

Original Paper

# Leveraging Big Data to Improve Health Awareness Campaigns: A Novel Evaluation of the Great American Smokeout

John W Ayers<sup>1</sup>, PhD, MA; J Lee Westmaas<sup>2</sup>, PhD; Eric C Leas<sup>3</sup>, MPH; Adrian Benton<sup>4</sup>, MS; Yunqi Chen<sup>5</sup>, BA; Mark Dredze<sup>4</sup>, PhD; Benjamin M Althouse<sup>6,7</sup>, PhD, ScM

<sup>1</sup>Graduate School of Public Health, San Diego State University, Chula Vista, CA, United States

<sup>2</sup>American Cancer Society, Behavioral Research Center, Atlanta, GA, United States

<sup>3</sup>University of California San Diego School of Medicine, La Jolla, CA, United States

<sup>4</sup>Human Language Technology Center of Excellence, Johns Hopkins University, Baltimore, MD, United States

<sup>5</sup>Bryn Mawr College, Bryn Mawr College, Philadelphia, PA, United States

<sup>6</sup>Santa Fe Institute, Santa Fe, NM, United States

<sup>7</sup>New Mexico State University, Las Cruces, NM, United States

**Corresponding Author:**

John W Ayers, PhD, MA

Graduate School of Public Health, San Diego State University

2967 Four Corners St

Chula Vista, CA, 91914-5210

United States

Phone: 1 6193711846

Fax: 1 6193711846

Email: [ayers.john.w@gmail.com](mailto:ayers.john.w@gmail.com)

## Abstract

**Background:** Awareness campaigns are ubiquitous, but little is known about their potential effectiveness because traditional evaluations are often unfeasible. For 40 years, the “Great American Smokeout” (GASO) has encouraged media coverage and popular engagement with smoking cessation on the third Thursday of November as the nation’s longest running awareness campaign.

**Objective:** We proposed a novel evaluation framework for assessing awareness campaigns using the GASO as a case study by observing cessation-related news reports and Twitter postings, and cessation-related help seeking via Google, Wikipedia, and government-sponsored quitlines.

**Methods:** Time trends (2009-2014) were analyzed using a quasi-experimental design to isolate spikes during the GASO by comparing observed outcomes on the GASO day with the simulated counterfactual had the GASO not occurred.

**Results:** Cessation-related news typically increased by 61% (95% CI 35-87) and tweets by 13% (95% CI -21 to 48) during the GASO compared with what was expected had the GASO not occurred. Cessation-related Google searches increased by 25% (95% CI 10-40), Wikipedia page visits by 22% (95% CI -26 to 67), and quitline calls by 42% (95% CI 19-64). Cessation-related news media positively coincided with cessation tweets, Internet searches, and Wikipedia visits; for example, a 50% increase in news for any year predicted a 28% (95% CI -2 to 59) increase in tweets for the same year. Increases on the day of the GASO rivaled about two-thirds of a typical New Year’s Day—the day that is assumed to see the greatest increases in cessation-related activity. In practical terms, there were about 61,000 more instances of help seeking on Google, Wikipedia, or quitlines on GASO each year than would normally be expected.

**Conclusions:** These findings provide actionable intelligence to improve the GASO and model how to rapidly, cost-effectively, and efficiently evaluate hundreds of awareness campaigns, nearly all for the first time.

(*JMIR Public Health Surveill* 2016;2(1):e16) doi: [10.2196/publichealth.5304](https://doi.org/10.2196/publichealth.5304)

**KEYWORDS**

big data; evaluation; health communication; mass media; social media; tobacco control; infodemiology; infoveillance; twitter; smoking cessation

## Introduction

Public health awareness campaigns are annual events designed to foster knowledge of health risks and engagement during specific days, weeks, or months. The number of public awareness campaigns has substantially risen, with one catalogue listing more than 200 ongoing domestic campaigns [1].

The ubiquity of awareness campaigns likely stems from the ease of their implementation. While most mass media campaigns require substantial advertising budgets [2], awareness campaigns require minimal assets. For example, many awareness campaigns are accomplished by issuing a press release that is freely carried across news media outlets and potentially on social media.

The ease of implementing awareness campaigns, however, sharply contrasts with the pragmatic and methodological barriers that have made assessing their potential effectiveness unfeasible. As an example, a recent systematic review found that only 5 of 79 published reports on awareness campaigns were evaluation studies [3]. One reason for the absence of evaluations is that the driving incentive for awareness campaigns is to minimize operational expenses, which can make traditional evaluations too costly [4]. In addition, because most awareness campaigns occur on a single day, traditional evaluations may not be possible, such as collecting survey responses for a pre-post analysis. Regardless, traditional modalities alone cannot altogether inform how awareness campaigns are expected to work, as they often cannot assess how campaign messages are diffused via news media, are shared on social media, or how the population engages with these messages (such as seeking out information online). Alternative evaluation approaches that rely on metrics assessing both media and popular reactions to media are needed to ensure that awareness campaigns' important public health messages are scientifically supported.

Big data may fill the knowledge gap for evaluating health awareness campaigns. It has to be noted that other public health surveillance gaps are already being filled by big data's inexpensive, fine-grained, and real-time capabilities, especially in cases where little or no traditional data exist [5-7]. However, to date, big data have rarely been applied to studying health communication campaigns [8-14], or even health communication research more broadly [15-18]. The few studies harnessing big data have often relied on a single outcome, such as Internet search queries, and loosely interpreted these outcomes in the absence of any framework for campaign evaluations. We take a step forward by proposing a simple and actionable approach to monitoring trends in the most frequently mentioned kinds of big data (news media, social media, and help seeking on Google, Wikipedia, and telephone counseling services). Our analyses are presented in the context of a novel evaluation framework that takes into account how awareness campaigns are expected to diffuse [3], and thus provide actionable intelligence for program managers at each stage of campaign diffusion from news media at the population level to individual help seeking regarding campaign outcomes.

In this report, we specifically developed a framework to evaluate the *American Cancer Society's* "Great American Smokeout" (GASO). This past year (2015) marked the 40th commemoration

of the GASO, held annually on the third Thursday of November, making it the world's longest running health awareness campaign. The GASO aims to generate and diffuse media encouraging individuals to go smoke-free for a day, and to consider quitting for good. Although activities promoting the GASO have varied over time, GASO is currently promoted by scheduling interviews with selected news outlets; producing media fact sheets with information about tobacco harms and quitting tips; disseminating toolkits with infographics, email templates, and event ideas such as quitting competitions for partnering organizations or businesses to celebrate the GASO; and more recently, by posting similar messages on social media.

Our aim was to develop a method for studying how a message diffuses through news and social media and ultimately leads to engagement via information seeking with the core message. Our novel evaluation framework was based on the key intended aims of the GASO and available big data sources. First, we examined the extent to which the GASO's cessation-related messages for 2009-2014 were carried in news media; second, how similar messages were shared on Twitter; and third, the extent to which these messages were associated with help seeking via Internet searches for quitting smoking, information retrieval on Wikipedia's cessation page, and calling state-sponsored quitlines, all informed by an analysis of 2191 days of data. We hypothesized that trends would spike on the day of the GASO. Moreover, the interrelationships between these data (eg, news media's relationship with social media) were explored, under the assumption that news media begets social media, progressing through individual help-seeking actions for cessation.

## Methods

Data were compiled into a daily time series for the period between January 1, 2009, and December 31, 2014. Each data source was purposively selected to provide insights on a different level of analysis or specificity. For example, news media trends indicate how well GASO's message of quitting smoking is disseminated, whereas search query trends indicate that the searcher is thinking about cessation or that the searcher is taking some action toward investigating behavior change. Table 1 presents the diverse data sources, as briefly described in the following section.

Cessation-related news media coverage that mentioned "quit" or "stop" and "smoking" was retrieved from News Library, a comprehensive repository of US newspaper articles, including online and print content. This included stories with titles such as "The last cigarette: Nine ex-smokers who quit the habit for good," "Quit smoking, one of the most important ways to improve health," and "Quitting smoking reduces heart risk faster than previously thought." Stories mentioning cessation were normalized relative to all news stories to adjust for potential biases arising from changes in media volume over time.

Social media shares about cessation were obtained from the public Twitter application programming interface. This did not include any data that had been marked as "private" by the author or any direct messages. Initially, the data were derived from 2 feeds, each a random sample of 10% of all tweets. We retained

600,000 tweets mentioning “quit” or “stop” and “smoking” (including variations on “quit” like “quitting”) while not mentioning “fire,” “marijuana,” “mj,” “pot,” “pott,” or “weed.” Subsequently, we randomly selected 10,000 of these tweets and a single reader labeled them as to whether the tweeter was indeed tweeting about smoking cessation. We then trained an automated classifier on that data to mimic human coding, obtaining about 90% precision and 70% recall. The resulting classifier was ran on all 600,000 tweets to obtain the analyzed trend. This included tweets such as “tryn quit smoking waiste to much money” (sic) or “quit smoking and start exercising because healthy body healthy mind” (sic). Tweets were measured in raw volume, but trending (increases in tweets over

time) was corrected for in the analysis as detailed in the following section.

Internet searches for cessation information were obtained from *Google Trends*, a public archive of Google search queries. Queries that included “quit” or “stop” and “smoking” (such as “quit smoking methods,” “help stop smoking,” or “quit smoking”) after excluding searches that also included “fire,” “marijuana,” “mj,” “pot,” “pott,” or “weed,” were monitored relative to all queries each day, reported as a relative search volume (RSV), where RSV=100 is the day with the highest search proportion, and RSV=50 is a day with 50% of that highest proportion.

**Table 1.** Big data codebook.

Type	Source	Description	Privacy	Aim
News coverage	News Library [19]: Available with a paid subscription, including papers since 1903.	Includes news articles from 5689 US newspapers. Described as “virtually all” US newspapers.	All publicly available data. Complete text not available for some papers behind pay walls.	To evaluate how frequently the primary GASO message was carried by news media.
Social media	Twitter [20]: Available prospectively with a paid subscription.	Includes either samples of all tweets or all tweets for specific keywords/phrases on the global Twitter platform.	Derived from public tweets (excluding direct messages or tweets marked “private”). Includes username and location for users sharing their location.	To evaluate how frequently cessation-related tweets were shared on social media (a secondary aim of the GASO).
Internet searches	Google [21]: Available to any user with a Google account.	Includes global search trends for investigator-selected keywords or phrases.	Derived from private search activity on Google. Disaggregated to city, national, or global units without any identifying information to protect privacy.	To evaluate how the GASO motivated some to engage in behavior change by seeking out additional information on cessation.
Information retrieval	Wikipedia [22]: Raw access logs are available here; those used in the study were aggregated by a third party.	Includes counts of all Wikipedia page visits. Aggregated by language, but unable to aggregate by location.	Derived from private online activity. Disaggregated to preserve privacy.	To evaluate how the GASO motivated specific information retrieval on the most popular cessation resource.
Quitline calls	Sourced privately, covering 29 US states. Additional states are only available from other providers or state representatives.	Includes call language, frequency, and duration to US state-sponsored quitlines.	Derived from privately obtained calls. Data were aggregated nationally to protect privacy.	To evaluate how the GASO motivated some smokers to engage in quit counseling.

Wikipedia visits to the primary English language “smoking cessation” entry were also counted. Wikipedia was selected because it is typically the first or second result on Internet search engines. Web page views for this Wikipedia page were normalized by dividing by all Wikipedia English language page views reported per 1,000,000 each day.

Quitline call records were obtained from aggregating daily call logs for 28 states whose quitline service is provided by Alere Wellbeing, Inc. Demographic details for each call are recorded (eg, language of the call), but here we focused on all English-language calls. The raw call counts were used because there was no suitable denominator for normalization; however, trending (and other problems normalization typically resolves) was corrected for in the analysis as detailed in the following section.

Our statistical approach was quasi-experimental, focusing on deriving a single treatment effect for any potential spike proximal to the GASO in the time series for each of the 5 data sources [18,23]. The treatment period was the day of the GASO

and the reference was based on the empirically derived counterfactual had the GASO not occurred. For the overall estimate we used autoregressive integrated moving average (ARIMA) models that included all the days in each time series with a regression coefficient for the days of GASO fit to maximize Akaike information criterion using the algorithm outlined by Hyndman and Khandakar [24]. For the individual years, we fit the same ARIMA models to the year prior to each GASO event. The counterfactual value for the GASO was imputed based on this model for each of the 5 analyzed time series. The ratio of the observed values (treatment) and derived values from the ARIMAs (counterfactual) was computed, thereby making the effect estimates comparable across each time series [eg, (Treatment – Counterfactual)/Counterfactual]. These models are robust to the most well-known biases. For instance, the models are adjusted for seasonal [25] and circaseptan (day of the week) [26,27] periodicities, and trending in the data, such as declines in smoking-cessation Internet search queries as a proportion of all queries due to shrinking smoking populations or how cessation tweets may be growing as

Twitter's user base grows. Confidence bounds ( $\alpha=.05$ ) around these estimates were estimated using 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 [28].

To examine how increases in each time series were related to increases in media, we fit a linear model with the percent increases on GASO for media as the predictor and the percent increases for tweets, search, Wikipedia, and quitline calls as the outcome. Coefficients of these models were used to estimate the percent increase in outcome for a given increase in media. Confidence intervals were calculated analogously to those above.

Last, we replicated the aforementioned models focusing on estimating treatment effects for New Year's Day, which typically sees the greatest spike in cessation-related activity. These estimates were used as a baseline to judge the practical impacts of the GASO relative to New Year's Day (eg,  $\Delta\text{NYD}$  divided by  $\Delta\text{GASO}$ ).

## Results

Three patterns were observed from visual inspection of the time series across outcomes (Figure 1). There was substantial variability across outcomes in the size of the spikes (if any) around the GASO. For example, the GASO concurred with large spikes in cessation-related news media in 2009 but it was not replicated in tweets. There was substantial year-to-year variability within outcomes in the size of the spikes around the GASO. For example, news media coverage of cessation appeared to be following a downward trend with potential spikes appearing smaller later in the study period. Last, it appeared that the timing of the GASO closely preceded or occurred during lows in many of the outcomes. For example, Google searches were at their annual low around the holiday season, just days after the GASO.

Statistical analysis indicated that across all years, smoking cessation news stories were 61% (95% CI 35-87) higher than would be expected if the GASO did not occur. The largest increases occurred earlier in the analyzed period and the smallest in 2013 (Figure 2). Tweets were 13% (95% CI -21 to 48) higher

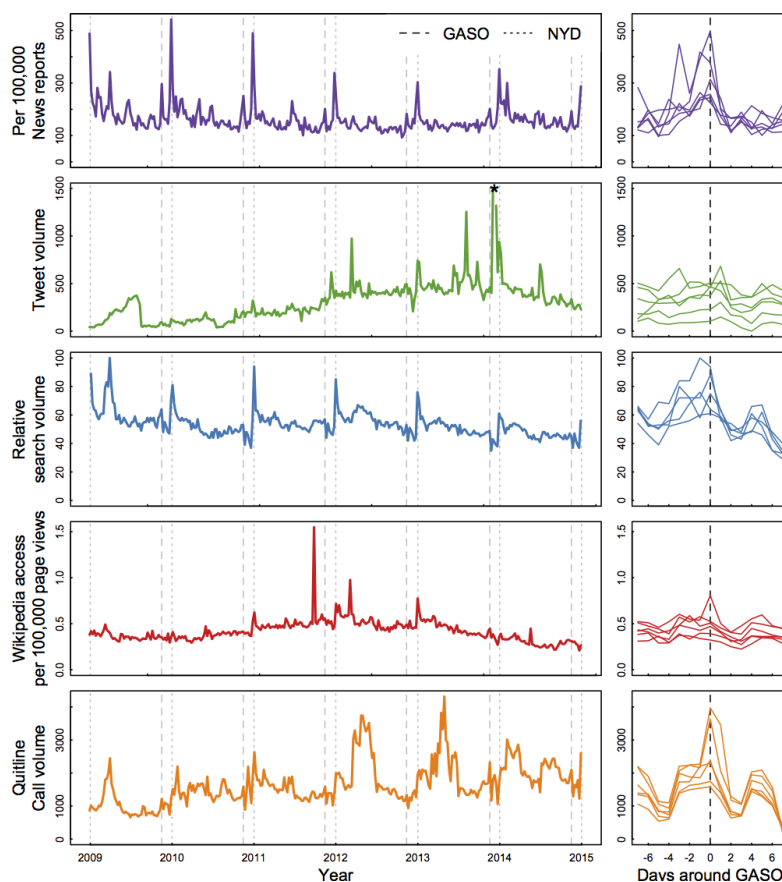
than would be expected, although these differences were also decreasing over the study period, ranging from an increase of 48% (95% CI 9-131) in 2011 to a nonsignificant 3% decrease (95% CI -478 to 47) in 2013. Analyses of the year-by-year relationship between news and social media coverage indicated that a 50% increase in news media was associated with an increase in tweets (28%; 95% CI -2 to 59;  $P<.06$ ).

Google searches for cessation information were 25% higher (95% CI 10-40) across all years, with a peak increase of 52% (95% CI 26-92) in 2011, when news media and Twitter interest also peaked. Wikipedia views increased by 22% (95% CI -26 to 67) across all years but the increase was not statistically significant overall or for any one year. Call volume to quitlines increased by 42% (95% CI 19-64) across all years, with the largest increases in 2011 (76%; 95% CI 12-308), 2013 (110%; 95% CI 32-412), and 2014 (92; 95% CI 24-322). In absolute terms over the entire 5 years, this amounts to about 42,600 more Google queries, 4600 more Wikipedia visits, and 13,400 more quitline calls than would be expected had the GASO not occurred.

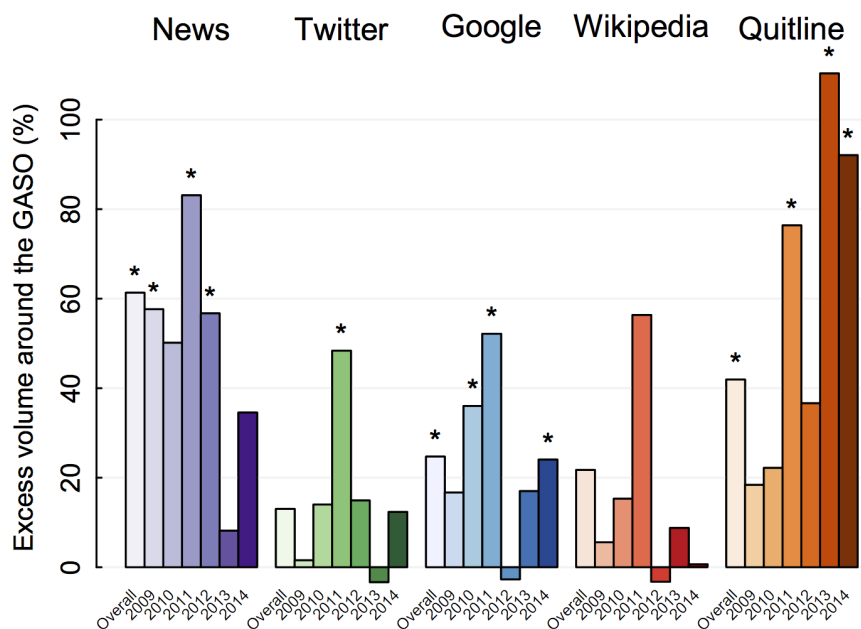
Increased cessation-related news coverage was positively associated with tweets and cessation-related help seeking on Wikipedia and Google (Figure 3). Specifically, a 50% increase in cessation-related news media predicted a 12% (95% CI -22 to 46) increase in Internet cessation queries and 26% (95% CI -22 to 73) increase for Wikipedia cessation page visits, albeit neither trend was statistically significant based on a small sample of 6 years. Quitline calls were inversely related to news media, a relationship largely driven mute by the 2 most recent years when quitline call increases were greatest and news volume was lowest.

To provide a comparison for the GASO, we fitted similar models for increases around New Year's Day, the only other period with visually significant spikes. Over all New Years (2009-2014), for example, news media coverage of cessation increased by 121% (95% CI 101-139) and Internet searches were 37% (95% CI 26-47) higher around New Year's Day. Thus, the GASO approximates 70% of the expected increase in news media and 68% of the expected increase in cessation searches for New Year's Day ( $\Delta\text{GASO}$  divided by  $\Delta\text{NYD}$ ).

**Figure 1.** Trends in all outcome measures (news volume, Tweets, Google searches, Wikipedia article views, and quitline call volume, top to bottom panels) across the entire study period (2009-2014). Gray lines indicate GASO (dashed lines) and New Years Day (NYD, dotted lines). The right panels display the same data focused around the days before and after the GASO for each year of the study period.

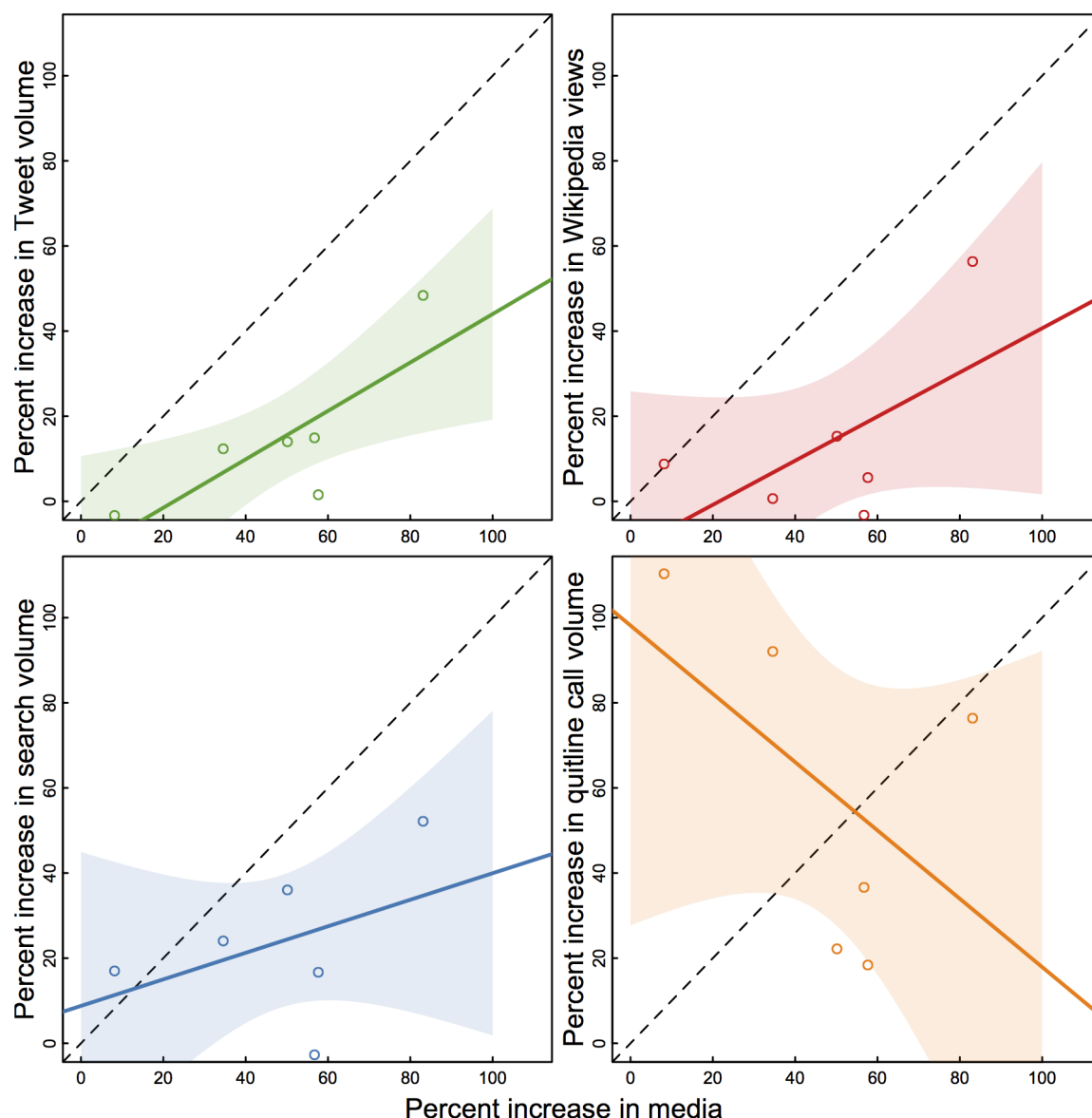


**Figure 2.** Percent increases in outcomes on the GASO compared to the counterfactual had the GASO not occurred, as detailed in the text. Asterisks indicate statistical significance ( $P < .05$ ).





**Figure 3.** Relationships between increases in cessation-related news media and help-seeking online (Google and Wikipedia) and on quitlines. Each circle captures the co-occurring annual mean effect estimate for the outcomes as displayed on the (y) and (x) axis, with a plotted line and confidence interval as derived from a linear regression. The slope of the dashed reference line indicates a perfect one-to-one relationship between the outcomes.



## Discussion

This study demonstrated a novel big data approach to rapidly, cost-effectively, and efficiently evaluate hundreds of awareness campaigns, nearly all for the first time. In doing so we found that the GASO often corresponded with increased cessation-related news coverage and cessation-related help seeking on Google, Wikipedia, and quitlines.

Our case study provided insights into how the GASO may have filtered through to the public and influenced cessation-related behavior in a way that has not been possible with traditional data. For example, we found that the GASO corresponded with increased cessation-related news coverage. Although this increase was not associated with more cessation-related tweets, it did correspond with about 61,000 more instances of cessation-related help seeking on Google, Wikipedia, and quitlines each year. These increases also compared favorably with New Year's Day, representing about two-thirds of the

effectiveness of New Year's Day for news media and Google searches. Most critically, these novel effectiveness estimates offer suggestions for how GASO can be expanded or modified to further enhance its impact.

Because media messages need to be timed to reach receptive audiences [29], one suggestion to better ensure GASO's cessation messages reach a more receptive audience would be to relocate these on the calendar. GASO is currently celebrated close to the holiday season when the lowest cessation-related help-seeking outcomes were observed. Add to this the recent discoveries about repeating seasonal [25] and circaseptan (day of the week) [26,27] patterns in cessation contemplations with the greatest interest during late winter and particularly on Monday, an empirically justified date of observing GASO, for example, would be the last Monday in February. With more smokers primed to consider cessation already, the GASO may very well realize larger increases apart from any specific change to its implementation.

Another potential strategy to improve effectiveness would be to attempt to increase news coverage, which, although positively related to changes in Google and Wikipedia help seeking for cessation, has been down most recently. Based on a review of prior press releases, it appears that GASO uses a repeating theme focused on smoking cessation alone. Changing themes each year to highlight novel content, such as rare health risks as with the Tips from Former Smokers Campaign [30], may engender more widespread media attention. For example, World No Tobacco Day changes themes every year and a recent analysis of that campaign showed more consistent year-to-year spikes in news coverage and correspondingly more consistent help seeking for smoking cessation on Google [12].

An enhanced strategic focus for using Twitter could also potentially increase GASO's effect on discussions about quitting. A reliance on news media to propagate tweets is not likely to be effective given the weak relationship between news coverage and tweets. Only recently has the *American Cancer Society* addressed the GASO's social media presence, which has included a handful of tweets announcing the GASO, and/or a few tweets with quitting tips leading up to and including the day of the GASO. Most tobacco control programs have been slow to adopt online social marketing strategies [31] and some that have engaged in these activities were quickly met with opposition or "Twitter Bombing" [32]. How best to engage awareness campaigns on social media websites, such as Twitter, continues to remain an important open question. With cessation-related tweets highest in 2011, GASO implementers might further investigate the context of that year and why tweets spiked.

The increase in online help seeking was promising, but these increases need to be concomitant with strategies to better ensure that those seeking help have an online pathway to evidence-based cessation assistance. Given the frequency with which smokers seek and find dubious cessation treatments online, such as laser therapies [33,34], a strategy to direct smokers to evidence-based cessation aids (such as information available at [smokefree.gov](http://smokefree.gov), [becomeanex.org](http://becomeanex.org), or [cancer.org](http://cancer.org)) may be beneficial for smokers attempting to quit [35]. This can be achieved in several ways such as purchasing advertisements on search engines for reliable cessation support or curating Wikipedia entries to provide reliable cessation support. For the

latter, at the time we studied the general cessation Wikipedia entry, it did not include any reference to or links to proven cessation techniques, but an effort by the *Society for Nicotine and Tobacco Research* is underway to revise and update relevant Wikipedia entries. Similarly, tailoring of quitline call scripts to target GASO-motivated callers may encourage cessation by highlighting how more smokers are choosing to quit on the GASO. This could potentially increase conversions from callers contemplating to making a quit attempt via social norm mechanisms [36].

A limitation of our analyses is that (even with multiple data sources) we could not observe all the pathways by which the GASO could motivate help seeking. For example, radio discussions about cessation might motivate purchases of nicotine replacement therapies at brick-and-mortar stores, both of which we did not observe. However, given that television and radio news strongly correlate with print news in other settings [37] and the Internet is by far the most popular vehicle for health seeking, we would not expect the overall interpretation of the campaign's effectiveness to change substantially. Still, we are actively working on identifying and adding additional data sources to our novel evaluation framework, such as Facebook [38,39]. In the same vein, our analysis focused on a single trend from each data source and did not describe subtle variations in the content of news, tweets, Google searches, etc. In subsequent evaluations, we plan to enhance our framework by diving deeper into these data, as we have done with searches for syndromic surveillance [40]. Last, the validity of our metrics is unknown because we were not able to make comparisons with surveys or other traditional data. Nonetheless, where overlap exists, our novel data sources often compare favorably with more traditional data. For example, Google searches mirror both disease outbreaks [41,42] and behavioral outcomes [43], including tobacco-related behaviors [44,45].

This study serves as an example framework for how to leverage big data for novel evaluations of awareness campaigns. As a result, the goals of awareness campaigns, like GASO's vital aim to encourage quitting, can be better realized based on scientific data. Moreover, if GASO and related awareness campaigns are modified or enhanced based on the interpretation of our results, they can be swiftly reassessed using the same low-cost, fine-grained, and real-time big data framework.

## Acknowledgments

We wish to thank Nick Generous for helping obtain the Wikipedia access log data, and providing expert advice on the analysis of these data. We also thank Omar Kordahi and Mary Kokstis at Alere Wellbeing, Inc. for providing aggregated quitline call volume data. We thank Keith Schnakenberg for helpful advice on data modeling. Some of the research reported in this publication for Google searches was informed by work supported by RCA173299A from the National Cancer Institute and FDA Center for Tobacco Products (CTP). The content is solely the responsibility of the authors and does not represent the official views of the American Cancer Society, National Institutes of Health, or the Food and Drug Administration. The funders had no role in the design, conduct, or interpretation of the study nor in the preparation, review, or approval of the manuscript.

## Conflicts of Interest

Drs Ayers and Althouse share an equity stake in a consultancy, Directing Medicine LLC, that advises clinician-scientists on how to implement some of the methods embodied in this work. Dr Dredze has been paid by Directing Medicine LLC in the past 5 years. Dr Ayers owns an equity position in HealthWatcher Inc. a data analytics firm specializing in media monitoring. Neither

the data nor the methods described in this article are proprietary. There are no other potentially relevant conflicts. The other authors have no conflicts of interest to declare.

## References

1. US Department of Health and Human Services. National Health Observances. 2014 at a Glance. 2014. URL: <http://healthfinder.gov/NHO/nhoyear.aspx?year=2014> [accessed 2016-02-24] [WebCite Cache ID 6fXMxkOhU]
2. Leas EC, Myers MG, Strong DR, Hofstetter CR, Al-Delaimy WK. Recall of anti-tobacco advertisements and effects on quitting behavior: results from the California smokers cohort. *Am J Public Health*. Feb 2015;105(2):e90-e97. [doi: [10.2105/AJPH.2014.302249](https://doi.org/10.2105/AJPH.2014.302249)] [Medline: [25521871](https://pubmed.ncbi.nlm.nih.gov/25521871/)]
3. Purtle J, Roman LA. Health awareness days: sufficient evidence to support the craze? *Am J Public Health*. Jun 2015;105(6):1061-1065. [doi: [10.2105/AJPH.2015.302621](https://doi.org/10.2105/AJPH.2015.302621)] [Medline: [25879148](https://pubmed.ncbi.nlm.nih.gov/25879148/)]
4. Wakefield MA, Loken B, Hornik RC. Use of mass media campaigns to change health behaviour. *Lancet*. Oct 9, 2010;376(9748):1261-1271. [FREE Full text] [doi: [10.1016/S0140-6736\(10\)60809-4](https://doi.org/10.1016/S0140-6736(10)60809-4)] [Medline: [20933263](https://pubmed.ncbi.nlm.nih.gov/20933263/)]
5. Ayers JW, Althouse BM, Dredze M. Could behavioral medicine lead the web data revolution? *JAMA*. Apr 9, 2014;311(14):1399-1400. [FREE Full text] [doi: [10.1001/jama.2014.1505](https://doi.org/10.1001/jama.2014.1505)] [Medline: [24577162](https://pubmed.ncbi.nlm.nih.gov/24577162/)]
6. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. *Am J Prev Med*. May 2011;40(5 Suppl 2):S154-S158. [doi: [10.1016/j.amepre.2011.02.006](https://doi.org/10.1016/j.amepre.2011.02.006)] [Medline: [21521589](https://pubmed.ncbi.nlm.nih.gov/21521589/)]
7. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the Internet. *J Med Internet Res*. 2009;11(1):e11. [FREE Full text] [doi: [10.2196/jmir.1157](https://doi.org/10.2196/jmir.1157)] [Medline: [19329408](https://pubmed.ncbi.nlm.nih.gov/19329408/)]
8. Cooper CP, Mallon KP, Leadbetter S, Pollack LA, Peipins LA. Cancer Internet search activity on a major search engine, United States 2001-2003. *J Med Internet Res*. Jul 1, 2005;7(3):e36. [FREE Full text] [doi: [10.2196/jmir.7.3.e36](https://doi.org/10.2196/jmir.7.3.e36)] [Medline: [15998627](https://pubmed.ncbi.nlm.nih.gov/15998627/)]
9. Owen L, Youdan B. 22 years on: the impact and relevance of the UK No Smoking Day. *Tob Control*. Feb 2006;15(1):19-25. [FREE Full text] [doi: [10.1136/tc.2005.011254](https://doi.org/10.1136/tc.2005.011254)] [Medline: [16436399](https://pubmed.ncbi.nlm.nih.gov/16436399/)]
10. Glynn RW, Kelly JC, Coffey N, Sweeney KJ, Kerin MJ. The effect of breast cancer awareness month on internet search activity--a comparison with awareness campaigns for lung and prostate cancer. *BMC Cancer*. 2011;11:442. [FREE Full text] [doi: [10.1186/1471-2407-11-442](https://doi.org/10.1186/1471-2407-11-442)] [Medline: [21993136](https://pubmed.ncbi.nlm.nih.gov/21993136/)]
11. Davis NF, Smyth LG, Flood HD. Detecting internet activity for erectile dysfunction using search engine query data in the Republic of Ireland. *BJU Int*. Dec 2012;110(11 Pt C):E939-E942. [FREE Full text] [doi: [10.1111/j.1464-410X.2012.11237.x](https://doi.org/10.1111/j.1464-410X.2012.11237.x)] [Medline: [22587462](https://pubmed.ncbi.nlm.nih.gov/22587462/)]
12. Ayers JW, Althouse BM, Allem JP, Ford DE, Ribisl KM, Cohen JE. A novel evaluation of World No Tobacco day in Latin America. *J Med Internet Res*. 2012;14(3):e77. [FREE Full text] [doi: [10.2196/jmir.2148](https://doi.org/10.2196/jmir.2148)] [Medline: [22634568](https://pubmed.ncbi.nlm.nih.gov/22634568/)]
13. Thackeray R, Burton SH, Giraud-Carrier C, Rollins S, Draper CR. Using Twitter for breast cancer prevention: an analysis of breast cancer awareness month. *BMC Cancer*. 2013;13:508. [FREE Full text] [doi: [10.1186/1471-2407-13-508](https://doi.org/10.1186/1471-2407-13-508)] [Medline: [24168075](https://pubmed.ncbi.nlm.nih.gov/24168075/)]
14. DeVilbiss E, Lee B. Brief report: trends in US National autism awareness from 2004 to 2014: the impact of national autism awareness month. *J Autism Dev Disord*. Dec 2014;44(12):3271-3273. [doi: [10.1007/s10803-014-2160-4](https://doi.org/10.1007/s10803-014-2160-4)] [Medline: [24915931](https://pubmed.ncbi.nlm.nih.gov/24915931/)]
15. Noar SM, Ribisl KM, Althouse BM, Willoughby JF, Ayers JW. Using digital surveillance to examine the impact of public figure pancreatic cancer announcements on media and search query outcomes. *J Natl Cancer Inst Monogr*. Dec 2013;2013(47):188-194. [doi: [10.1093/jncimonographs/igt017](https://doi.org/10.1093/jncimonographs/igt017)] [Medline: [24395990](https://pubmed.ncbi.nlm.nih.gov/24395990/)]
16. Ayers JW, Althouse BM, Noar SM, Cohen JE. Do celebrity cancer diagnoses promote primary cancer prevention? *Prev Med*. Jan 2014;58:81-84. [doi: [10.1016/j.ypmed.2013.11.007](https://doi.org/10.1016/j.ypmed.2013.11.007)] [Medline: [24252489](https://pubmed.ncbi.nlm.nih.gov/24252489/)]
17. Noar SM, Althouse BM, Ayers JW, Francis DB, Ribisl KM. Cancer information seeking in the digital age: effects of Angelina Jolie's prophylactic mastectomy announcement. *Med Decis Making*. Jan 2015;35(1):16-21. [doi: [10.1177/0272989X14556130](https://doi.org/10.1177/0272989X14556130)] [Medline: [25349187](https://pubmed.ncbi.nlm.nih.gov/25349187/)]
18. Ayers JW, Althouse BM, Emery S. Changes in Internet searches associated with the "Tips from Former Smokers" campaign. *Am J Prev Med*. Jun 2015;48(6):e27-e29. [doi: [10.1016/j.amepre.2015.03.015](https://doi.org/10.1016/j.amepre.2015.03.015)] [Medline: [25998929](https://pubmed.ncbi.nlm.nih.gov/25998929/)]
19. News Library. News Library [Home]. URL: <http://www.newslibrary.com> [accessed 2015-10-29] [WebCite Cache ID 6ce1Xt22n]
20. Twitter. Twitter [Home]. URL: <http://www.twitter.com> [accessed 2016-02-22] [WebCite Cache ID 6ce1UyUXC]
21. Google Trends. Google Trends [Home]. URL: <https://www.google.com/trends/> [accessed 2015-10-29] [WebCite Cache ID 6ce1ZUur8]
22. Wikipedia/Smoking Cessation. URL: [https://en.wikipedia.org/wiki/Smoking\\_cessation](https://en.wikipedia.org/wiki/Smoking_cessation) [accessed 2015-10-29] [WebCite Cache ID 6ce1brv8K]
23. Shadish WR, Cook TD. The renaissance of field experimentation in evaluating interventions. *Annu Rev Psychol*. 2009;60:607-629. [doi: [10.1146/annurev.psych.60.110707.163544](https://doi.org/10.1146/annurev.psych.60.110707.163544)] [Medline: [18616391](https://pubmed.ncbi.nlm.nih.gov/18616391/)]



