(1):1–20. [PubMed: 23302815]
14. Gruzca RA, Abbacchi AM, Przybeck TR. Discrepancies in estimates of prevalence and correlates of substance use and disorders between two national surveys. *Addiction*. 2007; 102(4):623–9. [PubMed: 17309538]
15. [Accessed 6/20/2013] About Google Trends. 2009. <http://www.google.com/intl/en/trends/about.html#7>
16. [Accessed 10/18/2013] Tobacco: Euromonitor from trade sources/national statistics. 2012 <http://www.portal.euromonitor.com>
17. Gfroerer JC, Larson SL, Colliver JD. Drug use patterns and trends in rural communities. *J Rural Health*. 2007; 23:10–5. [PubMed: 18237319]
18. Hanson CL, Novilla M, Lelinneth LB, et al. Using the rural-urban continuum to explore adolescent alcohol, tobacco, and other drug use in Montana. *J Child Adoles Subst*. 2008; 18(1):93–105.
19. Rodu B, Cole P. Smokeless tobacco use among men in the United States, 2000 and 2005. *J Oral Pathol Med*. 2009; 38(7):545–50. [PubMed: 19453847]
20. Welte JW, Barnes GM, Tidwell MCO, et al. Tobacco use, heavy use, and dependence among adolescents and young adults in the United States. *Subst Use Misuse*. 2011; 46(9):1090–8. [PubMed: 21391810]
21. Kozlowski LT, Dollar KM, Giovino GA. Cigar/cigarillo surveillance: limitations of the U.S. department of agriculture system. *Am J Prev Med*. 2008; 34(5):424–6. [PubMed: 18407010]
22. Willard SD, Nguyen MM. Internet search trends analysis tools can provide real-time data on kidney stone disease in the United States. *Urology*. 2013; 81(1):37–42. [PubMed: 21676450]
23. Wilson K, Brownstein JS. Early detection of disease outbreaks using the Internet. *CMAJ*. 2009; 180(8):829–31. [PubMed: 19364791]
24. Centers for Disease Control and Prevention. Smoking-attributable mortality, years of potential life lost, and productivity losses—United States, 2000–2004. *MMWR Morb Mortal Wkly Rep*. 2008; 57(45):1226–28. [PubMed: 19008791]

25. Viswanath K, Herbst RS, Land SR, et al. Tobacco and cancer: An American association for cancer research policy statement. *Cancer Res.* 2010; 70(9):3419–30. [PubMed: 20388799]
26. Delnevo CD, Wackowski OA, Giovenco DP, et al. Examining market trends in the United States smokeless tobacco use: 2005–2011. *Tob Control.* 2012
27. Carpenter CM, Connolly GN, Ayo-Yusuf OA, et al. Developing smokeless tobacco products for smokers: an examination of tobacco industry documents. *Tob Control.* 2009; 18(1):54–9. [PubMed: 18948390]

Figure 1.

Choropleth maps of youth prevalence of current cigar use in 2011 (Figure 1a), adult prevalence of current cigar use in 2010–2011 (Figure 1b), and Google relative search volume for cigars in 2011 (Figure 1c). White states in Figure 1a represent states where no youth prevalence data were available in 2011.

Figure 2.

Scatterplots with regression line of Google relative search volume for cigars and current use of cigars among youth and adults by state. Both adult and youth searches are captured in the Google Trends data. Shaded area represents 95% confidence limits for mean predicted values. Figure 2a presents a scatterplot of 2011 Google relative search volume for cigars with 2011 prevalence of current use of cigars/cigarillos/little cigars among *youth* (source 2011 YRBS). Analysis was restricted to 36 states. YRBS data on prevalence of youth cigar use was not available for the remaining states. When excluding RI and UT, $r=0.26$, $R\text{-square}=0.070$, $p=.130$. Figure 2b presents a scatterplot of 2011 Google relative search volume for cigars with 2010–2011 prevalence of current use of any type of cigar among *adults* (source 2010–2011 NSDUH). Sample size includes 50 states plus the District of Columbia. When excluding NY and UT, results were similar ($r=0.48$, $R\text{-square}=0.230$, $p=.001$).

Figure 3. Choropleth maps of youth prevalence of current smokeless tobacco use in 2011 (Figure 3a), adult prevalence of current smokeless tobacco use in 2010–2011 (Figure 3b), and Google relative search volume for smokeless tobacco in 2011 (Figure 3c). White states in Figure 3a represent states where no youth prevalence data were available in 2011.

Figure 4.

Scatterplots with regression line of Google relative search volume for smokeless tobacco and current use of smokeless tobacco among youth and adults by state. Both adult and youth searches are captured in the Google Trends data. Shaded area represents 95% confidence limits for mean predicted values. Figure 4a presents a scatterplot of 2011 Google relative search volume for smokeless tobacco with 2011 prevalence of current use of chewing tobacco, snuff, or dip among *youth* (source 2011 YRBS). Analysis was restricted to 40 states. YRBS data on prevalence of youth smokeless tobacco use was not available for the remaining states. When excluding WY, $r=0.66$, $R\text{-square}=0.440$, $p<.001$. Figure 4b presents a scatterplot of 2011 Google relative search volume for smokeless tobacco with 2010–2011 prevalence of current use of chewing tobacco or snuff among *adults* (source 2010–2011

NSDUH). Sample size includes 50 states plus the District of Columbia. When excluding WY, $r=0.70$, $R\text{-square}=0.495$, $p<.001$.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript