Available online at www.sciencedirect.com

Public Health

journal homepage: www.elsevier.com/puhe

Original Research

Social media indicators of the food environment and state health outcomes

Q.C. Nguyen [a,*], H. Meng [a], D. Li [b], S. Kath [c], M. McCullough [d], D. Paul [c], P. Kanokvimankul [a], T.X. Nguyen [e], F. Li [c]

[a] *Department of Health, Kinesiology, and Recreation, College of Health, University of Utah, Salt Lake City, United States*
[b] *Center for Systems Integration and Sustainability, Michigan State University, East Lansing, United States*
[c] *School of Computing, University of Utah, Salt Lake City, United States*
[d] *Department of Geography, University of Utah, Salt Lake City, United States*
[e] *Department of Epidemiology and Biostatistics, UCSF School of Medicine, San Francisco, United States*

ARTICLE INFO

Article history:
Received 1 November 2016
Received in revised form
24 March 2017
Accepted 28 March 2017

Keywords:
Social determinants
Social media
Geography
Food and nutrition
Chronic disease
Data mining

ABSTRACT

Objectives: Contextual factors can influence health through exposures to health-promoting and risk-inducing factors. The aim of this study was to (1) build, from geotagged Twitter and Yelp data, a national food environment database and (2) to test associations between state food environment indicators and health outcomes.

Study design: This is a cross-sectional study based upon secondary analyses of publicly available data.

Methods: Using Twitter's Streaming Application Programming Interface (API), we collected and processed 4,041,521 food-related, geotagged tweets between April 2015 and March 2016. Using Yelp's Search API, we collected data on 505,554 unique food-related businesses. In linear regression models, we examined associations between food environment characteristics and state-level health outcomes, controlling for state-level differences in age, percent non-Hispanic white, and median household income.

Results: A one standard deviation increase in caloric density of food tweets was related to higher all-cause mortality (+46.50 per 100,000), diabetes (+0.75%), obesity (+1.78%), high cholesterol (+1.40%), and fair/poor self-rated health (2.01%). More burger Yelp listings were related to higher prevalence of diabetes (+0.55%), obesity (1.35%), and fair/poor self-rated health (1.12%). More alcohol tweets and Yelp bars and pub listings were related to higher state-level binge drinking and heavy drinking, but lower mortality and lower percent reporting fair/poor self-rated health. Supplemental analyses with county-level social media indicators and county health outcomes resulted in finding similar but slightly attenuated associations compared to those found at the state level.

Conclusions: Social media can be utilized to create indicators of the food environment that are associated with area-level mortality, health behaviors, and chronic conditions.

[*] *Corresponding author*. Department of Health, Kinesiology, and Recreation, University of Utah, 1901 E South Campus Drive, Annex B 2124, Salt Lake City, UT 84112, United States. Tel.: +1 (801) 585 5134.

E-mail address: quynh.nguyen@health.utah.edu (Q.C. Nguyen).

http://dx.doi.org/10.1016/j.puhe.2017.03.013
0033-3506/© 2017 The Royal Society for Public Health. Published by Elsevier Ltd. All rights reserved.

Introduction

Background

Food environment characteristics are critical contextual factors affecting how people access food.[1] Varying sociocultural conditions and physical features of the environment influence food choices.[2] For example, people are concerned about food quality and availability, locations of stores and restaurants, prices, customer service, and operating hours.[1,2] Food environments, which can be characterized by risk factors (such as exposure to high caloric foods) as well as health-promoting factors (such as availability of healthy food stores), can impact health. Higher prevalence of fast food restaurants have been found to be related to higher obesity rates at the state level.[3] Another state-level analysis found that higher per capita number of fast food and full-service restaurants and reduced price of meals correlated with higher obesity rates.[4] Conversely, areas with prominent access to healthy food outlets enable diets with fresh and healthy food. Studies have documented increased fruit and vegetable consumption[5] and lower body mass index[6] with more supermarket availability.[7] Poor and minority neighborhoods have fewer large supermarkets than wealthy and majority white neighborhoods,[8] which may increase health disparities.

Social media, such as Twitter, are online forms of communication where people create content, share information, and engage in social networking. Twitter can be used as a tool to examine individuals' food decision-making and how that is patterned by their built food environment.[9] Chen and Yang found that higher numbers of green retailer (grocery stores chains and local fruit and vegetable stores) within a buffered distance of the Twitter user's geotagged location was associated with more healthful food tweets. This significant difference may indicate that people living in healthier food environments may engage in healthier eating behaviors.[7] Ghosh and Guha found a strong positive correlation between tweets about high calorie foods/obesity and locations of McDonalds.[10] Widener and Li found that disadvantaged areas had fewer positive Twitter references for fruits and vegetables.[11]

