Digital Detection for Tobacco Control: Online Reactions to the 2009 U.S. Cigarette Excise Tax Increase

John W. Ayers PhD, MA[1], **Benjamin M. Althouse** PhD, ScM[2], **Kurt M. Ribisl** PhD[3,4], **Sherry Emery** PhD, MBA[5]

[1]*Graduate School of Public Health, San Diego State University, San Diego, CA;* [2]*Santa Fe Institute, Santa Fe, NM;* [3]*Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC;* [4]*Lineberger Comprehensive Cancer Center, Chapel Hill, NC;* [5]*Institute for Health Research and Policy, University of Illinois at Chicago, Chicago, IL*

Corresponding Author: John W. Ayers, PhD, MA, Graduate School of Public Health, San Diego State University, 9245 Sky Park Court, Suite 230, San Diego, CA 92123, USA. Telephone: 619-371-1846; Fax: 858-505-4770; E-mail: ayers.john.w@gmail.com

Received April 22, 2013; accepted October 15, 2013

ABSTRACT

Introduction: The Internet is revolutionizing tobacco control, but few have harnessed the Web for surveillance. We demonstrate for the first time an approach for analyzing aggregate Internet search queries that captures precise changes in population considerations about tobacco.

Methods: We compared tobacco-related Google queries originating in the United States during the week of the State Children's Health Insurance Program (SCHIP) 2009 cigarette excise tax increase with a historic baseline. Specific queries were then ranked according to their relative increases while also considering approximations of changes in absolute search volume.

Results: Individual queries with the largest relative increases the week of the SCHIP tax were "cigarettes Indian reservations" 640% (95% CI = 472–918), "free cigarettes online" 557% (95% CI = 432–756), and "Indian reservations cigarettes" 542% (95% CI = 414–733), amounting to about 7,500 excess searches. By themes, the largest relative increases were tribal cigarettes 246% (95% CI = 228–265), "free" cigarettes 215% (95% CI = 191–242), and cigarette stores 176% (95% CI = 160–193), accounting for 21,000, 27,000, and 90,000 excess queries. All avoidance queries, including those aforementioned themes, relatively increased 150% (95% CI = 144–155) or 550,000 from their baseline. All cessation queries increased 46% (95% CI = 44–48), or 175,000, around SCHIP; including themes for "cold turkey" 19% (95% CI = 11–27) or 2,600, cessation products 47% (95% CI = 44–50) or 78,000, and dubious cessation approaches (e.g., hypnosis) 40% (95% CI = 33–47) or 2,300.

Conclusions: The SCHIP tax motivated specific changes in population considerations. Our strategy can support evaluations that temporally link tobacco control measures with instantaneous population reactions, as well as serve as a springboard for traditional studies, for example, including survey questionnaire design.

INTRODUCTION

Annual telephone surveys have been the principal tobacco control sentinel for decades (Giovino et al., 2009). However, the value of such surveys is diminishing due to respondents' increasing unwillingness to participate (Curtin, Presser, & Singer, 2005; Groves, 2006) and rising administration costs (Boland, Sweeney, Scallan, Harrington, & Staines, 2006). This is further complicated by restricted data sharing models, with some survey datasets remaining hidden from public consumption for years or indefinitely (Chan et al., 2010; King, 2011). At the same time, public health officials are demanding more surveillance (Brownson, Fielding, & Maylahn, 2009; U.S. Department of Health and Human Services, n.d.). The *Centers for Disease Control and Prevention's Office on Smoking and Health*, for instance, named rapid surveillance a major research agenda (Centers for Disease Control,

2011). Such calls are being underscored by expanding *Food and Drug Administration* tobacco product regulatory authority and its surveillance needs (Leischow, Zeller, & Backinger, 2012).

Passively collected online, digital data are free, publicly available, and can be quickly analyzed (Ayers, Althouse, Allem, Rosenquist, & Ford, 2013; Eysenbach, 2011). Yet, tobacco control investigators are not harnessing these data and risk falling behind how the tobacco industry is using digital avenues to undermine control measures (Freeman, 2012; Ribisl & Jo, 2012) or promote their products (Yamin, Bitton, & Bates, 2010).

Pioneering work has laid a foundation for the development of digital detection for tobacco control. However, these early analyses have been limited, building on traditional descriptive analytic approaches adapted from survey-based surveillance where the aim has been to describe a single or maybe two conceptual outcomes, such as those from categorizing twitter

accounts (Prochaska, Pechmann, Kim, & Leonhardt, 2012), news articles (Ayers et al., 2012), or Internet search queries (Ayers, Althouse, Johnson, & Cohen, 2013; Ayers, Ribisl, & Brownstein, 2011b; Cobb, 2010). In one of the earliest applications of digital detection for tobacco control, for instance, we described a formative evaluation of the State Children's Health Insurance Program (SCHIP) April 1, 2009 U.S. cigarette excise tax increase using a pool of 24 investigator-selected queries, intended to be indicative of spikes in cessation or tax avoidance considerations (Ayers, Ribisl, & Brownstein, 2011a).

In this report, we describe a novel method designed specifically for applications to digital data in tobacco control by analyzing precise changes in Internet search queries around the SCHIP tax. The approach we propose deviates from the traditional model, where investigators specify their hypotheses and key outcome measures a priori. Informed by data mining strategies (Paul & Dredze, 2011; Shah & Tenenbaum, 2012), we aimed to identify how population, tobacco-related considerations changed around SCHIP as a case study. We did so by analyzing both the content and volume of hundreds of unique Google search queries. Processing 1,776 systematically selected individual search query trends, we analyzed deviations from preexisting patterns, ranked these deviations according to their relative increase while also noting potential raw increases in search volume, and clustered queries into thematic categories based on their content for data-driven explanations. These can then be used to not only evaluate the tax increase but also define new research questions and inform the design of traditional studies and interventions, better realizing the potential of digital data.

