American Journal of
Preventive Medicine

RESEARCH ARTICLE

Online Searching and Social Media to Detect Alcohol Use Risk at Population Scale

Elissa R. Weitzman, ScD, MSc,[1,2,3] Kara M. Magane, MS,[1] Po-Hua Chen, MD, MPH,[4] Hadi Amiri, PhD,[5] Timothy S. Naimi, MD, MPH,[6] Lauren E. Wisk, PhD[1,2,7]

Introduction: Harnessing engagement in online searching and social media may provide complementary information for monitoring alcohol use, informing prevention and policy evaluation, and extending knowledge available from national surveys.

Methods: Relative search volumes for 7 alcohol-related keywords were estimated from Google Trends (data, 2014–2017), and the proportion of alcohol use–related Twitter posts (data, 2014–2015) was estimated using natural language processing. Searching/posting measures were created for all 50 U.S. states plus Washington, D.C. Survey reports of alcohol use and summaries of state alcohol policies were obtained from the Behavioral Risk Factor Surveillance System (data, 2014–2016) and the Alcohol Policy Scale. In 2018–2019, associations among searching/posting measures and same state/year Behavioral Risk Factor Surveillance System reports of recent (past-30-day) alcohol use and maximum number of drinks consumed on an occasion were estimated using logistic and linear regression, adjusting for sociodemographics and Internet use, with moderation tested in regressions that included interactions of select searching/posting measures and the Alcohol Policy Scale.

Results: Recent alcohol use was reported by 52.93% of 1,297,168 Behavioral Risk Factor Surveillance System respondents, which was associated with all state-level searching/posting measures in unadjusted and adjusted models ($p<0.0001$). Among drinkers, most searching/posting measures were associated with maximum number of drinks consumed ($p<0.0001$). Associations varied with exposure to high versus low levels of state policy controls on alcohol.

Conclusions: Strong associations were found among individual alcohol use and state-level alcohol-related searching/posting measures, which were moderated by the strength of state alcohol policies. Findings support using novel personally generated data to monitor alcohol use and possibly evaluate effects of alcohol control policies.

Am J Prev Med 2020;58(1):79–88.

INTRODUCTION

Alcohol use is associated with leading causes of morbidity and mortality, including among youth.[1,2] Excessive drinking is associated with $249 billion annually in the U.S. (counting losses from workplace productivity and motor vehicle crashes and costs from healthcare treatment, law enforcement, and criminal justice); binge drinking—4 to 5 or more drinks on one occasion for female/male adults[3]—accounts for 77% of these costs.[4] The magnitude and costs of alcohol-related problems make it especially vital to identify trends and leverage points for preventive interventions.

From the [1]Division of Adolescent/Young Adult Medicine, Boston Children's Hospital, Boston, Massachusetts; [2]Department of Pediatrics, Harvard Medical School, Boston, Massachusetts; [3]Computational Health Informatics Program, Boston Children's Hospital, Boston, Massachusetts; [4]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, Massachusetts; [5]Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts; [6]Section of General Internal Medicine, Boston Medical Center, Boston, Massachusetts; and [7]Division of General Internal Medicine and Health Services Research, David Geffen School of Medicine at the University of California, Los Angeles, California

Address correspondence to: Elissa R. Weitzman, ScD, MSc, Division of Adolescent/Young Adult Medicine, Boston Children's Hospital, 300 Longwood Avenue, BCH 3187, Boston MA 02115. E-mail: elissa.weitzman@childrens.harvard.edu.

0749-3797/$36.00
https://doi.org/10.1016/j.amepre.2019.08.027

Existing surveys routinely monitor alcohol use behaviors throughout the U.S.,[5–7] providing prevalence estimates among school-based and household samples[8–10] and enabling characterization of disorders, comorbidities, and treatment patterns.[11,12] These invaluable resources are, however, resource intensive, and may be constrained to households with a landline telephone or availability/willingness to participate in an interview. They also may be less informative for detecting patterns and problems over short time periods or small geographies.

Monitoring approaches that harness population engagement in online searching and social media may offer complementary information about alcohol use. Termed "infodemiology" or "infoveillance,"[13] these approaches include tracking content entered into online search platforms and published through social media. Using them, researchers have been able to illuminate health attitudes and beliefs as well as disease incidence and spread, delivering timely insight into diverse phenomena, including sentiments toward health-protective behaviors (e.g., vaccination),[14,15] outbreaks of infectious disease (e.g., influenza),[16,17] and health problem awareness relevant to evaluating public health campaigns.[18,19]

Engagement with online searching and social media is ubiquitous,[20,21] and alcohol-related searching and posting patterns may elucidate alcohol use at national, state, or local levels at flexible time scales. Online searching about alcohol use has been shown to predict state-level changes in alcohol-induced death rates in the U.S.[22] Temporal patterns of social media posting about alcohol use are similar to known daily and holiday drinking patterns.[23] Source locations of Twitter posts about alcohol correspond to alcohol outlet density.[24] Plus, college student survey reports about alcohol use correspond to content on their social media profiles.[25] Promising findings do not yet confirm the correspondence among structured survey measures and alcohol-related searching and posting at a population level, nor do they shed light on whether the ambient policy environment moderates associations, which would support using these data for program and policy evaluation.

This study sought to test whether annual state-based measures of alcohol use derived from alcohol-related online searching and posting would be associated with individual-level alcohol use reported by same-state participants in a structured, U.S. survey composed of population-based surveys of all 50 states and Washington, D.C. Alignment between these measures would constitute important evidence for using web-based big data as complementary sources of information for monitoring alcohol use behaviors. Further, this study sought to ascertain whether associations among state-level searching/posting and individually reported measures of alcohol use would vary in the context of state alcohol policies. Strong policy controls on access to alcohol and beverage pricing and promotion are associated with reduced levels of heavy alcohol use and harms (e.g., accident and injury reports).[26–28] It is not known whether these patterns hold in analyses that employ searching/posting measures. Here, individual reports of alcohol use in a U.S. national survey were hypothesized to be associated with state-level measures of alcohol searching/posting, with associations additionally hypothesized to be sensitive to policies with attenuation in the setting of high policy controls.

METHODS

Study Sample

This study tested whether individual reports of alcohol use measured for subjects from the 50 U.S. states and Washington, D.C. surveyed in the Behavioral Risk Factor Surveillance System (BRFSS) were associated with state-level measures of online searching and social media posting about alcohol and used measures drawn from the Alcohol Policy Scale (APS)[29] to test whether associations would change in the context of variations in the state alcohol policy environment. Each surveyed individual in BRFSS for the 2014, 2015, and 2016 survey years was assigned their state's value for alcohol-related keywords or Twitter posting measures for the year in which they completed their survey (BRFSS participants from a given survey year who completed their survey in the subsequent calendar year were assigned relative search volumes [RSVs] from that subsequent year) and their state's APS value (consistent over time). This study was reviewed by the Boston Children's Hospital IRB and deemed exempt.

Measures

Search activity metrics for alcohol-related keywords were derived from the Google Trends (GT) application, an online, public, keyword research tool. GT computes the RSV for a specified term, defined as the number of searches for the specified term relative to the total number of searches. RSVs are displayed on a 0–100 scale, with 100 representing the highest RSV for that term. RSVs are normalized to account for areas with high Internet search activity and differences in the population of searches. RSVs can be calculated for a specified location (e.g., Alabama) and time (e.g., 2016).[30] Using GT, RSVs were generated for the following alcohol-related keywords for the 50 U.S. states and Washington, D.C.: *alcohol*, *alcoholic*, *alcoholism*, *drinking*, *beer*, *liquor*, and *wine*. Keywords were selected to encompass the major beverage categories of alcohol and a parsimonious set of action and problem descriptors.

A computer classifier was developed that identified first-person reports of alcohol use in Twitter posts (tweets). The classifier, which had an area under the curve performance of 92%,[31] was developed following a standard process in which a random sample of 10,000 tweets were manually labeled and used to train a classifier. The classifier was run on a Twitter data set consisting of 16 million geo-spatially tagged tweets posted from the contiguous U.S. for the January 2014–December 2015 period, which contained at least 1 of 600 alcohol-relevant keywords generated by expanding a previously used list[23] and drawing on terms in online alcohol dictionaries (e.g., www.onlineslangdictionary.com) and the word2vec toolkit (Mikolov et al., 2013. Distributed Representations of Words and Phrases and their Compositionality. In: Advances in Neural Information Processing Systems 26 [NIPS 2013]). This keyword-matched data set was drawn from the full Twitter firehose of all geo-spatially tagged tweets posted during the described period. After labeling all 16 million tweets using this classifier, state-specific summary measures were created to reflect the proportion of all tweets in a state that were identified as referring to first-person alcohol consumption by the classifier. The 2014 and 2015 BRFSS respondents were assigned the Twitter proportion from their state; 2016 BRFSS respondents were not assigned a Twitter proportion owing to the lack of temporal overlap between the 2 data sources.

Survey reports of alcohol use were obtained using data from the 2014, 2015, and 2016 BRFSS (including all 50 states and Washington, D.C.).[32] Alcohol use outcomes were past-30-day use (recent use) and maximum drinks on an occasion during the past 30 days (maximum drinks) among those who reported recent use. The BRFSS asked: *During the past 30 days, how many days per week or per month did you have at least one drink of any alcoholic beverage such as beer, wine, a malt beverage, or liquor?* Individuals who consumed alcohol on $\geq$1 day in the past 30 days classified as recent users and those reporting 0 days of alcohol consumption in the past 30 days as nonusers. The BRFSS also asked: *During the past 30 days, what is the largest number of drinks you had on any occasion?* Participants reported number of drinks (continuous) or *don't know/not sure* or *refused*. For recent use, participants who responded *don't know/not sure* (n=9,434) or *refused* (n=8,552) and participants who were missing responses (n=53,051; 5.2% of the total sample) were dropped from the analytic sample. For maximum drinks on an occasion, participants who responded *don't know/not sure* (n=17,077) or *refused* (n=3,772) or were missing responses (n=3,928) were recoded to the weighted median of each variable. Responses were Winsorized above the 99th percentile.[33−35] Sociodemographic variables included the following: sex, age (in years), race, ethnicity, urbanicity, education, marital status, past-30-day Internet use (any vs none), and BRFSS survey year.

Measures of U.S. state alcohol policy environments as of 2011 were based on APS scores.[29] The APS is a state by year−level measure of the relative restrictiveness of the alcohol policy environment, based on 29 alcohol policies developed based on expert nomination, efficacy, and implementation rating and prospectively validated against BRFSS data for people aged $\geq$18 years. Higher APS scores, representing more restrictive state policy environments, have been found to be associated with lower odds of binge drinking among adults and lower odds of drinking and binge drinking among underage youth.[27,29]

Statistical Analysis

Analyses were conducted in 2018−2019 using Stata, version 15. Statistical significance was considered at p<0.05. Analyses accounted for the complex sampling design of BRFSS and applied sampling weights. Summary statistics were computed to characterize the study sample overall and by recent use. Differences in sociodemographic characteristics, online searching/posting, and policy were compared by recent use using chi-square tests and t-tests, as appropriate. Univariate and multivariable logistic regression models were used to estimate individual odds of recent use among BRFSS respondents as independently predicted by state-level RSVs or Twitter metrics in a series of regression analyses where the association of each keyword RSV or Twitter metric was modeled separately, adjusting for individual sociodemographic factors and Internet use. Maximum number of drinks was modeled similarly using linear regression. For regression analyses, RSVs and Twitter metrics were standardized to represent the change in odds of any recent use, or the change in the β-coefficient for the maximum number of drinks consumed, at the individual level predicted by an SD increase in the level of searching or Twitter posting at the state level. A second set of regressions examined the interaction of selected online searching/posting measures with state alcohol policy scores; predicted margins and their associated 95% CIs were plotted to depict the interactions visually. Select RSVs illustrative of different dimensions of interest in alcohol use (i.e., the product *beer* and the problem *alcoholism*) were interacted with state alcohol policy scores, and plots are shown for these RSVs and Twitter to illustrate results and aid interpretation. A sensitivity analysis evaluated associations among BRFSS measures and state-level RSVs estimated at the month level (i.e., BRFSS respondents were assigned an RSV for their state that corresponded to their interview month and year; Appendix Table 1, available online).

RESULTS

Overall, of the final, weighted sample, 52.93% of BRFSS respondents were recent alcohol users; recent use was associated with all demographic characteristics (p<0.0001; Table 1). GT RSVs, Twitter prevalence, and the APS were also associated with recent use in bivariate analyses (p<0.0001).

In adjusted analyses, all RSVs and the Twitter measure were positively associated with recent use (for all, p<0.0001; Table 2), with the RSV for *beer* showing the largest effect and *liquor* showing the smallest effect; there was a 3%−17% adjusted increase in participants' odds of reporting recent use for an increase of 1 SD in RSV at the state level. There was a 6% increase in the odds of recent use associated with exposure to higher prevalence of state-level tweeting about alcohol use. Results were similar for estimates of these associations at the state/month level (Appendix Table 1, available online).

In adjusted analyses, *alcohol*, *drinking*, *alcoholic*, *alcoholism*, and *beer* were positively associated with maximum number of drinks on an occasion (all RSVs except for liquor p<0.0001; Twitter, p=0.97). The RSV for *wine* was negatively associated with maximum number of drinks on an occasion (p<0.001). Results were similar for estimates of these associations at the state/month level (Appendix Table 1, available online).

The magnitude of associations among state-level RSVs and Twitter metrics and individual survey reports of alcohol use varied by levels of alcohol control policies. For example, whereas higher RSVs were associated with

Table 1. Associations Among Sociodemographic Characteristics of BRFSS Respondents' Online Searching/Posting Measures by BRFSS Alcohol Use

	Recent alcohol use[a]			
Variable	**Total**	**Yes**	**No**	***p*-value**[b]
Total N (unweighted)	1,297,168	662,080	635,088	
Total % (weighted)		52.93	47.07	
Sociodemographics				
Sex, %				**<0.0001**
Female	51.46	46.93	53.07	
Male	48.54	59.29	40.71	
Age, years,[c] mean (SD)	47.18 (17.89)	45.50 (16.63)	49.06 (19.10)	**<0.0001**
Race, %				**<0.0001**
White only	73.97	55.54	44.46	
Black or African American	12.15	46.09	53.91	
American Indian or Alaskan native	1.67	43.95	56.05	
Asian	4.91	43.22	56.78	
Native Hawaiian or other Pacific Islander	0.31	49.34	50.66	
Other race only	2.48	45.88	54.12	
Multiracial	1.66	52.36	47.64	
Unknown	2.85	43.18	56.82	
Ethnicity, %				**<0.0001**
Hispanic	14.87	45.67	54.33	
Non-Hispanic	85.13	54.20	45.80	
Urbanicity, %				**<0.0001**
Urban	16.69	49.70	50.30	
Periurban	9.83	52.45	47.55	
Suburban	5.27	49.84	50.16	
Rural	7.39	40.42	59.58	
Unknown	60.82	55.68	44.32	
Education, %				**<0.0001**
Never attended school or only kindergarten	0.28	32.18	67.82	
Grades 1–8	4.41	29.18	70.82	
Grades 9–11	9.11	34.77	65.23	
Grade 12 or GED	28.06	46.10	53.90	
College 1–3 years	31.21	55.95	44.05	
College ≥4 years	26.63	67.17	32.83	
Unknown	0.30	33.41	66.59	
Marital status, %				**<0.0001**
Married	50.88	54.80	45.20	
Divorced	10.81	51.03	48.97	
Widowed	6.86	35.32	64.68	
Separated	2.52	46.13	53.87	
Never married	23.85	54.49	45.51	
A member of an unmarried couple	4.53	59.80	40.20	
Unknown	0.56	43.62	56.38	
Internet use, past 30 days, %				**<0.0001**
Any	82.69	57.63	42.37	
None	17.31	30.50	69.50	
RSV (tabulated at the person-level from state-year estimates)				
Alcohol, mean (SD)	82.16 (6.23)	82.41 (6.20)	81.89 (6.25)	**<0.0001**
Drinking, mean (SD)	81.14 (6.30)	81.33 (6.27)	80.93 (6.31)	**<0.0001**
Alcoholic, mean (SD)	73.94 (7.49)	74.16 (7.52)	73.69 (7.45)	**<0.0001**

(*continued on next page*)

Table 1. Associations Among Sociodemographic Characteristics of BRFSS Respondents' Online Searching/Posting Measures by BRFSS Alcohol Use (*continued*)

	Recent alcohol use[a]			
Variable	**Total**	**Yes**	**No**	***p*-value[b]**
Alcoholism, mean (SD)	57.59 (9.77)	58.04 (9.77)	57.07 (9.74)	**<0.0001**
Beer, mean (SD)	68.90 (11.65)	69.68 (11.52)	68.02 (11.71)	**<0.000 1**
Liquor, mean (SD)	41.51 (12.23)	41.77 (12.29)	41.23 (12.12)	**<0.0001**
Wine, mean (SD)	67.41 (12.63)	68.16 (12.22)	66.57 (13.05)	**<0.0001**
Twitter, mean (SD)	46.86 (3.60)	46.98 (3.56)	46.73 (3.62)	**<0.0001**
Alcohol policy score, mean (SD)	43.92 (6.32)	43.72 (5.88)	44.14 (6.79)	**<0.0001**

[a]The percentages shown are weighted.
[b]All statistics calculated at the individual level.
[c]Age in BRFSS collapsed above 80 to meet the criteria for public release.
BRFSS, Behavioral Risk Factor Surveillance System; RSV, relative search volume.

Table 2. Associations Among State-Level Google RSV, Twitter Alcohol Posting Prevalence, and BRFSS Alcohol Use Reports

Variable	**Recent alcohol use OR (95% CI)**	**Maximum number of drinks on an occasion B (95% CI)**
Alcohol-related searching/posting measures, unadjusted models[a]		
Alcohol	1.10 (1.10, 1.11)	0.04 (0.03, 0.06)
Drinking	1.08 (1.07, 1.08)	0.10 (0.08, 0.11)
Alcoholic	1.08 (1.07, 1.08)	0.10 (0.08, 0.11)
Alcoholism	1.15 (1.14, 1.16)	0.03 (0.01, 0.05)
Beer	1.17 (1.16, 1.18)	0.03 (0.01, 0.05)
Liquor	1.06 (1.05, 1.07)	−0.04 (−0.05, −0.02)
Wine	1.14 (1.13, 1.15)	−0.13 (−0.15, −0.11)
Twitter	1.09 (1.08, 1.11)	0.01 (−0.02, 0.03)
Alcohol-related searching/posting measures, adjusted models[b]		
Alcohol	1.09 (1.08, 1.09)	0.04 (0.02, 0.05)
Drinking	1.08 (1.08, 1.09)	0.09 (0.08, 0.11)
Alcoholic	1.07 (1.06, 1.07)	0.08 (0.07, 0.10)
Alcoholism	1.12 (1.11, 1.13)	0.04 (0.02, 0.05)
Beer	1.17 (1.16, 1.18)	0.05 (0.04, 0.07)
Liquor	1.03 (1.02, 1.04)	−0.01 (−0.02, 0.01)
Wine	1.17 (1.16, 1.18)	−0.05 (−0.07, −0.04)
Twitter	1.06 (1.05, 1.07)	0.00 (−0.02, 0.02)

[a]Unadjusted models adjust only for individual RSV or Twitter and are standardized to reflect the odds of individual use per a 1-SD (at the state-level) increase in RSV or Twitter.
[b]Adjusted models adjust for individual RSV or Twitter, age, sex, race, ethnicity, education, marital status, urbanicity, Internet use, and survey year; and are standardized to reflect the odds of individual use per a 1-SD (at the state-level) increase in RSV or Twitter.
BRFSS, Behavioral Risk Factor Surveillance System; RSV, relative search volume.

greater probability of recent use, the strength of that association varied across values of the APS. As an illustration, in Figure 1a−c, for recent use, the strength of the association is greatest in states with higher policy controls/enforcement and possibly low ambient supply (i.e., there is greater dispersion among the curves at high versus low APS values). Additionally, RSVs had a stronger association with maximum number of drinks in states with less restrictive policies, as illustrated in Figure 2a−c.

DISCUSSION

Drinking behaviors reported by individuals in a nationally representative survey were associated with alcohol-related online searching/posting measures. Similar results were seen for a subset of RSVs and for the Twitter measure in relation to reports of the maximum number of drinks consumed on an occasion among recent drinkers. Analyses were robust across 2 approaches for creating state-level measures of alcohol searching/posting from personally

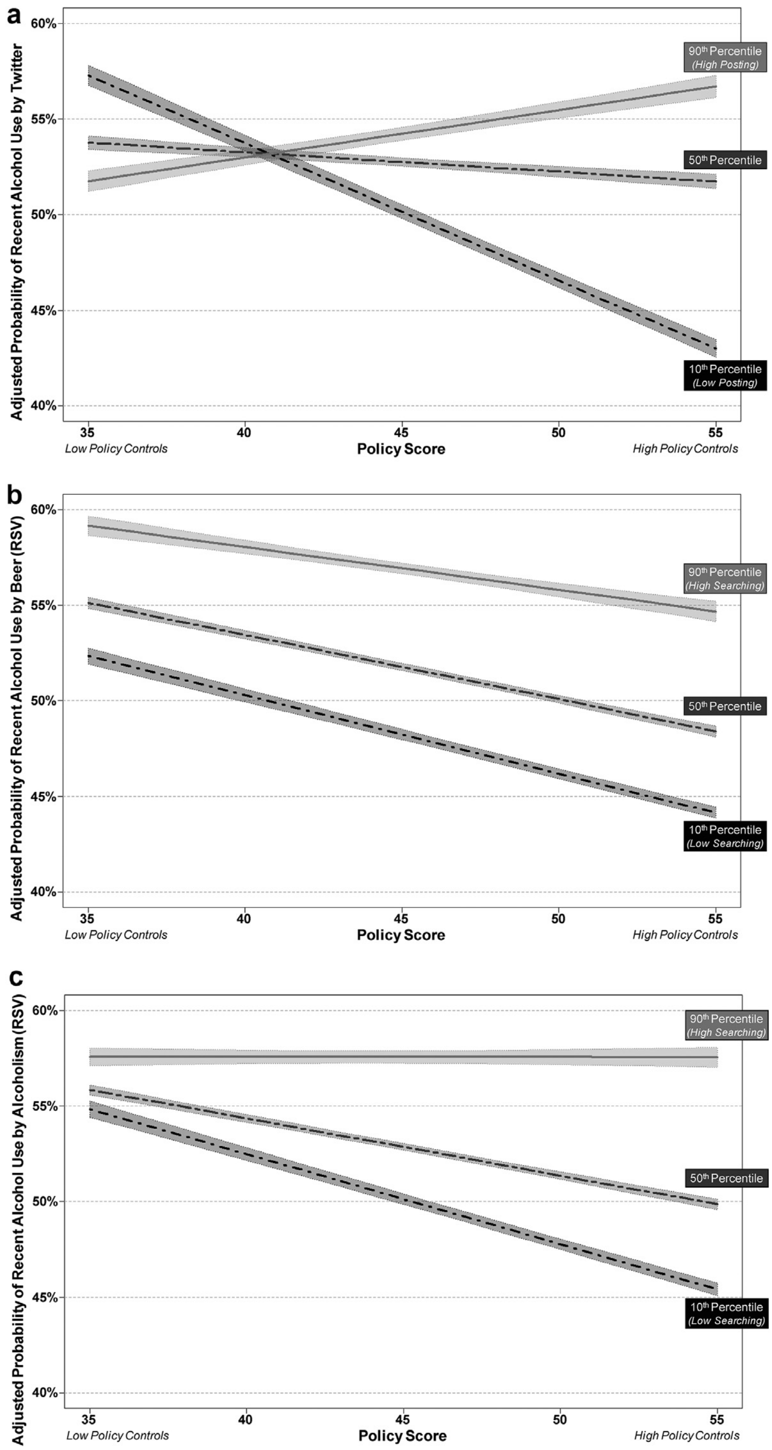

Figure 1. Adjusted predicted probability of recent alcohol use for Behavioral Risk Factor Surveillance System respondents at different levels of exposure to state Alcohol Policy Score (APS) and state posting/searching measures. (**a**) Predicted probabilities, with associated 95% confidence intervals (CIs), of recent alcohol use by APS and Twitter posting (10th, 50th, and 90th percentile values) output from covariate-adjusted logistic regression models. (**b**) Predicted probabilities, with associated 95% CIs, of recent alcohol use by APS and online searching about beer (10th, 50th, and 90th percentile values) output from covariate-adjusted logistic regression models. (**c**) Predicted probabilities, with associated 95% CIs, of recent alcohol use by APS and online searching about alcoholism (10th, 50th, and 90th percentile values) output from covariate-adjusted logistic regression models.
RSV, relative search volume.

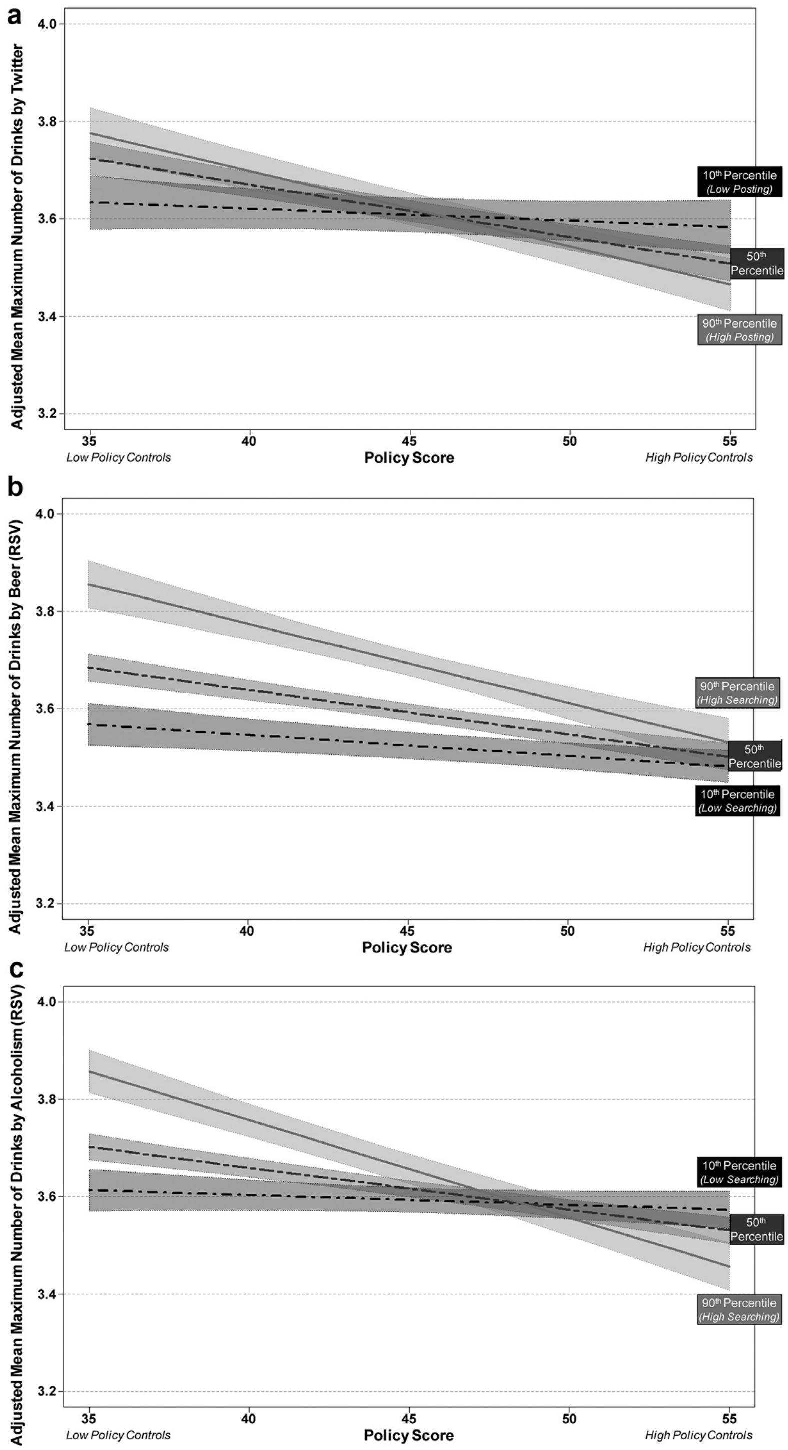

Figure 2. Adjusted predicted mean maximum number of drinks consumed on a single occasion for Behavioral Risk Factor Surveillance System respondents at different levels of exposure to state Alcohol Policy Score (APS) and state posting/searching measures. (**a**) Predicted means, with associated 95% confidence intervals (CIs), of the maximum number of drinks by APS and Twitter posting (10th, 50th, and 90th percentile values) output from covariate-adjusted linear regression models. (**b**) Predicted means, with associated 95% CIs, of the maximum number of drinks by APS and online searching about beer (10th, 50th, and 90th percentile values) output from covariate-adjusted linear regression models. (**c**) Predicted means, with associated 95% CIs, of the maximum number of drinks by APS and online searching about alcoholism (10th, 50th, and 90th percentile values) output from covariate-adjusted linear regression models.
RSV, relative search volume.

generated data gleaned from 2 platforms using multivariable regression to adjust for the effects of participant sociodemographic characteristics and past-30-day Internet use. Findings support the use of online search activity and social media data as complementary sources of information about alcohol use and constitute an important proof of concept for extension to the monitoring of alcohol use approaches to health surveillance developed to track sentiments and infectious disease.[36−39]