Moreover, social processes may influence health behaviors. Social processes can affect health via (1) the maintenance of norms around healthy behaviors, (2) stimulation of interest in new activities, (3) emotional support for making healthy choices, (4) the dispersal of knowledge about health promotion practices, and (5) political advocacy and collective action around health.[12–16] Ghosh and Guha found obesity-prevention-themed tweets positively correlated with the number of policies related to obesity, nutrition, and physical activity at the state level,[10] possibly indicating higher levels of health advocacy in certain areas. Children who live in states with weaker competitive food and beverage laws are at greater risk of being overweight or obese than their peers who live in states with strong laws.[17] The social environment can not only offer opportunities for social control, in regulating unhealthy behaviors and facilitating the social learning of healthy behaviors but can also promote risky behaviors. The spread of health behaviors such as food consumption, health screening, smoking, alcohol consumption, drug use, and sleep has been observed to spread through social networks.[18–21]

Social media data have also been analyzed to understand how individuals communicate health topics, the popularity of topics, and sentiment towards current health topics (e.g. vaccines).[22] For instance, Myslín et al. analyzed tweets to examine sentiment towards various tobacco products and found that hookah and electronic cigarettes were characterized by more positive sentiment than references to traditional tobacco products.[23] Social learning theory posits that learning is a cognitive process that occurs in a social context. Views and activities described via social media can help shape perceived norms, attitudes, beliefs, and subsequently behaviors of people. Liking or following alcohol marketing social media pages has been found to be associated with early age at first alcohol use and heavier alcohol consumption among youth.[24] Social media have been utilized for health education and behavioral change interventions such as those aimed to increase physical activity and decrease smoking. Social media can be used for health promotion campaigns to provide health information and social support.[25,26] In addition, user-driven websites and applications such as Yelp have emerged to provide a platform for people to post reviews and testimonies of local businesses and services. In 2016, Yelp's mobile app averaged 65 million users per month.[27] Yelp reviews have been leveraged to understand patient experiences at health facilities—information which can be utilized to improve quality of care.[28] Yelp data can be used to understand the types of food businesses in a community and the popularity of various foods.

In this study, we examine factors related to the food environment. From Twitter data, we obtain indicators of socially modeled eating and drinking behaviors, possibly capturing prevalent norms and preferences around food. From Yelp data, we assess the availability and popularity of cuisines as perceived by visitors to restaurants. The widespread use of the internet and the abundance of openly shared personal opinions with geotagged check-ins at various locations enable researchers to understand area characteristics, which are unique strengths of utilizing social media data over traditional means of data collection.

Study aims

The present study constructs a national database of food environment indicators from publicly available Twitter and Yelp data. We then test associations between state-level food environment indicators and health outcomes, accounting for differences in state demographic characteristics via census data that may act as potential confounders related to both food environment indicators and health outcomes.

Methods

Twitter data collection and spatial join

For approximately one year, from April 2015–March 2016, we utilized Twitter's Streaming Application Programming Interface (API) to continuously collect a random 1% sample of publicly available tweets with latitude and longitude coordinates. In total, we collected 79,848,992 geotagged tweets from 603,363 unique Twitter users in the contiguous United States (excluding Alaska and Hawaii). The median number of tweets per user was four. Each geotagged tweet was assigned to its corresponding state location, based on the latitude and longitude coordinates of where the tweet was sent. This spatial join procedure was implemented in Python. An R-tree was used to build a spatial index[29] on the national polygon data to speed up computation. We linked 99.8% of tweets to their respective state locations. Further description of our methodology can be found here.[30]

Sentiment analysis

To conduct sentiment analysis, we utilized MAchine Learning for LanguagE Toolkit (MALLET)—a Java-based package for statistical natural language processing. We leveraged the Maximum Entropy text classifier in MALLET to classify tweets as happy and not happy.[31] MALLET assigns to each tweet a predicted probability from 0 to 100% that a tweet is happy based upon word-level features. During our pilot testing of the classifier, we compared manual labeling of a random subset of tweets with tweets classified as 'happy' or 'not happy'. We concluded a predicted probability ≥80% achieved the highest accuracy between manual labels and machine-learning algorithm labels, and thus that cut point was used to label tweets as 'happy' vs 'not happy.' An example of a happy tweet is 'best day of my life. eating all the cheese.' An example of a unhappy tweet is: 'my stomach hurts … i ate too much pizza and wings.' The accuracy of the sentiment algorithm was 78%.[30]

Food analysis

We compiled a list of over 1430 popular foods and beverages from the US Department of Agriculture's National Nutrient Database.[32] Each food item was associated with a measure of caloric density, operationalized as calories per 100 g. Fruits, vegetables, nuts, and lean proteins (e.g. fish, chicken, and turkey) were labeled as 'healthy foods' (340 food terms in total). Fried foods were not considered healthy foods. We also tracked mentions of popular alcoholic beverages (e.g. martini) and tweets referencing drinking (66 terms).

We computed caloric density by summing up all the foods mentioned in the tweet. We tracked healthy food references for each tweet. Moreover, we leveraged our sentiment analysis algorithms to assess sentiment towards food. These variables (i.e. any food references, healthy food references, alcohol references, caloric density, and sentiment towards healthy foods and alcohol) were then aggregated and summarized at the state level to create state indicators of food culture. Over the study period, we collected and processed 4,041,521 geotagged food tweets. Accuracy between algorithm-labeled tweets and manually generated labels was 83% food and 88% alcohol.