METHODS

Data were downloaded from *Google Trends* (google.com/trends), a public facing database of Google queries. To populate our list of queries, we began with a list of 16 root terms adapted from our prior analysis of the SCHIP tax: "quit smoking," "stop smoking," "nicotine replacement," "nicotine patch," "nicotine gum," "chantix," "quit smoking hyponosis," "quit smoking cold turkey," "quit smoking lasers," "discount cigarettes," "cheap cigarettes," "cigarette coupon," "Indian cigarettes," "online cigarettes," "tax-free cigarettes," and "duty-free cigarettes. We then downloaded the top 10 most related terms for each of the root 16 as selected by *Google Trends* based on both the content of the queries (e.g., containing the same or similar terms as the root queries) and the tendency for a user to also search for these terms in the same session he/she searched for the root term. From these 160, the next top 10 most related queries were selected for an additional 1,600 queries. Related-term selection is done automatically by *Google Trends*. This approach produced 1,776 potential queries for analysis, from which 92 very low volume terms (as indicated by missing data), 845 duplicates, and 658 unrelated or unclear (e.g., "the patch") terms were purged, yielding 181 unique and relevant queries for analysis.

Each term (or query) was captured on a daily scale, normalized (relative search volume, RSV) to the highest proportion of searches for the term relative to all searches. RSV is reported on a 0 to 100 scale with 100 being the day with the highest proportion of searches for the term relative to all queries and RSV = 50 meaning another day had 50% of that highest search proportion. This relative approach corrects for trending (e.g., all queries may be increasing) and therefore if

we see a RSV spike for a tobacco-related term, it is a logical deduction that other queries did not also spike (Dutka & Hanson, 1989). We captured data for about 2,550 days (7 years) for each of the queries but focus our analyses on the beginning of 2009, the window in which SCHIP took effect.

Since our earlier analyses (Ayers et al., 2011a) suggested that the SCHIP tax produced a pulse effect (a spike in search volume) around the tax, and our goal is to describe the impact of the SCHIP across numerous individual and thematic groups of search queries, we focused on estimating a single-treatment effect describing any spike in search volume during the week SCHIP took effect. Quantitatively, we focused on estimating the relative mean search volume for the week of the tax increase compared with a typical search volume from a period before the tax for each of the 181 retained search queries. This approach is commonly referred to as an interrupted time series (Lewis-Beck, 1986) or quasi-experimental design (Shadish & Cook, 2009), since we are comparing how an outcome (RSV) changed around an intervention/treatment. Specifically, we specified regression equations to estimate differences in the mean search volume aggregating daily trends for the week of SCHIP (seven daily estimates, March 29 to April 4) compared with an aggregated period before SCHIP (28 daily estimates, February 1–28, 2009), excluding a washout period (March 1–28) and any periods before February 2009. These exclusions were necessary to ensure the reference period would be indicative of both usual and recent RSV, unaffected by SCHIP (washout) and seasonal trends, such as New Years Day, or other interventions occurring nationally or in states (omission of pre-February 2009).

We modeled the difference between the SCHIP period and the pre period as a percent increase, scaling the difference in mean RSV between SCHIP and the pre period by the mean RSV for the pre period (e.g., [SCHIP-Pre]/Pre). We then ranked individual queries by the estimated percent increase to identify which queries had the largest relative spike in search volume. To better understand these rankings, we then combined queries into thematic groups for qualitative interpretations, based on expert agreement among the authors, estimating the relative mean increase for these clusters of queries. CIs around these quotients were estimated by simulating 10,000 random draws from the multivariate normal sampling distribution with mean equal to the maximum-likelihood point estimates and variance equal to the variance–covariance matrix, that is, $\hat{\beta}_{est} \sim MVN(\hat{\beta}_{mle}, V(\hat{\beta}_{mle}))$, where $\hat{\beta}_{mle}$ is the vector of the regression coefficients and $V(\hat{\beta}_{mle})$ its variance–covariance matrix from the regression (King, Tomz, & Wittenberg, 2000).

Finally, to supplement our understanding of the relative rankings of queries and to demonstrate practical significance, we crudely estimated raw search volume from *Google Adwords* (google.com/adwords), a user interface that provides raw search volume for online marketers. The volume estimates from *Adwords* represent the typical monthly search volume for a specific query estimated from a long time series, as a result, we transformed this monthly volume by multiplying the value by 12 (months), then dividing by 52 (weeks), and then multiplying this quotient by the estimated relative increase during the SCHIP week for the queries analyzed. This approach assumes that the baseline period to which we compare the week of SCHIP had a raw volume that was close to the overall mean RSV. We evaluated this assumption by comparing relative query volumes from the

baseline period to other time periods before and after SCHIP. This assumption largely holds, but interpretations of raw volumes should be viewed as supplemental to those involving changes in RSV. *Google Adwords* did not return volume for two terms and this has been noted in the results and corresponding figure.

RESULTS

Figure 1 shows a spaghetti plot of all 181 daily tobacco-related search queries around SCHIP. Nearly all individual query trends peaked the day of (or around) SCHIP as indicated by the overlapping of trends near RSV = 100 on April 1. Averaging across all terms tobacco-related queries relatively increased about 79% (95% CI = 77–82) for the week around SCHIP (including 3 days before and after the tax hike) compared with a usual period before the tax. However, this aggregate approach masks which tobacco-related queries were driving the mean spike around SCHIP.

Figure 2 shows rankings by relative increase the week of SCHIP for the statistically significantly increasing queries, totally roughly 962,400 excess searches for the 162 statistically significantly spiking queries. All tax avoidance–related queries had the largest spike in relative query volume. "Cigarettes Indian reservations" rose 640% (95% CI = 472–918) compared with their relative baseline mean, "free cigarettes online" 557% (95% CI = 432–756) and "Indian reservations cigarettes" 542% (95% CI = 414–733), totaling to about 7,500 excess queries. "Nicotine in system" rose relatively 12% (95% CI = 4–21) or 600 in absolute terms and "quit smoking tips" 12% (95% CI = 3–22) or 230 were among the lesser relatively changed but significantly spiking queries. Focusing on raw volume alone, the largest increasing queries were "how to quit," "why stop smoking," "cigarettes online," "online cigarettes," and "quit smoking," totaling about 225,000 excess queries the week of SCHIP.

Among the entire list of statistically significant relatively spiking queries, several themes emerged, including general interest in the tax, online shopping for cigarettes, coupons for both cigarettes and cessation products, cessation aids or techniques, smoking risks, etc. But the two most common overarching themes were general tax avoidance, such as "discount cigarettes" relatively increasing 159% (95% CI = 116–211) or about 5,500 in raw queries, and general cessation, such as "tips quitting smoking" relatively increasing 67% (95% CI = 32–106) or about 1,200 in raw queries.

Aggregated tax avoidance–related queries relatively increased 150% (95% CI = 144–155) from their baseline or about 550,000 in absolute excess queries, with trends shown in Figure 3. In comparison, cessation-related queries relatively increased 47% (95% CI = 44–49) from their baseline or about 175,000, suggesting avoidance queries were 2.6 times more searched the week of SCHIP ($p < .0001$). After the top three queries ("cigarettes Indian reservations," "free cigarettes online," and "Indian reservations cigarettes"), the top five avoidance queries with the largest relative increases also included "Indian cigarettes online" rising 491% (95% CI = 394–626) or about 800 queries and "Indian cigarettes" 426% (95% CI = 327–572) or about 8,000 queries. The top five relatively increasing cessation queries around SCHIP included "nicoderm cq coupons" rising 426% (95% CI = 286–682) or about 2,400 in absolute terms, "chantix coupons" 285% (95% CI = 205–401) or 5,300, "help quitting smoking" 252% (95% CI = 161–388) or 5,300, "nicoderm coupons" 220% (95% CI = 172–279) or 2,200, and "stop smoking patches" 208% (95% CI = 138–309) or 750. Among the highest relatively increasing avoidance-related queries, there was a focus on attempting to purchase cigarettes from Native American Web sites, many of whom claim they can sell "tax free." Among the most relatively increasing cessation-related queries, there was a focus on cessation products, including discounted medications, and to a lesser extent general cessation resources. All the queries can be further subcategorized like this to capture such specific themes, including both those related and unrelated to avoidance and cessation altogether.

Figure 4 displays rankings among smaller thematic groupings of queries by their relative increase in search volume the week of SCHIP compared with their baseline means, omitting clusters with fewer than two search terms. These provide evidence of relative increases in general interest in the tax, including terms such as "cigarette tax," rising 350% (95% CI = 255–492) or about 22,000 the week of SCHIP and interest
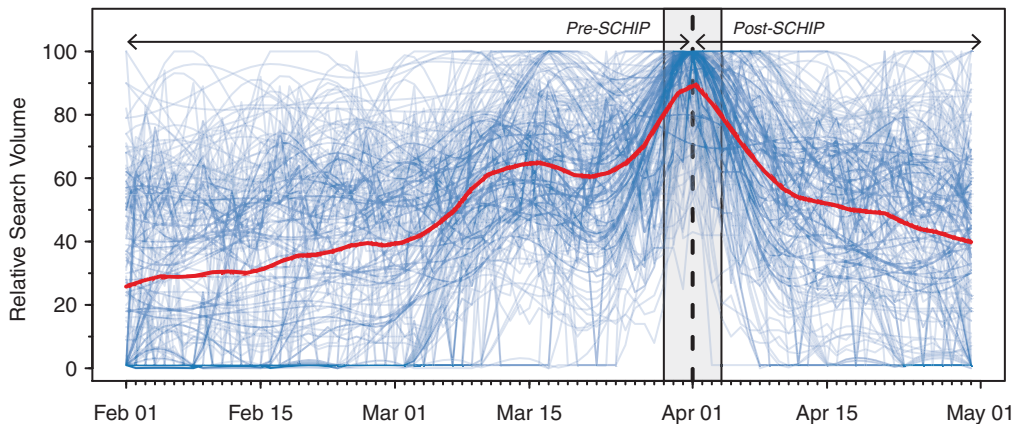