24. Hyndman RJ, Khandakar Y. Automatic Time Series Forecasting: The Forecast Package for R 7. 2008. URL: <https://www.jstatsoft.org/article/view/v027i03> [accessed 2016-02-24] [WebCite Cache ID 6fXN6Gfuy]
25. Zhang Z, Zheng X, Zeng DD, Leischow SJ. Information seeking regarding tobacco and lung cancer: effects of seasonality. *PLoS One*. 2015;10(3):e0117938. [FREE Full text] [doi: [10.1371/journal.pone.0117938](https://doi.org/10.1371/journal.pone.0117938)] [Medline: [25781020](https://pubmed.ncbi.nlm.nih.gov/25781020/)]
26. Ayers JW, Althouse BM, Johnson M, Cohen JE. Circaseptan (weekly) rhythms in smoking cessation considerations. *JAMA Intern Med*. Jan 2014;174(1):146-148. [FREE Full text] [doi: [10.1001/jamainternmed.2013.11933](https://doi.org/10.1001/jamainternmed.2013.11933)] [Medline: [24166181](https://pubmed.ncbi.nlm.nih.gov/24166181/)]
27. Ayers JW, Althouse BM, Johnson M, Dredze M, Cohen JE. What's the healthiest day?: Circaseptan (weekly) rhythms in healthy considerations. *Am J Prev Med*. Jul 2014;47(1):73-76. [doi: [10.1016/j.amepre.2014.02.003](https://doi.org/10.1016/j.amepre.2014.02.003)] [Medline: [24746375](https://pubmed.ncbi.nlm.nih.gov/24746375/)]
28. King G, Tomz M, Wittenberg J. Making the most of statistical analyses: improving interpretation and presentation. *Am J Polit Sci*. 2000;44:341-355.
29. Bauer R. The obstinate audience: the influence process from the point of view of social communication. *Am Psychol*. 1964;19:319.
30. Rigotti NA, Wakefield M. Real people, real stories: a new mass media campaign that could help smokers quit. *Ann Intern Med*. Dec 18, 2012;157(12):907-909. [doi: [10.7326/0003-4819-156-1-2012101010-00541](https://doi.org/10.7326/0003-4819-156-1-2012101010-00541)] [Medline: [23007853](https://pubmed.ncbi.nlm.nih.gov/23007853/)]
31. Duke JC, Hansen H, Kim AE, Curry L, Allen J. The use of social media by state tobacco control programs to promote smoking cessation: a cross-sectional study. *J Med Internet Res*. 2014;16(7):e169. [FREE Full text] [doi: [10.2196/jmir.3430](https://doi.org/10.2196/jmir.3430)] [Medline: [25014311](https://pubmed.ncbi.nlm.nih.gov/25014311/)]
32. Ayers JW, Althouse BM, Ribisl KM, Emery S. Digital detection for tobacco control: online reactions to the 2009 U.S. cigarette excise tax increase. *Nicotine Tob Res*. May 2014;16(5):576-583. [FREE Full text] [doi: [10.1093/ntr/ntt186](https://doi.org/10.1093/ntr/ntt186)] [Medline: [24323570](https://pubmed.ncbi.nlm.nih.gov/24323570/)]
33. Cobb NK. Online consumer search strategies for smoking-cessation information. *Am J Prev Med*. Mar 2010;38(3 Suppl):S429-S432. [doi: [10.1016/j.amepre.2009.12.001](https://doi.org/10.1016/j.amepre.2009.12.001)] [Medline: [20176318](https://pubmed.ncbi.nlm.nih.gov/20176318/)]
34. Ayers J, Althouse B, Ribisl K, Emery S. Digital detection for tobacco control: online reactions to the 2009 U.S. cigarette excise tax increase. *Nicotine Tob Res*. May 2014;16(5):576-583. [FREE Full text] [doi: [10.1093/ntr/ntt186](https://doi.org/10.1093/ntr/ntt186)] [Medline: [24323570](https://pubmed.ncbi.nlm.nih.gov/24323570/)]
35. Westmaas JL, Abrams L, Bontemps-Jones J, Bauer JE, Bade J. Using the internet to understand smokers' treatment preferences: informing strategies to increase demand. *J Med Internet Res*. 2011;13(3):e58. [FREE Full text] [doi: [10.2196/jmir.1666](https://doi.org/10.2196/jmir.1666)] [Medline: [21873150](https://pubmed.ncbi.nlm.nih.gov/21873150/)]
36. Cialdini RB, Demaine LJ, Sagarin BJ, Barrett DW, Rhoads K, Winter PL. Managing social norms for persuasive impact. *Soc Influ*. 2006;1:3-15.
37. Cowling D, Modayil M, Stevens C. Assessing the relationship between ad volume and awareness of a tobacco education media campaign. *Tob Control*. Apr 2010;19 Suppl 1:i37-i42. [FREE Full text] [doi: [10.1136/tc.2009.030692](https://doi.org/10.1136/tc.2009.030692)] [Medline: [20382649](https://pubmed.ncbi.nlm.nih.gov/20382649/)]
38. Chunara R, Bouton L, Ayers JW, Brownstein JS. Assessing the online social environment for surveillance of obesity prevalence. *PLoS One*. 2013;8(4):e61373. [FREE Full text] [doi: [10.1371/journal.pone.0061373](https://doi.org/10.1371/journal.pone.0061373)] [Medline: [23637820](https://pubmed.ncbi.nlm.nih.gov/23637820/)]
39. Gittelman S, Lange V, Gotway Crawford Carol A, Okoro CA, Lieb E, Dhingra SS, et al. A new source of data for public health surveillance: Facebook likes. *J Med Internet Res*. 2015;17(4):e98. [FREE Full text] [doi: [10.2196/jmir.3970](https://doi.org/10.2196/jmir.3970)] [Medline: [25895907](https://pubmed.ncbi.nlm.nih.gov/25895907/)]
40. Althouse BM, Allem JP, Childers MA, Dredze M, Ayers JW. Population health concerns during the United States' Great Recession. *Am J Prev Med*. Feb 2014;46(2):166-170. [doi: [10.1016/j.amepre.2013.10.008](https://doi.org/10.1016/j.amepre.2013.10.008)] [Medline: [24439350](https://pubmed.ncbi.nlm.nih.gov/24439350/)]
41. Santillana M, Zhang DW, Althouse BM, Ayers JW. What can digital disease detection learn from (an external revision to) google flu trends? *Am J Prev Med*. Sep 2014;47(3):341-347. [doi: [10.1016/j.amepre.2014.05.020](https://doi.org/10.1016/j.amepre.2014.05.020)] [Medline: [24997572](https://pubmed.ncbi.nlm.nih.gov/24997572/)]
42. Eysenbach G. Infodemiology: tracking flu-related searches on the web for syndromic surveillance. *AMIA Annu Symp Proc*. 2006.:244-248. [FREE Full text] [Medline: [17238340](https://pubmed.ncbi.nlm.nih.gov/17238340/)]
43. Goel S, Hofman JM, Lahaie S, Pennock DM, Watts DJ. Predicting consumer behavior with Web search. *Proc Natl Acad Sci U S A*. Oct 12, 2010;107(41):17486-17490. [FREE Full text] [doi: [10.1073/pnas.1005962107](https://doi.org/10.1073/pnas.1005962107)] [Medline: [20876140](https://pubmed.ncbi.nlm.nih.gov/20876140/)]
44. Ayers J, Althouse B. "Tips From Former Smokers" can benefit from considering all available data: Reply to McAfee et al. *Am J Prev Med*. Dec 2015;49(6):e133-e134. [FREE Full text] [doi: [10.1016/j.amepre.2015.08.005](https://doi.org/10.1016/j.amepre.2015.08.005)] [Medline: [26456875](https://pubmed.ncbi.nlm.nih.gov/26456875/)]
45. Cavazos-Rehg PA, Krauss MJ, Spitznagel EL, Lowery A, Grucza RA, Chaloupka FJ, et al. Monitoring of non-cigarette tobacco use using Google Trends. *Tob Control*. May 2015;24(3):249-255. [FREE Full text] [doi: [10.1136/tobaccocontrol-2013-051276](https://doi.org/10.1136/tobaccocontrol-2013-051276)] [Medline: [24500269](https://pubmed.ncbi.nlm.nih.gov/24500269/)]