Yelp data collection

We used Yelp's search API with public end-points (/v2/search) to obtain the top 20 restaurants, based on ratings for a particular location. Yelp provides a radius filter parameter, which enables search for Yelp listings within a boundary. We generated a set of search locations in terms of longitudes and latitudes covering the entire Unites States using a radius filter of 1000 m. We also added restaurant location data from OpenStreetMap to boost coverage of our existing search set. As our location set was large (composed of entire United States) we created a streaming API, which used the Yelp public end-points described above to obtain the data. Because there was a daily limit of approximately 25,000 queries that can be performed a day, we designed our system to conduct 24,000 unique searches on a daily basis.

Our Yelp data collection resulted in finding 505,554 unique food-related business entries collected between February–April 2016. We processed these entries into broad food themes including: International Cuisine (Thai, Indian, Chinese, Japanese, French, Italian, Mexican), Cafés and Bakeries (coffee and tea, cafes, bakeries, bagels, bubble tea), and bars and pubs (beer, wine, spirits, sports bar, sports pub). To create Yelp state-level indicators, we calculated the percent of each food theme out of all food-related Yelp entries for that state.

Other publicly state-level data

We obtained state-level health outcome data—including all-cause mortality and homicide rate from the 2013 National Vital Statistics Reports. Vital statistics data were based on information from resident death certificates filed in the 50 states and the District of Columbia. Death certificates are generally completed by funeral directors, attending physicians, medical examiners, and coroners. Age-adjusted death rates expressed per 100,000 population were based on the 2000 US standard population. Causes of death statistics were classified by the 10th Revision of the International Classification of Diseases and based on the underlying cause of death.

We collected age-adjusted prevalence of health-related risk behaviors and chronic health conditions of US adult residents at state level from the Behavioral Risk Factor Surveillance System (BRFSS), the nation's premier system of health-related telephone surveys. The BRFSS questionnaires were created by BRFSS state coordinators and Centers for Disease Control and Prevention (CDC) staff. Each BRFSS survey has three parts: the core component including a set of questions on demographic characteristics and current health behaviors, optional modules, and state-added questions. Our study data was based on the BRFSS 2014 Questionnaire, which included assessments of alcohol consumption, tobacco use, physical activity, self-rated health status, self-reported body mass index, and diagnosis by a healthcare professional for diabetes and high cholesterol.

Analytic approach

Twitter characteristics were standardized to have a mean of 0 and standard deviation of 1. In adjusted linear regression models, we examined associations between food environmental characteristics (constructed from Twitter and Yelp data) and state health outcomes, controlling for demographic characteristics. Models were run separately for each health outcome. Demographic characteristics were obtained from the 2010–2014 American Community Survey 5-year estimates and included the following: median age, percentage non-Hispanic white, and median household income to capture information on compositional and economic characteristics of a community. We evaluated statistical significance at $P < 0.05$. Processing and statistical analysis tasks were performed with Stata MP13 (StataCorp LP, College Station, TX). The study was approved by the authors' Institutional Review Board.

Results

Table 1 presents descriptive statistics of our social media database. About 5% of tweets were food-related. The average caloric density of food tweets was 234 (per 100 g). About one in six (16%) food tweets included healthy food (lean proteins, fruits, vegetables) references. About 28% of healthy food tweets and 35% of alcohol tweets were categorized as happy (Table 1) when compared to 19% of tweets overall. Regarding Yelp data, about 27% of popular Yelp listings were for international cuisines, 10% for burger places, 9.4% for bars and pubs, and 5.7% for cafes and bakeries (Table 1).

Supplementary Table 1 (see Appendix A for a link to the supplementary data) presents state rankings of Twitter-derived variables. For happiness, Montana ranked the highest and Louisiana the lowest (Supplementary Fig. 1). Interestingly Louisiana also held positions near the bottom for tweets about healthy foods (Fig. 1). Montana had the most healthy food tweets. Mississippi had the highest caloric density of food mentions. Vermont, Maine, and Wisconsin were the states with the most alcohol tweets (Supplementary Table 1).

Table 1 – Descriptive statistics of social media characteristics.

Social media-derived indicators	Mean (standard deviation)
Twitter data[a]	
Percentage tweets about food	5.0 (1.0)
Calories density (calories per 100 g)	233.7 (14.1)
Percentage food tweets about healthy foods	16.3 (2.4)
Percentage tweets about alcohol	0.8 (0.2)
Percentage food tweets that are happy	27.0 (2.3)
Percentage healthy food tweets that are happy	28.0 (2.1)
Percentage alcohol tweets that are happy	35.0 (3.9)
Yelp data[b]	
Percentage café and bakeries	5.7 (1.4)
Percentage bars and pubs	9.4 (2.1)
Percentage burger places	10.0 (1.8)
Percentage international cuisines	27.0 (4.2)

[a] Twitter indicators created from 79,848,992 geotagged tweets collected between March 2015 and April 2016. $n = 49$. Data from the contiguous United States including District of Columbia; does not include Alaska and Hawaii.