Figure 1. Time trends for all tobacco-related queries. Figure shows time trends before and after the SCHIP tax for all tobacco-related queries analyzed (*N* = 181). Each line indicates estimated search volume (relative search volume) for a specific query. The red line indicates the mean across all queries, and the shaded region highlights the 7-day period around, with the dashed line indicating the day of, SCHIP.
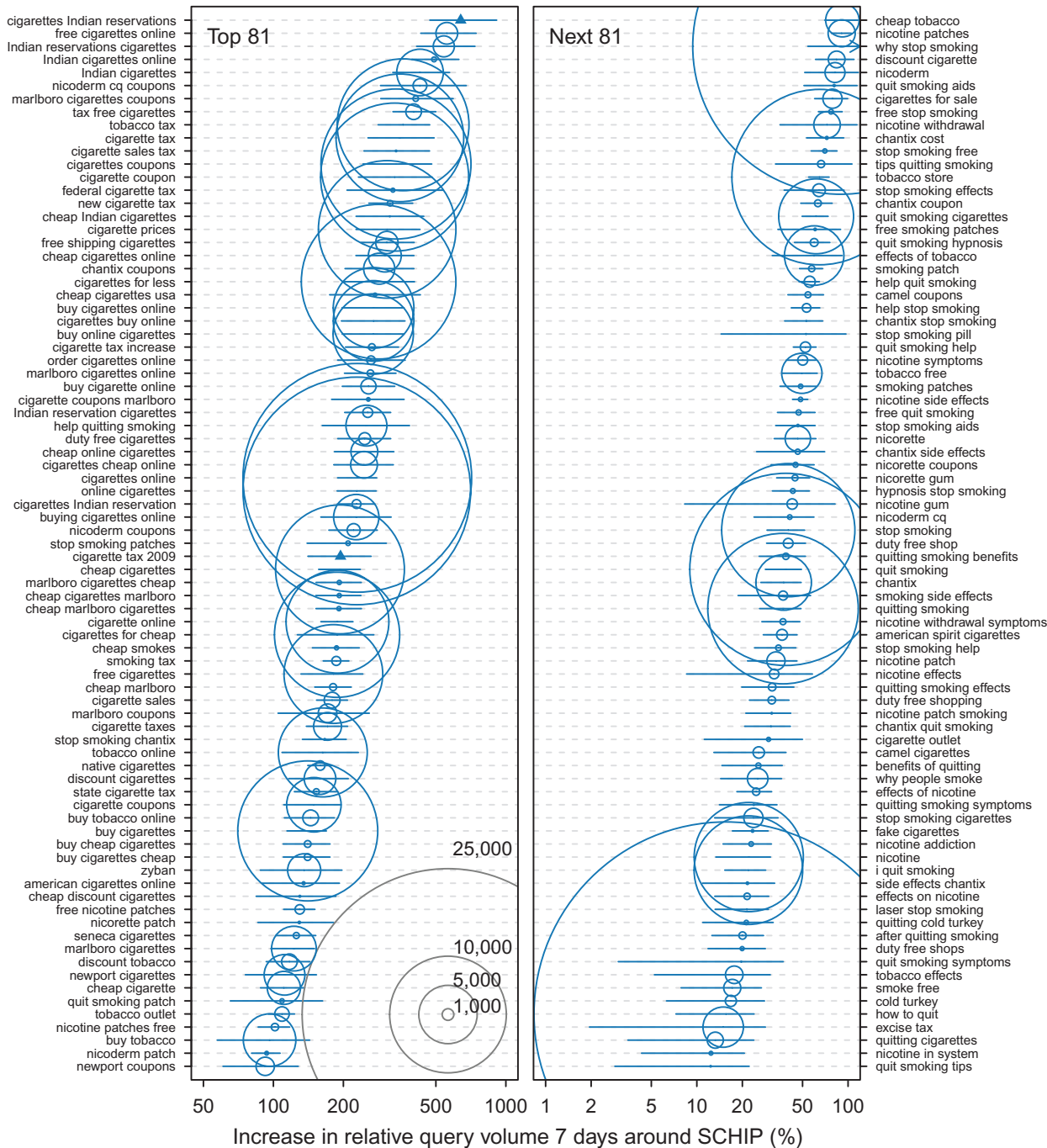
Figure 2. Ranking of relative and absolute changes for specific tobacco-related queries around the SCHIP tax. Each node represents the mean percent increase (with node size corresponding to estimates of excess absolute volume from *Google Adwords*, with triangle nodes indicating no volume estimate was returned by *Adwords*), and each line represents the 95% CI for that increase, as estimated from an interrupted time series, comparing the ratio of search volume (relative search volume, RSV) for the 7 days (including three before and three after) around SCHIP by search volume before SCHIP. Queries are ranked by the mean percent increase according to RSV, with only statistically significant associations shown (N = 162).

in cigarette prices, including terms such as "cigarette prices," rising 309% (95% CI = 227–427) or about 23,500.

Overall, queries for Native American or tribal cigarettes had the largest relative increases around SCHIP (246%; 95% CI = 228–265 or about 21,000 in raw terms). "Free" cigarette queries relatively rose 215% (95% CI = 191–242) or about 27,000, which included queries with the combination of