Alcohol use varies in relation to a range of individual and environmental factors.[40,41] Tracking trends in alcohol use is complex,[42] as population estimates are influenced by age, period, and cohort effects that encompass the cumulative impacts of demand- and supply-side contexts bearing on consumption.[43] This complexity challenges program and policy evaluation. Harnessing the volume and velocity of online data may aid efforts to identify and explain shifting patterns of use. A growing number of studies provide support for this enterprise, including studies of the temporal patterns in posting about alcohol on Twitter, which are higher on weekends than weekdays, during late night/early morning hours, and on holidays,[23,44] consistent with known patterns.[45,46] Further, surveys of community-residing and social media−engaged adolescents and young adults in Australia found strong positive associations among drinking risk and self-reported patterns of liking and following alcohol-related social media brands.[47,48] This study builds on these investigations and provides the first report of the extent to which individual drinking reported in a national survey is associated with state-level alcohol searching/posting derived from independently collected, personally generated electronic data. As alcohol use−related preventive interventions may be local or regional,[49,50] impacts of these efforts may be missed if their effects on alcohol use beliefs or behaviors are benchmarked against state or national data and on annual or longer timelines. Harnessing online searching/posting as complementary sources of information may be highly significant to evaluations that are focused on detecting patterns over brief time intervals, including in small areas, such as cities or counties, for which population-based surveys are not informative. Indeed, sensitivity analyses suggest that associations are similar though attenuated when modeled for exposure to monthly search activity (Appendix Table 1, available online).

Although individual patterns of alcohol use varied in relation to measures drawn from ambient patterns of alcohol-related online searching and posting (individual risk rises with increases in these measures), associations varied for different search terms and behaviors. For example, searching for *wine* was positively associated with recent use and negatively associated with maximum number of drinks consumed on an occasion among recent drinkers. Perhaps not surprisingly, this was not found for *beer*, for which high-volume consumption may be more typical. Associations also differ in the context of exposure to varying strengths of state alcohol policies. The approach overall is informative, including for surveillance of heavy alcohol use, and finds protective effects of strong alcohol policies on both outcomes, even where ambient measures of searching/posting are high. Findings are consistent with studies using national survey data where the state-level prevalence of binge drinking among adults predicted individual binge drinking among college students, with relationships moderated by strong state alcohol control policies.[51] This study provides additional evidence that alcohol control policies moderate consumption, including by constraining the volume of alcohol consumed even among participants from states with high levels of online searching/posting about alcohol. This is an especially noteworthy finding as heavy consumption, including high levels of per-occasion consumption as measured by the maximum number of drinks consumed, predicts harms,[52] making it an important target of prevention. Still, patterns are nuanced, as seen in the higher probability of recent use for individuals exposed to high levels of searching/posting measures and stronger policies in the illustrative cases. This pattern may reflect imposition of controls in response to heavy demand/use. Longitudinal work is needed to clarify these issues.

Limitations

Limitations apply. First, BRFSS is subject to recall and reporting biases; however, BRFSS measures of alcohol and substance use have been found to be moderately valid and reliable.[53] Second, a limited number of drinking outcomes were investigated. Future studies could include other measures, harms, or comorbidities. Third, personally generated searching/posting data have limitations around correctly inferring meaning from unstructured text, understanding the quality and transparency of algorithms, and accounting for the potential that differential Internet access and use may introduce bias and influence findings.[54] Steps to limit validity threats were taken by triangulating across multiple data sources and search terms, which strengthens the plausibility of results, even where insight is impeded into search term algorithms used to generate RSVs via a third party (Google), a known constraint.[55] Differential Internet access could introduce bias; however, the GT platform normalizes RSV estimates for this.[30] The natural language processing method used to generate the Twitter measure employed a high-performance classifier.[31] Per standard practice, this classifier was implemented on a very large keyword-enriched, though nonpopulation representative, data set. Fourth, both spatial and temporal overlap across all data sources (BRFSS, GT, and Twitter) was achieved for all spatial and all RSV analyses; temporal overlap was achieved for Twitter analyses for 2014 and 2015 only.

CONCLUSIONS

Strong associations were found among state-level searching/posting measures of alcohol use and individuals' alcohol consumption reported over 3 years of a national health survey. Findings were robust across 2 different approaches to operationalizing alcohol measures from personally generated data. Findings support using search activity and social media data to complement traditional public health monitoring of alcohol use and auger well for future efforts to extend traditional health surveillance systems toward use of online searching/posting to improve understanding of alcohol use and its prevention at population scale.

ACKNOWLEDGMENTS

The authors acknowledge Drs. John Brownstein and Jared Hawkins for contributing Twitter data, Mauricio Santillana for guidance on the collection and use of Google Trends data, and Guergana Savova for guidance on development of the classifier to detect alcohol use from Twitter posts.

The content is solely the responsibility of the authors and does not necessarily represent the official views of NIH.

Research reported in this report was supported by the National Institute on Alcohol Abuse and Alcoholism of NIH under award # R21AA023901.

ERW conceptualized and designed the study; supervised data collection; directed data analyses; and drafted, reviewed, and revised the manuscript. KMM assisted with data collection and analyses, and contributed to drafting, reviewing, and revising the manuscript. HA developed the natural language processing classifier and critically reviewed the manuscript. TSN contributed the Alcohol Policy Scale and critically reviewed the manuscript. P-HC collected data, conducted data analyses, and reviewed and revised the manuscript. LEW conducted data analyses and contributed to drafting, reviewing, and revising the manuscript.

No financial disclosures were reported by the authors of this paper.

SUPPLEMENTAL MATERIAL

Supplemental materials associated with this article can be found in the online version at https://doi.org/10.1016/j.amepre.2019.08.027.