[b] Yelp state indicators created from 505,554 food-related entries collected between February and April 2016.

Supplementary Table 2 presents state rankings of Yelp-derived variables of food-theme businesses. Vermont, Montana, Maine, Oregon, and Washington have the top positions for cafes and bakeries (8–10%) (Fig. 2). Wisconsin, Minnesota, and the District of Colombia have high percentages of popular Yelp listings that are bars/pubs (13–15%) (Fig. 3). Utah, Nebraska, Alabama, and Indiana have the highest percentages of popular Yelp listings that are burger places (12–13%). For international cuisine, California, Arizona, New Mexico, and Texas top the list with 34–38% of popular Yelp listings for this category (Supplementary Fig. 2).

Moreover, we investigated whether state-level food environment variables derived from social media data are associated with health outcomes (Table 2). We found that higher caloric density of food tweets were related to higher mortality, chronic conditions, and fair/poor self-rated health. In addition, more burger Yelp listings were related to higher prevalence of diabetes, obesity, and fair/poor self-rated health. Conversely, more cafe and bakery Yelp listings were associated with lower mortality, lower prevalence of chronic conditions, and better self-rated health (Table 2). Higher percentages of alcohol tweet were related to lower mortality, lower homicide rates, and better self-rated health but were also related to more binge drinking and heavy drinking (Table 3). Similarly, higher percentages of popular Yelp entries for bars and pubs were related to lower mortality and better self-rated health, but more binge drinking and heavy drinking (Table 3).

Additionally, given that younger-aged individuals are over-represented among Twitter and Yelp users than in the general population, we ran sensitivity analyses with obesity prevalence among 26–44 year olds as the outcome variable. Associations were moderately stronger compared to those seen for obesity prevalence among all adults (Supplementary Table 3). We also examined associations between Twitter- and Yelp-derived food environmental characteristics and health outcomes at the county level. Generally, we found that associations are in the same direction but attenuated at the county level when compared to the state level (Supplementary Tables 4 and 5).

Discussion

In this study, we utilize innovative data collection and processing techniques to characterize the food environments of states with data from Twitter and Yelp. We do not expect these data to perfectly capture dietary behaviors—because for instance, not everyone tweets all the foods they consume. Nonetheless, these data may be useful indicators of area-level social norms and preferences. Moreover, public posts about certain foods and behaviors may influence the opinions and behaviors of others, as posited by theoretical frameworks on the role of social networks.[33] A variety of health behaviors such as dietary patterns, health screening, substance use and sleep have been observed to spread through social networks.[18–21]

In addition to creating a national state-level database, we investigated whether food environment characteristics are associated with state health outcomes. We found that Twitter and Yelp characteristics indicative of higher caloric foods were related to higher mortality, higher prevalence of chronic

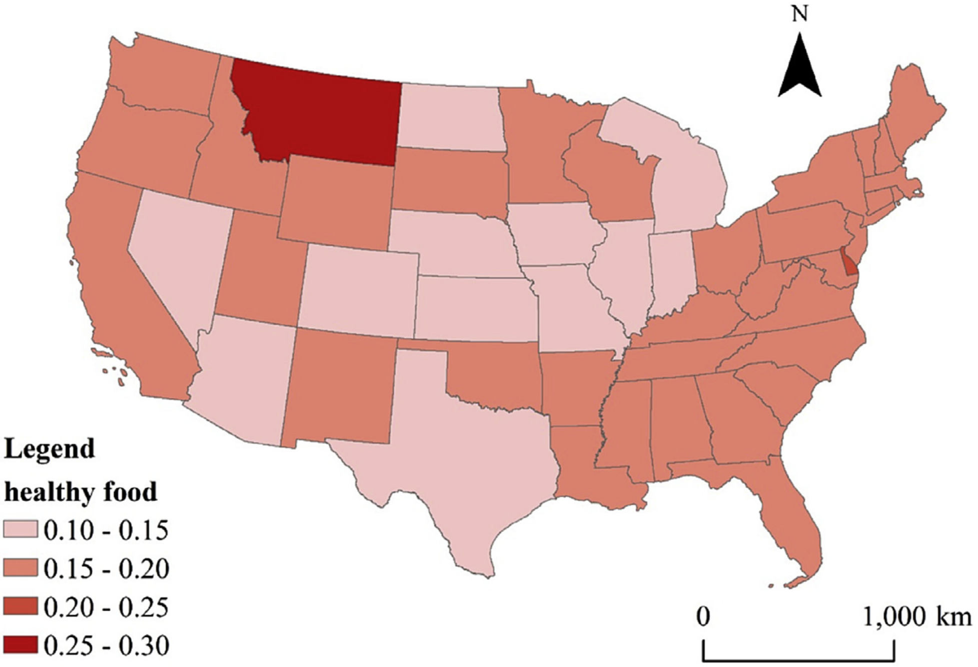

Fig. 1 – **National distribution of healthy food tweets, by state. Figure displays proportion of food tweets that contained healthy food mentions, defined as the following: fruits, vegetables, nuts, and lean proteins (e.g. fish, chicken, and turkey). Fried foods were not considered healthy foods.**