"free" with cigarette(s), for example, "free shipping cigarettes" or simply "free cigarettes." It was also of interest to identify other terms commonly used in conjunction with cigarettes as thematic groups, such as "buy" relatively increasing 137% (95% CI = 128–147) or about 200,000, "cheap" 78% (95% CI = 69–86) or 42,000, and "tax-free"/"duty-free" 77% (95% CI = 69–86) or about 42,000. Queries for cigarette retailers
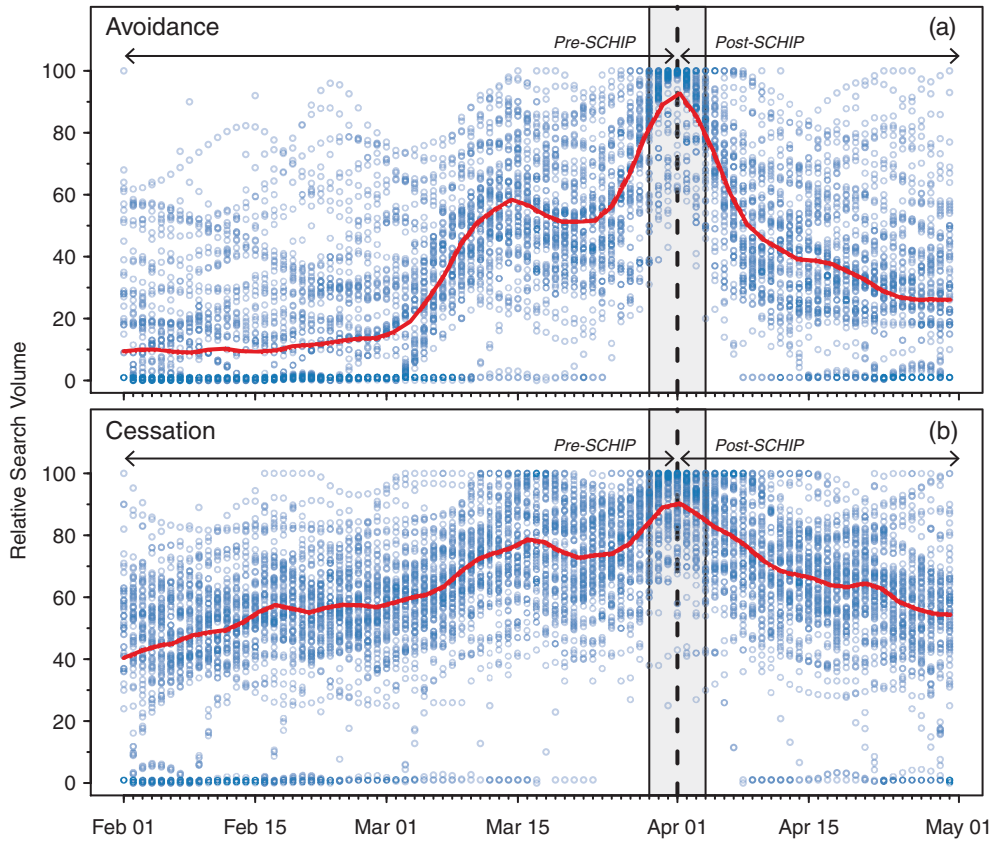
Figure 3. Time trends for tax avoidance and smoking cessation queries around the SCHIP tax. Each panel shows time trends before and after SCHIP for (a) tax avoidance–related queries and (b) smoking cessation–related queries. Each dot indicates an estimated search volume (relative search volume) for a specific query, including 73 and 74 related queries for (a) and (b). Dark red solid lines indicates the mean trend from specific queries, and the shaded region highlights the 7-day period around, with the dashed line indicating the day of, the SCHIP tax.
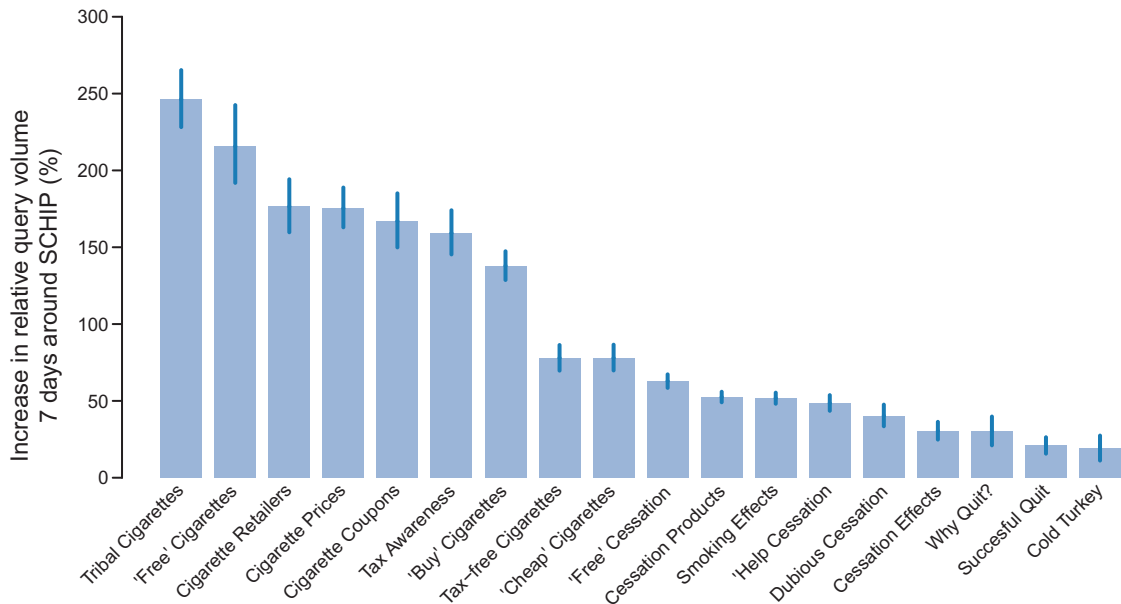

Figure 4. Ranking of relative changes for post-hoc identified categories of tobacco-related queries around the SCHIP tax. Each bar represents the mean percent increase in search volume, and each line represents the 95% CI for that increase, as estimated from an interrupted time series, comparing the ratio of search volume (relative search volume, RSV) for the 7 days (including three before and three after) around SCHIP by search volume before SCHIP. Queries are ranked by the mean percent increase according to RSV. Each bar is indicated by a group of similar queries, with the specific queries included in each category described in the text.

relatively rose 176% (95% CI = 160–193) and cigarette coupons for potential use at point of sale relatively rose 166% (95% CI = 149–184), representing about 90,000 and 70,000 excess queries the week of SCHIP, respectively.