REFERENCES

1. Eaton DK, Kann L, Kinchen S, et al. Youth risk behavior surveillance - United States, 2009. *MMWR Surveill Summ.* 2010;59(5):1–142. Accessed January 15, 2019.
2. GBD 2016 Alcohol Collaborators. Alcohol use and burden for 195 countries and territories, 1990−2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet.* 2018;392(10152):1015–1035. https://doi.org/10.1016/S0140-6736(18)31310-2.
3. NIAAA. Drinking levels defined. National Institute on Alcohol Abuse and Alcoholism. www.niaaa.nih.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking. Published 2016. Accessed December 28, 2018.
4. CDC. Excessive drinking is draining the US economy. www.cdc.gov/features/costsofdrinking/. Published 2018. Accessed December 28, 2018.
5. CDC. Behavioral Risk Factor Surveillance System. www.cdc.gov/brfss/. Accessed December 28, 2018.
6. CDC. Youth Risk Behavior Surveillance System (YRBSS). www.cdc.gov/healthyyouth/data/yrbs/. Accessed January 2, 2019.
7. Substance Abuse and Mental Health Services Administration. National Survey on Drug Use and Health (NSDUH). https://nsduhweb.rti.org/respweb/homepage.cfm. Accessed December 28, 2018.
8. Monitoring the Future. Purpose and design of MTF. www.monitoringthefuture.org/purpose.html. Accessed January 10, 2019.
9. Substance Abuse and Mental Health Administration. NSDUH - about the survey. https://nsduhweb.rti.org/respweb/about_nsduh.html. Published 2018. Accessed January 10, 2019.
10. CDC. About BRFSS. www.cdc.gov/brfss/about/index.htm. Published 2014. Accessed January 10, 2019.
11. Grant BF, Goldstein RB, Saha TD, et al. Epidemiology of DSM-5 alcohol use disorder results from the National Epidemiologic Survey on Alcohol and Related Conditions III. *JAMA Psychiatry.* 2015;72(8):757–766. https://doi.org/10.1001/jamapsychiatry.2015.0584.
12. Grant BF, Chou SP, Saha TD, et al. Prevalence of 12-month alcohol use, high-risk drinking, and DSM-IV alcohol use disorder in the United States, 2001−2002 to 2012−2013: results from the National Epidemiologic Survey on Alcohol and Related Conditions. *JAMA Psychiatry.* 2017;74(9):911–923. https://doi.org/10.1001/jamapsychiatry.2017.2161.
13. 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. https://doi.org/10.2196/jmir.1157.
14. Salathé M, Khandelwal S. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLoS Comput Biol.* 2011;7(10):e1002199. https://doi.org/10.1371/journal.pcbi.1002199.
15. Kang GJ, Ewing-Nelson SR, Mackey L, et al. Semantic network analysis of vaccine sentiment in online social media. *Vaccine.* 2017;35(29):3621–3638. https://doi.org/10.1016/j.vaccine.2017.05.052.
16. Marques-Toledo C A, Degener CM, Vinhal L, et al. Dengue prediction by the web: tweets are a useful tool for estimating and forecasting dengue at country and city level. *PLoS Negl Trop Dis.* 2017;11(7):e0005729. https://doi.org/10.1371/journal.pntd.0005729.
17. Brownstein JS, Chu S, Marathe A, et al. Combining participatory influenza surveillance with modeling and forecasting: three alternative approaches. *JMIR Public Health Surveill.* 2017;3(4):e83. https://doi.org/10.2196/publichealth.7344.
18. Hopkins ZH, Secrest AM. Public health implications of Google searches for sunscreen, sunburn, skin cancer, and melanoma in the United States. *Am J Health Promot.* 2019;33(4):611–615. https://doi.org/10.1177/0890117118811754.
19. Phillips CA, Barz Leahy A, Li Y, Schapira MM, Bailey LC, Merchant RM. Relationship between state-level Google online search volume and cancer incidence in the United States: retrospective study. *J Med Internet Res.* 2018;20(1):e6. https://doi.org/10.2196/jmir.8870.
20. Pew Research Center. Social media fact sheet. www.pewinternet.org/fact-sheet/social-media/. Published 2018. Accessed January 10, 2019.
21. Purcell K, Brenner J, Rainie L. *Search engine use 2012.* www.pewinternet.org/2012/03/09/search-engine-use-2012/. Published 2012. Accessed January 10, 2019.