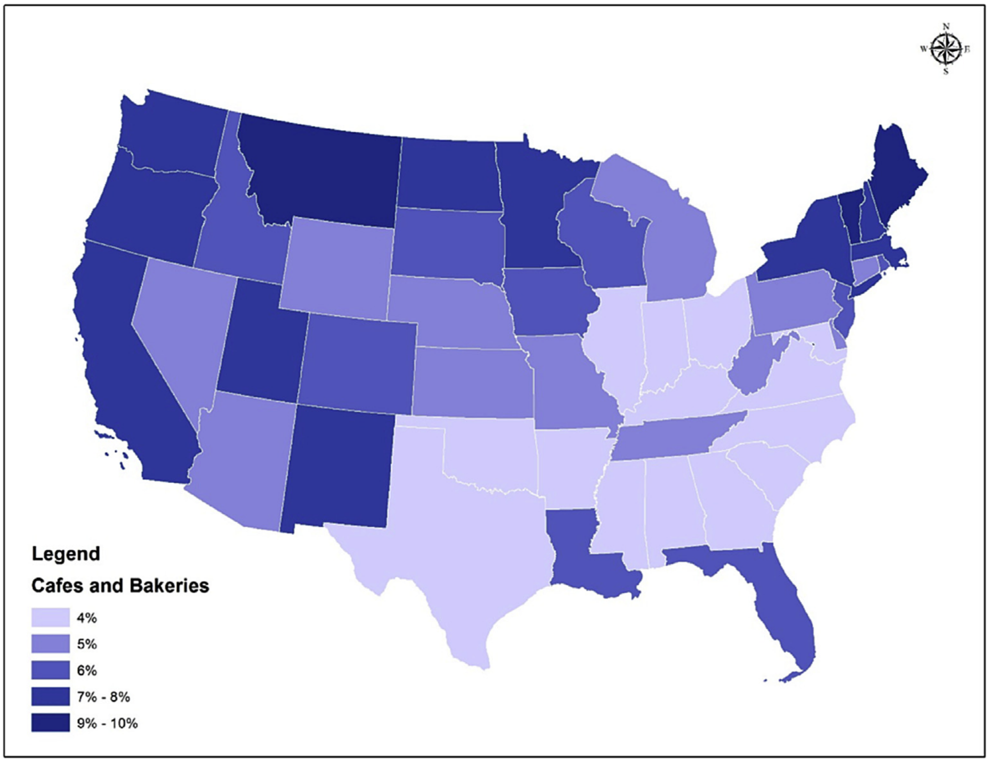

Fig. 2 – **Yelp entries for cafes and bakeries, by state. Figure displays percent of popular Yelp entries at the state level that were cafes and bakeries.**

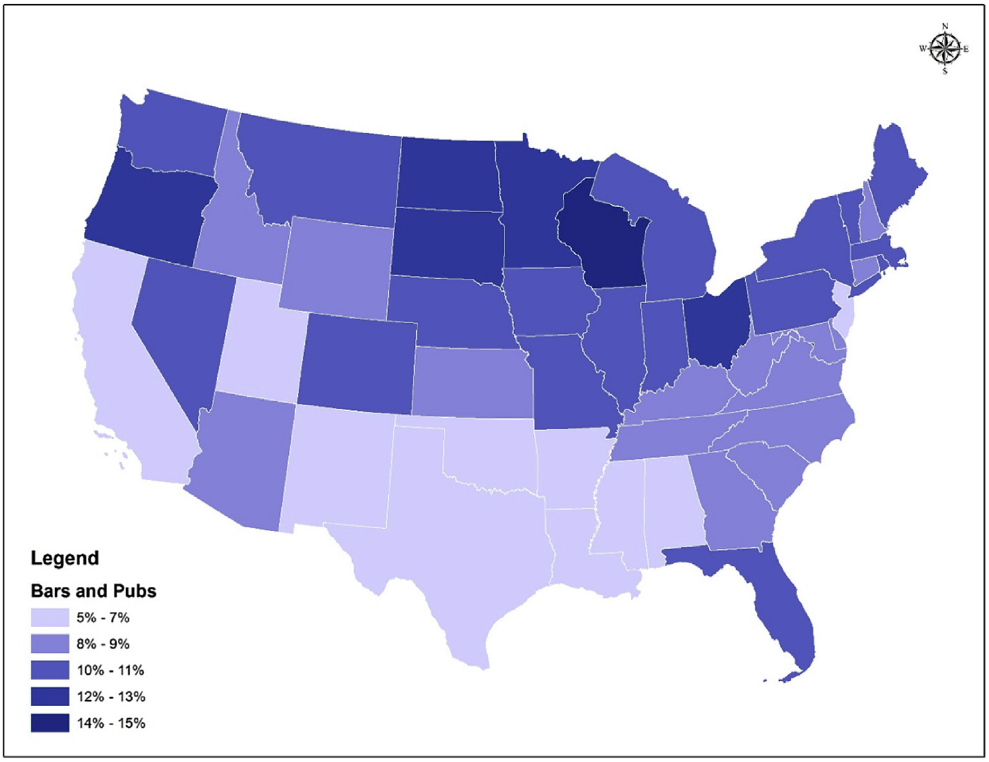

Fig. 3 – Yelp food entries for bars and pubs, by state. Figure displays percent of popular Yelp entries at the state level that were bars and pubs.