Among cessation-related queries, we also found common tag words, such as "free" cessation or "help" cessation with each relatively rising 60% (95% CI = 55–64) and 48% (95% CI = 43–53) or in absolute terms 55,000 and 6,000. Beyond these general searches, interest in specific cessation aids also had strong relative increases. Queries for evidence-based cessation aids, such as nicotine replacement therapies, relatively increased about 53% (95% CI = 49–56) or about 78,000 the week of SCHIP. However, interest in dubious cessation products or approaches also relatively increased 40% (95% CI = 33–47) or about 2,300, including queries such as "quit smoking hypnosis" or "laser stop smoking." At the same time, relative spikes in interest for the most commonly used cessation approach quitting "cold turkey" rose 19% (95% CI = 11–27) amounting to about 2,600 more queries the week of SCHIP.

The week of SCHIP, health interest spiked relatively. Interest in the effects of smoking relatively increased 51% (95% CI = 47–54) or about 75,000 in total queries, including queries such as "nicotine side effects," "smoking side effects," and "effects of nicotine." At the same time, interest in the side effects of cessation also increased, collectively rising by 31% (95% CI = 25–36) or about 10,000 queries, including "Nicotine withdrawal [sp]," "stop smoking effects," and "quitting smoking symptoms" queries. Queries focused on reasons for quitting relatively increased 30% (95% CI = 21–39) or about 55,000 collectively, including queries like "why stop smoking." Last, a few relatively increasing queries indicated the person searching had already successfully quit (or knew someone who had). Specifically, "I quit smoking" relatively increased 22% (95% CI = 15–28) with the overall category of similar queries rising 20% (95% CI = 15–26) or about 20,000 queries.

DISCUSSION

Digital detection represents a potentially powerful surveillance supplement for tobacco control, helping to achieve a timely and cost-effective evaluation of tobacco-related contemplations. Building on our earlier work (Ayers et al., 2011b, 2012; Ayers, Althouse, Johnson, et al., 2013) that also showed SCHIP produced immediate changes in smoking cessation and tax avoidance contemplations (Ayers et al., 2011a), our expanded analyses of unique, systematically collected queries allowed us to figuratively get inside the head of searchers. Changes ranged from the broad to the specific, comprising general interest in the tax, cigarette prices, tax avoidance (including seeking tribal cigarettes, "free" cigarettes, cigarette stores, cigarette coupons, "buying" cigarettes, "cheap" cigarettes, and "tax-free" cigarettes), cessation (including seeking "free" cessation, "help" with cessation, cessation products, dubious cessation methods, and cold turkey methods), health (including smoking effects, cessation side effects, and reasons for quitting), and indications the searcher had quit themselves or knew someone who quit. In total, these changes amounted to about 962,400 excess Google searches, or contemplations, relating to smoking behaviors the week of SCHIP.

Generally, our findings highlight the utility of digital detection for tobacco control. The availability of reliable annual or semiannual tobacco trends from the Behavioral Risk Factor Surveillance System (BRFSS) and Tobacco Use Supplement to the Current Population Survey (TUS-CPS) has significantly enhanced our understanding of population tobacco use behavior (Pierce, Messer, White, Cowling, & Thomas, 2011). However, when multiple interventions (sometimes dozens) occur between two cross-sectional survey-based assessments, attributing year-to-year changes to any single intervention is impossible. Between 2008 and 2010, for example, what proportion of tobacco-related behaviors (or contemplations) may be attributable to SCHIP or the numerous smoking bans and tax increases in several U.S. states? With daily trends, changes in the tobacco control environment can be temporally associated with proximal changes in population information seeking. In the case of SCHIP, we observed the highest relative peak in information seeking the day of the tax. As a result, each tobacco control measure may be judged by how they relatively motivate information seeking in the immediate period around the intervention, and these can be further linked to crudely estimated raw changes in search volume to judge practical significance.

Identifying finer grained trends, especially for the intervening periods between national surveys, has been a major agenda in tobacco control. By harnessing search query archives (or other online digital data), researchers are no longer limited to single outcomes or temporally disconnected evidence. As evidenced by the hundreds of unique outcomes identified in this report, digital surveillance supports multiple inquiries, with little cost and instantaneous data availability and sharing. Moreover, our hypothesis-free approach, as popularized in films like *Moneyball*, can potentially identify changes in population information seeking (and tobacco contemplations) that are unexpected to investigators and therefore serve as a formative base for traditional inquiry (Hastie, Tibshirani, & Friedman, 2009). Specifically, analyses similar to those we present can be used to select questions for the BRFSS and TUS-CPS surveys, identify new tobacco products (Ayers et al., 2011b), and new tobacco-related contemplations for intervention. In these cases, digital detection for tobacco control, in addition to supplanting traditional methods, can make them both better and more cost-effective. For instance, our approach can be scaled up to analyze tens of thousands of queries over large expanses of time, including prospectively.