22. Parker J, Cuthbertson C, Loveridge S, Skidmore M, Dyar W. Forecasting state-level premature deaths from alcohol, drugs, and suicides using Google Trends data. *J Affect Disord.* 2017;213:9–15. https://doi.org/10.1016/j.jad.2016.10.038.
23. Liu J, Weitzman ER, Chunara R. Assessing behavior stage progression from social media data. *CSCW Conf Comput Support Coop Work.* 2017;2017:1320–1333. https://doi.org/10.1145/2998181.2998336.
24. Hossain N, Hu T, Feizi R, White AM, Luo J, Kautz H. Inferring fine-grained details on user activities and home location from social media: detecting drinking-while-tweeting patterns in communities. In: Paper presented at: The Tenth International AAAI Conference on Web and Social Media (ICWSM 2016); May 17−20; 2016. https://arxiv.org/abs/1603.03181. Accessed September 5, 2019.
25. Moreno MA, Cox ED, Young HN, Haaland W. Underage college students' alcohol displays on Facebook and real-time alcohol behaviors. *J Adolesc Health.* 2015;56(6):646–651. https://doi.org/10.1016/j.jadohealth.2015.02.020.
26. Hadland SE, Xuan Z, Sarda V, et al. Alcohol policies and alcohol-related motor vehicle crash fatalities among young people in the US. *Pediatrics.* 2017;139(3):e20163037. https://doi.org/10.1542/peds.2016-3037.
27. Xuan Z, Blanchette JG, Nelson TF, et al. Youth drinking in the United States: relationships with alcohol policies and adult drinking. *Pediatrics.* 2015;136(1):18–27. https://doi.org/10.1542/peds.2015-0537.
28. Naimi TS, Xuan Z, Sarda V, et al. Association of state alcohol policies with alcohol-related motor vehicle crash fatalities among US adults. *JAMA Intern Med.* 2018;178(7):894–901. https://doi.org/10.1001/jamainternmed.2018.1406.
29. Naimi TS, Blanchette J, Nelson TF, et al. A new scale of the U.S. alcohol policy environment and its relationship to binge drinking. *Am J Prev Med.* 2014;46(1):10–16. https://doi.org/10.1016/j.amepre.2013.07.015.
30. Foroughi F, Lam AK-Y, Lim MSC, Saremi N, Ahmadvand A. "Googling" for cancer: an infodemiological assessment of online search interests in Australia, Canada, New Zealand, the United Kingdom, and the United States. *JMIR Cancer.* 2016;2(1):e5. https://doi.org/10.2196/cancer.5212.
31. Amiri H, Magane KM, Wisk LE, Savova GK, Weitzman ER. *Toward large-scale and multi-facet analysis of first person alcohol drinking.* https://scholar.harvard.edu/files/hadi/files/amiri_amia18.pdf. Published 2018. Accessed June 15, 2019.
32. CDC. 2016 BRFSS survey data and documentation. www.cdc.gov/brfss/annual_data/annual_2016.html. Accessed January 10, 2019.
33. Balsa AI, Giuliano LM, French MT. The effects of alcohol use on academic achievement in high school. *Econ Educ Rev.* 2011;30(1):1–15. https://doi.org/10.1016/j.econedurev.2010.06.015.
34. Dixon WJ. Simplified estimation from censored normal samples. *Ann Math Statist.* 1960;31(2):385–391. https://doi.org/10.1214/aoms/1177705900.
35. Siegel M, DeJong W, Naimi TS, et al. Brand-specific consumption of alcohol among underage youth in the United States. *Alcohol Clin Exp Res.* 2013;37(7):1195–1203. https://doi.org/10.1111/acer.12084.
36. Brownstein JS, Freifeld CC, Madoff LC. Digital disease detection — harnessing the web for public health surveillance. *N Engl J Med.* 2009;360(21):2153–2155. 2157. https://doi.org/10.1056/NEJMp0900702.
37. Lu FS, Hou S, Baltrusaitis K, et al. Accurate influenza monitoring and forecasting using novel internet data streams: a case study in the Boston metropolis. *J Med Internet Res.* 2018;4(1):e4. https://doi.org/10.2196/publichealth.8950.
38. Nagar R, Yuan Q, Freifeld CC, et al. A case study of the New York City 2012−2013 influenza season with daily geocoded Twitter data from temporal and spatiotemporal perspectives. *J Med Internet Res.* 2014;16(10):e236. https://doi.org/10.2196/jmir.3416.
39. Hswen Y, Sewalk KC, Alsentzer E, Tuli G, Brownstein JS, Hawkins JB. Investigating inequities in hospital care among lesbian, gay, bisexual, and transgender (LGBT) individuals using social media. *Soc Sci Med.* 2018;215:92–97. https://doi.org/10.1016/j.socscimed.2018.08.031.
40. Levy S, Campbell MD, Shea CL, DuPont R. Trends in abstaining from substance use in adolescents: 1975−2014. *Pediatrics.* 2018;142(2):e20173498. https://doi.org/10.1542/peds.2017-3498.
41. Sudhinaraset M, Wigglesworth C, Takeuchi DT. Social and cultural contexts of alcohol use influences in a social−ecological framework. *Alcohol Res.*;38(1):35–45. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4872611/. Accessed June 15, 2019.
42. Kerr WC, Mulia N, Zemore SE. U.S. trends in light, moderate, and heavy drinking episodes from 2000 to 2010. *Alcohol Clin Exp Res.* 2014;38(9):2496–2501. https://doi.org/10.1111/acer.12521.
43. Kerr WC, Greenfield TK, Bond J, Ye Y, Rehm J. Age-period-cohort modelling of alcohol volume and heavy drinking days in the US National Alcohol Surveys: divergence in younger and older adult trends. *Addiction.* 2009;104(1):27–37. https://doi.org/10.1111/j.1360-0443.2008.02391.x.
44. West JH, Hall PC, Prier K, et al. Temporal variability of problem drinking on Twitter. *Open J Prev Med.* 2012;2(1):43–48. https://doi.org/10.4236/ojpm.2012.21007.
45. Kushnir V, Cunningham JA. Event-specific drinking in the general population. *J Stud Alcohol Drugs.* 2014;75(6):968–972. https://doi.org/10.15288/jsad.2014.75.968.
46. Neighbors C, Atkins DC, Lewis MA, et al. Event-specific drinking among college students. *Psychol Addict Behav.* 2011;25(4):702–707. https://doi.org/10.1037/a0024051.
47. Carrotte ER, Dietze PM, Wright CJ, Lim MS. Who "likes" alcohol? Young Australians' engagement with alcohol marketing via social media and related alcohol consumption patterns. *Aust N Z J Public Health.* 2016;40(5):474–479. https://doi.org/10.1111/1753-6405.12572.
48. Jones SC, Robinson L, Barrie L, Francis K, Lee JK. Association between young Australian's drinking behaviours and their interactions with alcohol brands on Facebook: results of an online survey. *Alcohol Alcohol.* 2016;51(4):474–480. https://doi.org/10.1093/alcalc/agv113.
49. George MD, Holder HD, McKenzie PN, Mueller HR, Herchek DC, Faile BS. Replication of a controlled community prevention trial: results from a local implementation of science-based intervention to reduce impaired driving. *J Prim Prev.* 2018;39(1):47–58. https://doi.org/10.1007/s10935-017-0499-y.
50. Bermea AM, Lardier DT, Forenza B, Garcia-Reid P, Reid RJ. Communitarianism and youth empowerment: motivation for participation in a community-based substance abuse prevention coalition. *J Commun Psychol.* 2019;47(1):49–62. https://doi.org/10.1002/jcop.22098.
51. Nelson TF, Naimi TS, Brewer RD, Wechsler H. The state sets the rate: the relationship among state-specific college binge drinking, state binge drinking rates, and selected state alcohol control policies. *Am J Public Health.* 2005;95(3):441–446. https://doi.org/10.2105/AJPH.2004.043810.
52. Hingson RW, Zha W, White AM. Drinking beyond the binge threshold: predictors, consequences, and changes in the U.S. *Am J Prev Med.* 2017;52(6):717–727. https://doi.org/10.1016/j.amepre.2017.02.014.
53. Pierannunzi C, Hu SS, Balluz L. A systematic review of publications assessing reliability and validity of the Behavioral Risk Factor Surveillance System (BRFSS), 2004−2011. *BMC Med Res Methodol.* 2013;13:49. https://doi.org/10.1186/1471-2288-13-49.
54. Chunara R, Wisk LE, Weitzman ER. Denominator issues for personally generated data in population health monitoring. *Am J Prev Med.* 2017;52(4):549–553. https://doi.org/10.1016/j.amepre.2016.10.038.
55. Mavragani A, Ochoa G, Tsagarakis KP. Assessing the methods, tools, and statistical approaches in Google trends research: systematic review. *J Med Internet Res.* 2018;20(11):e270. https://doi.org/10.2196/jmir.9366.