Table 2 – State-level food environment and health outcomes.

State-level adult health outcomes	State-level Twitter variables[a]		
	Caloric density of Twitter food mentions	Percent Yelp listing, burgers	Percent Yelp listing, café and bakeries
	Beta (95% CI)[b]	Beta (95% CI)[b]	Beta (95% CI)[b]
All-cause mortality per 100,000	46.50 (25.81–67.20)**	16.85 (−9.89 to 43.59)	−31.06 (−48.69 to −13.44)**
Percent diabetes	0.75 (0.42–1.09)**	0.55% (0.14–0.96)**	−0.66% (−0.92 to −0.41)**
Percent obesity	1.78 (0.89–2.67)**	1.35% (0.29–2.40)*	−1.92% (−2.51 to −1.32)**
Percent high cholesterol	1.40 (0.79–2.00)**	0.36% (−0.43 to 1.16)	−1.09% (−1.58 to −0.60)**
Percent poor/fair self-rated health	2.01 (1.40–2.61)**	1.12% (0.25–1.98)*	−1.06% (−1.66 to −0.45)**

CI, confidence interval.

*P < 0.05; **P < 0.01.

[a] Twitter-derived variables (independent variables in regression models) were standardized to have a mean of 0 and standard deviation of 1. $n = 49$. States in the contiguous United States, including District of Columbia.

[b] Adjusted linear regression models were run for each outcome separately. Models controlled for state-level demographics: median age, % non-Hispanic white, median household income. Data sources for health outcomes: 2013 National Vital Statistics Reports, 2014 Behavioral Risk Factor Surveillance System.

conditions, and worse self-rated health. Alternatively, popularity of cafes and bakeries were associated with better health outcomes. Lastly, higher alcohol mentions and popularity of bars and pubs were associated with lower mortality but higher binge drinking and heavy drinking.

Study findings in context

Regarding alcohol characteristics, we find that tweets with alcohol mentions are happier than tweets in general, which aligns with the results of a recent study that analyzed nearly 12 million alcohol-related tweets and found that the number of pro-drinking tweets exceeded anti-drinking tweets by 10 times.[34] While an alcohol-related tweet does not necessarily indicate actual alcohol consumption, greater alcohol use is associated with greater alcohol content shared online.[35] Also alcohol consumption like other health behaviors can be influenced by social norms and a ‘culture of drinking.’[20,36] For instance, Ahern et al. found that neighborhoods with permissive norms around drunkenness had higher rates of

Table 3 – State-level alcohol characteristics and health outcomes.

State-level adult health outcomes	State-level Twitter variables[a]	
	Percent tweets about alcohol	Percent Yelp listing, bars and pubs[a]
	Beta (95% CI)[b]	Beta (95% CI)[b]
All-cause mortality per 100,000	−43.30 (−63.56 to −23.04)**	−21.60 (−40.51 to −2.68)*
Homicide per 100,000	−0.85 (−1.59 to −0.12)*	0.05 (−0.57 to 0.67)
Suicide per 100,000	0.22 (−0.93 to 1.38)	−0.50 (−1.45 to 0.44)
Unintentional injury death	−2.17 (−4.63 to 0.30)	−1.45 (−3.52 to 0.62)
Percent poor/fair self-rated health	−1.66 (−2.32 to −1.01)**	−1.23 (−1.80 to −0.66)**
Percent binge drinking	1.86 (0.85 to 2.88)**	2.30 (1.63 to 2.96)**
Percent heavy drinking	0.75 (0.40 to 1.10)**	0.62 (0.33 to 0.91)**
Percent current smoking	−1.39 (−2.31 to −0.48)**	0.09 (−0.75 to 0.93)

CI, confidence interval.

*P < 0.05; **P < 0.01.

[a] Twitter-derived variables (independent variables in regression models) were standardized to have a mean of 0 and standard deviation of 1. $n = 49$. States in the contiguous United States, including District of Columbia.

[b] Adjusted linear regression models were run for each outcome separately. Models controlled for state-level demographics: median age, % non-Hispanic white, median household income. Data sources for health outcomes: 2013 National Vital Statistics Reports, 2014 Behavioral Risk Factor Surveillance System.

moderate and binge drinking.[36] The inverse relationship between alcohol characteristics and health outcomes may be related to health benefits associated with moderate drinking.[37] Alcohol consumption has been found to be associated with reduced mortality and cardiovascular disease risk.[38,39] Nonetheless, alcohol consumption also increases the risk of a range of adverse health outcomes including accidental injury, alcohol-related family, and occupation problems, fetal alcohol syndrome, liver diseases, and cancers.[40–43] Surprisingly, we did not find links between alcohol characteristics derived from social media and state outcomes relating to suicide and accidental injury. The culture of drinking captured via public social media posts may be more weighted towards those who engage in social drinking rather than substance abuse and dependence. Communities with higher social drinking may see some health benefits.