Such an approach can yield new insights or reinforce known implications, such as those detailed within regarding how the population immediately reacted to a significant increase in the federal tax on cigarettes. First, queries for tribal cigarettes (Choi, Hennrikus, Forster, & St Claire, 2012; Samuel, Ribisl, & Williams, 2012) had substantially larger relative increases around SCHIP than all other queries, as many as 75 times larger than some queries, and in absolute terms, this category of queries was also among the most absolute searched the week of SCHIP. This stresses the need for closing loopholes that allow smokers to avoid one of the most proven cessation promoting policies: increased taxes (Chaloupka, Yurekli, & Fong, 2012). Second, even when the population seeks out cessation resources, gaps in how they conceptualize this interest exist. Queries for dubious (or potentially dangerous) cessation methods relatively increased at fast rates. Even though our subsample of queries only suggest few excess dubious cessation method searches, estimates from another study suggest only

34% of cessation search sessions ever result in being linked to a professional cessation service (Cobb, 2010). Unless advocates take into account the need for guiding how the population thinks about cessation (as we saw in this study), modest rates of conversion may persist.

Digital surveillance has several ongoing limitations that outline additional areas of refinement in our approach. Our approach necessitates the population of interest be connected to the Internet. Likewise, changes in information seeking among the population of Google searchers must correspond to the population of smokers to derive valid trends. Factors like younger age, more income, and more education have been associated with using the Web as a health resource (Cotten & Gupta, 2004). However, recent work suggests that individuals 60+ years of age and adolescents have similar tendencies to search online for health information and nearly all age-by-demographic breakdowns consume some health information online, calling into question the assumption that Internet users differ dramatically from the entire population (McMullan, 2006; Ybarra & Suman, 2008). Unfortunately no contemporary studies exist comparing search propensity among smokers. Smokers who used the Internet in 2003 were more likely to have a high school or greater education, have an annual household income in excess of $50K, and be 10 years younger than off-line smokers (Stoddard & Augustson, 2006), but given increased Internet access (especially through smartphones, popular among low-income populations), potential differences have likely eroded but need to be verified (Backinger & Augustson, 2011). Second, evaluations interested in measuring smoking prevalence cannot be achieved using our approach *at this time*, as systems have not been developed to validly forecast a population tobacco trend as with other infectious disease outcomes, such as Dengue (Althouse, Ng, & Cummings, 2011).

Tobacco control is poised for rapid change on the heels of technologic development. Already, we have seen cessation-counseling move from brick-and-mortar settings and telephone quitlines, to the Internet (Myung, McDonnell, Kazinets, Seo, & Moskowitz, 2009), with little or no degradation in effectiveness (Graham et al., 2011). Extending our use of online resources to include population surveillance for tobacco control is a logical next step. Strong examples from other health domains and computer science can be adapted to tobacco control research (Paul & Dredze, 2011; Shah & Tenenbaum, 2012). However, applications of digital surveillance to tobacco control will require unique approaches that have not been developed in other fields, such as those used herein. In this report, we highlight how mining Internet search queries can shed light on a diverse range of immediate population responses to a tobacco control measure. More importantly, given the instantaneous availability of our data and the cost-effective manner by which our data are investigated, the approach we utilize has strong potential for routinized population tobacco control evaluations and formative data generation working hand-in-hand with research derived from more traditional sources. In the future, our approach will need to be expanded and refined, but this is just the initial step toward harnessing the Web and fulfilling the promise of "big data" for tobacco control.

FUNDING

DECLARATION OF INTERESTS

REFERENCES

Althouse, B. M., Ng, Y. Y., & Cummings, D. A. (2011). Prediction of dengue incidence using search query surveillance. *PLoS Neglected Tropical Diseases*, *5*, e1258.

Ayers, J. W., Althouse, B. M., Allem, J. P., Ford, D. E., Ribisl, K. M., & Cohen, J. E. (2012). A novel evaluation of world no tobacco day in Latin America. *Journal of Medical Internet Research*, *14*, e77.

Ayers, J. W., Althouse, B. M., Allem, J. P., Rosenquist, J. N., & Ford, D. E. (2013). Seasonality in seeking mental health information on google. *American Journal of Preventive Medicine*, *44*, 520–525.

Ayers, J. W., Althouse, B. M., Johnson, M., & Cohen, J. E. (2013). Weekly "circaseptan" rhythms in smoking cessation contemplations. *JAMA Internal Medicine*. Published online October 28, 2013. doi:10.1001/jamainternmed.2013.11933

Ayers, J. W., Ribisl, K., & Brownstein, J. S. (2011a). Using search query surveillance to monitor tax avoidance and smoking cessation following the United States' 2009 "SCHIP" cigarette tax increase. *PloS One*, *6*, e16777.

Ayers, J. W., Ribisl, K. M., & Brownstein, J. S. (2011b). Tracking the rise in popularity of electronic nicotine delivery systems (electronic cigarettes) using search query surveillance. *American Journal of Preventive Medicine*, *40*, 448–453.

Backinger, C. L., & Augustson, E. M. (2011). Where there's an app, there's a way? *American Journal of Preventive Medicine*, *40*, 390–391.

Boland, M., Sweeney, M. R., Scallan, E., Harrington, M., & Staines, A. (2006). Emerging advantages and drawbacks of telephone surveying in public health research in Ireland and the U.K. *BMC Public Health*, *6*, 208.

Brownson, R. C., Fielding, J. E., & Maylahn, C. M. (2009). Evidence-based public health: A fundamental concept for public health practice. *Annual Review of Public Health*, *30*, 175–201.

Centers for Disease Control. (2011). Research priorities identified by the office on smoking and health—fall 2011. Unpublished letter.

Chaloupka, F. J., Yurekli, A., & Fong, G. T. (2012). Tobacco taxes as a tobacco control strategy. *Tobacco Control*, *21*, 172–180.