## Abbreviations

**ARIMA:** autoregressive integrated moving average  
**GASO:** Great American Smokeout  
**RSV:** relative search volume

*Edited by G Eysenbach; submitted 29.10.15; peer-reviewed by J Unger; comments to author 20.11.15; revised version received 17.01.16; accepted 19.01.16; published 31.03.16*

*Please cite as:*

Ayers JW, Westmaas JL, Leas EC, Benton A, Chen Y, Dredze M, Althouse BM

*Leveraging Big Data to Improve Health Awareness Campaigns: A Novel Evaluation of the Great American Smokeout*

*JMIR Public Health Surveill* 2016;2(1):e16

URL: <http://publichealth.jmir.org/2016/1/e16/>

doi: [10.2196/publichealth.5304](https://doi.org/10.2196/publichealth.5304)

PMID: [27227151](https://pubmed.ncbi.nlm.nih.gov/27227151/)

©John W Ayers, J Lee Westmaas, Eric C Leas, Adrian Benton, Yunqi Chen, Mark Dredze, Benjamin M Althouse. Originally published in JMIR Public Health and Surveillance (<http://publichealth.jmir.org>), 31.03.2016. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Public Health and Surveillance, is properly cited. The complete bibliographic information, a link to the original publication on <http://publichealth.jmir.org>, as well as this copyright and license information must be included.