Our findings regarding ties between popularity of cafes and better health outcomes are in alignment with previous studies finding health benefits of coffee and tea for the prevention of chronic conditions[44] and cancer.[45] Additionally cafes are among the public spaces in a community that provide opportunities for social gatherings outside the office and home.[46] Thus, having more cafes in a community may boost social interactions, trust, and willingness of local residents to intervene for the common good—characteristics that have been related to better health outcomes.[14,47–49] Nevertheless, cafes also differ with regard to the quality of food items served, which may impact area-level health outcomes.

Study strengths and limitations

We built a national dataset using publicly available Twitter and Yelp data. Previous studies involving social media have demonstrated great potential for making area comparisons and investigating differences in health outcomes at selected locations.[50,51] This study contributes by showcasing that large scale comparisons at a national level can be achieved using interdisciplinary approaches that leverage machine learning, geographical information systems, and regression modeling.

However, this study is subject to limitations. First, Twitter users are not representative of the general US population. Currently about 23% of adults online (20% of US adults) utilize Twitter.[52] Usage is higher among African Americans (28%) and Hispanics (28%) compared to white, non-Hispanics (20%). Twitter usage is also higher among those aged 18–29 years (32%) and 30–49 years (29%) than to those 50–64 years (13%) and 65+ years (6%).[52] More men online use Twitter than women (25% vs 21%).[52] The health and health behaviors of social media users may differ from those who do not use social media.

Although characteristics constructed from Twitter and Yelp data may not be representative of the general population, nonetheless, the data may still have utility in providing information on the social environment of a community. That is, while not everyone is represented, the large collection of tweets sent out by members of a community, may still help us understand social influences on health. For instance, an analysis of Twitter data by Eichstaedt et al. found that psychological language on Twitter predicted county-level heart disease mortality—a condition that typically does not afflict young individuals.[53] Hence, even though Twitter data may be skewed towards adolescents and young adults, their online expressions may still reflect shared environmental features of the community at large.

In this study, we utilized Twitter data to create indicators of socially modeled food behaviors. We utilized Yelp to create indicators of food access and popularity of cuisines. However, we did not examine other important sociocultural factors that influence food choices. For example, people are concerned about prices, customer service, operating hours, food quality and availability, and locations of stores and restaurants. Preferred locations for shopping include safe and low crime neighborhoods.[1] Mobility is another contextual factor that has been overlooked by many researchers. People might go outside their neighborhoods to get food when traveling to meet families and friends or to go to work.[2]

The study explored associations between area-level characteristics and area-level health outcomes. Patterns observed at the area level may not apply to the individual level (ecological fallacy). Causal inference is limited by the observational nature of the study and the lack of control for confounding by individual-level characteristics. Analyses were conducted at the state and county levels. Analyses at smaller levels of geographies are limited by the lack of publicly available health outcome data at those levels. Also at smaller levels of geographies, merging over several years of data may be necessary to produce reliable estimates and to protect the confidentiality of participants' data.

Additionally, tweets were geocoded to the locations of where they were sent rather than to the home locations of Twitter users. Thus, tweets identified as coming from an area could include data from local residents and visitors. However, we believe that including data from visitors is important because visitors can influence the social environment in which they interact. Additionally, the data are naturally weighted to those who spend more time in an area.

Another limitation of the study is the time periods for the various data sources differed depending on data availability of publicly available data sources. Nonetheless, resident compositional characteristics, mortality rates, and prevalence of health conditions have been found to be relatively stable over a short time period.[54] Our study design was cross-sectional in nature; we were not able to examine longitudinal trends.

Characteristics of the built and social environment can impact health and overall well-being by determining access to resources or exposures to risk. In this study, we demonstrate that social media can be utilized to create indicators of the food environment that are associated with state-level mortality, health behaviors, and chronic conditions. Social media represents an untapped resource for public health research and practice.

Author statements

Acknowledgments

The authors would like to thank Drs. Ming Wen, James VanDerslice, and Ken Smith for their guidance and input on this project.

Ethical approval

The study was approved by the University of Utah Institutional Review Board.

Funding

This study is supported by a grant from NIH's Big Data to Knowledge Initiative, 5K01ES025433.

Competing interests

None declared.