Chan, M., Kazatchkine, M., Lob-Levyt, J., Obaid, T., Schweizer, J., Sidibe, M., & Yamada, T. (2010). Meeting the demand for results and accountability: A call for action on health data from eight global health agencies. *PLoS Medicine*, *7*, e1000223.

Choi, K., Hennrikus, D., Forster, J., & St Claire, A. W. (2012). Use of price-minimizing strategies by smokers and their effects on subsequent smoking behaviors. *Nicotine & Tobacco Research*, *14*, 864–870.

Cobb, N. K. (2010). Online consumer search strategies for smoking-cessation information. *American Journal of Preventive Medicine*, *38*(3 Suppl.), S429–S432.

Cotten, S. R., & Gupta, S. S. (2004). Characteristics of online and offline health information seekers and factors that discriminate between them. *Social Science & Medicine (1982)*, *59*, 1795–1806.

Curtin, R., Presser, S., & Singer, E. (2005). Changes in telephone survey nonresponse over the past quarter century. *Public Opinion Quarterly*, *69*, 87–98.

Dutka, A. F., & Hanson, H. H. (1989). *Fundamentals of data normalization*. Reading, MA: Addison-Wesley Pub. Co.

Eysenbach, G. (2011). Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. *American Journal of Preventive Medicine*, *40*(5 Suppl. 2), S154–S158.

Freeman, B. (2012). New media and tobacco control. *Tobacco Control*, *21*, 139–144. doi:10.1136/tobaccocontrol-2011-050193

Giovino, G. A., Biener, L., Hartman, A. M., Marcus, S. E., Schooley, M. W., Pechacek, T. F., & Vallone, D. (2009). Monitoring the tobacco use epidemic I. Overview: Optimizing measurement to facilitate change. *Preventive Medicine*, *48*(1 Suppl.), S4–S10.

Graham, A. L., Cobb, N. K., Papandonatos, G. D., Moreno, J. L., Kang, H., Tinkelman, D. G., & Abrams, D. B. (2011). A randomized trial of internet and telephone treatment for smoking cessation. *Archives of Internal Medicine*, *171*, 46–53.

Groves, R. M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*, 646–675.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. New York, NY: Springer.

King, G. (2011). Ensuring the data-rich future of the social sciences. *Science*, *331*, 719–721.

King, G., Tomz, M., & Wittenberg, J. (2000). Making the most of statistical analyses: Improving interpretation and presentation. *American Journal of Political Science*, *44*, 341–355.

Leischow, S. J., Zeller, M., & Backinger, C. L. (2012). Research priorities and infrastructure needs of the family smoking prevention and tobacco control act: Science to inform FDA policy. *Nicotine & Tobacco Research*, *14*, 1–6.

Lewis-Beck, M. S. (1986). Interrupted time series. In W. Berry & M. Lewis-Beck (Eds.) *New tools for social Scientists* (pp. 209–240). Beverly Hills, CA: Sage Publications.

McMullan, M. (2006). Patients using the internet to obtain health information: How this affects the patient-health professional relationship. *Patient Education and Counseling*, *63*, 24–28.

Myung, S. K., McDonnell, D. D., Kazinets, G., Seo, H. G., & Moskowitz, J. M. (2009). Effects of web- and computer-based smoking cessation programs: Meta-analysis of randomized controlled trials. *Archives of Internal Medicine*, *169*, 929–937.

Paul, M. J., & Dredze, M. (2011). You are what you tweet: Analyzing Twitter for public health. *Fifth International AAAI Conference on Weblogs and Social Media (ICWSM 2011)*, Barcelona, Spain .

Pierce, J. P., Messer, K., White, M. M., Cowling, D. W., & Thomas, D. P. (2011). Prevalence of heavy smoking in California and the United States, 1965–2007. *Journal of the American Medical Association*, *305*, 1106–1112.

Prochaska, J. J., Pechmann, C., Kim, R., & Leonhardt, J. M. (2012). Twitter=quitter? An analysis of twitter quit smoking social networks. *Tobacco Control*, *21*, 447–449.

Ribisl, K. M., & Jo, C. (2012). Tobacco control is losing ground in the web 2.0 era: Invited commentary. *Tobacco Control*, *21*, 145–146.

Samuel, K. A., Ribisl, K. M., & Williams, R. S. (2012). Internet cigarette sales and Native American sovereignty: Political and public health contexts. *Journal of Public Health Policy*, *33*, 173–187.

Shadish, W. R., & Cook, T. D. (2009). The renaissance of field experimentation in evaluating interventions. *Annual Review of Psychology*, *60*, 607–629.

Shah, N. H., & Tenenbaum, J. D. (2012). The coming age of data-driven medicine: Translational bioinformatics' next frontier. *Journal of the American Medical Informatics Association*, *19*, e2–e4.

Stoddard, J. L., & Augustson, E. M. (2006). Smokers who use internet and smokers who don't: Data from the health information and national trends survey (HINTS). *Nicotine & Tobacco Research*, *8*(Suppl. 1), S77–S85.

U.S. Department of Health and Human Services. (n.d.). *Ending the tobacco epidemic: A tobacco control strategic action plan for the U.S. Department of health and human services*. Washington, DC: Office of the Assistant Secretary for Health. November 2010.

Yamin, C. K., Bitton, A., & Bates, D. W. (2010). E-cigarettes: A rapidly growing internet phenomenon. *Annals of Internal Medicine*, *153*, 607–609

Ybarra, M., & Suman, M. (2008). Reasons, assessments and actions taken: Sex and age differences in uses of internet health information. *Health Education Research*, *23*, 512–521.