REFERENCES

1. Greenwood S, Perrin A, Duggan M. *Social media update 2016.* Washington, DC: Pew Research Center; 2016.
2. Chen X, Kwan M-P. Contextual uncertainties, human mobility, and perceived food environment: the uncertain geographic context problem in food access research. *Am J Public Health* 2015;**105**:1734–7.
3. Maddock J. The relationship between obesity and the prevalence of fast food restaurants: state-level analysis. *Am J Health Promot* 2004;**19**:137–43.
4. Chou S-Y, Grossman M, Saffer H. An economic analysis of adult obesity: results from the Behavioral Risk Factor Surveillance System. *J Health Econ* 2004;**23**:565–87.
5. Morland K, Wing S, Roux AD. The contextual effect of the local food environment on residents' diets: the atherosclerosis risk in communities study. *Am J Public Health* 2002;**92**:1761–7.
6. Brown AF, Vargas RB, Ang A, Pebley AR. The neighborhood food resource environment and the health of residents with chronic conditions. *J General Intern Med* 2008;**23**:1137–44.
7. Black C, Moon G, Baird J. Dietary inequalities: what is the evidence for the effect of the neighbourhood food environment? *Health Place* 2014;**27**:229–42.
8. Morland K, Wing S, Diez-Roux AV, Poole C. Neighborhood characteristics associated with the location of food stores and food service places. *Am J Prev Med* 2001;**22**:23–9.
9. Chen X, Yang X. Does food environment influence food choices? A geographical analysis through "tweets". *Appl Geogr* 2014;**51**:82–9.
10. Ghosh D, Guha R. What are we 'tweeting' about obesity? Mapping tweets with Topic Modeling and Geographic Information System. *Cartogr Geogr Inf Sci* 2013;**40**:90–102.
11. Widener MJ, Li W. Using geolocated twitter data to monitor the prevalence of healthy and unhealthy food references across the US. *Appl Geogr* 2014;**54**:189–97.
12. Ali MM, Amialchuk A, Heiland FW. Weight-related behavior among adolescents: the role of peer effects. *PLoS One* 2011;**6**:e21179.
13. Vartanian LR, Sokol N, Herman CP, Polivy J. Social models provide a norm of appropriate food intake for young women. *PLoS One* 2013;**8**:e79268.
14. Cohen DA, Finch BK, Bower A, Sastry N. Collective efficacy and obesity: the potential influence of social factors on health. *Soc Sci Med* 2006;**62**:769–78.
15. Kim D, Subramanian SV, Gortmaker SL, Kawachi I. US state- and county-level social capital in relation to obesity and physical inactivity: a multilevel, multivariable analysis. *Soc Sci Med* 2006;**63**:1045–59.
16. Berkman L, Syme S. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. *Am J Educ* 1979;**190**:186–204.
17. Hennessy E, Oh A, Agurs-Collins T, Chriqui JF, Mâsse LC, Moser RP, et al. State-level school competitive food and beverage laws are associated with children's weight status. *J Sch Health* 2014;**84**:609–16.
18. Pachucki MA, Jacques PF, Christakis NA. Social network concordance in food choice among spouses, friends, and siblings. *Am J Public Health* 2011;**101**:2170–7.
19. Keating NL, O'Malley AJ, Murabito JM, Smith KP, Christakis NA. Minimal social network effects evident in cancer screening behavior. *Cancer* 2011;**117**:3045–52.
20. Rosenquist JN, Murabito J, Fowler JH, Christakis NA. The spread of alcohol consumption behavior in a large social network. *Ann Intern Med* 2010;**152**:426–33.

21. Mednick SC, Christakis NA, Fowler JH. The spread of sleep loss influences drug use in adolescent social networks. *PLoS One* 2010:e9775.
22. Bahk CY, Cumming M, Paushter L, Madoff LC, Thomson A, Brownstein JS. Publicly available online tool facilitates real-time monitoring of vaccine conversations and sentiments. *Health Aff* 2016;**35**:341–7.
23. Myslín M, Zhu S-H, Chapman W, Conway M. Using twitter to examine smoking behavior and perceptions of emerging tobacco products. *J Med Internet Res* 2013;**15**:e174.
24. 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**:474–9.
25. Kousoulis AA, Kympouropoulos SP, Pouli DK, Economopoulos KP, Vardavas CI. From the classroom to facebook a fresh approach for youth tobacco prevention. *Am J Health Promot* 2016;**30**:390–3.
26. Wilson D, Bopp M, Colgan J, Sims D, SA M, Rovniak L, et al. A social media campaign for promoting active travel to a university campus. *J Healthc Commun* 2016;**1**(2):11.
27. Yelp.com. *An introduction to Yelp Metrics as of December 30, 2016*. Yelp.com; 2016 [October 1, 2016]; Available from: https://www.yelp.com/factsheet.
28. Rastegar-Mojarad M, Ye Z, Wall D, Murali N, Lin S. Collecting and analyzing patient experiences of health care from social media. *JMIR Res Protoc* 2015;**4**:e78.
29. Guttman A. R-trees: a dynamic index structure for spatial searching. In: *Proceedings of the 1984 ACM SIGMOD International Conference on Management of Data***14**; 1984. p. 47–57.
30. Nguyen QC, Li D, Meng H-W, Kath S, Nsoesie EO, Li F, et al. Building a national neighborhood dataset from geotagged Twitter data for indicators of happiness, diet, and physical activity. *JMIR Public Health Surveill* 2016;**2**:e158.
31. Nigam K, Lafferty J, McCallum A, editors. *Using maximum entropy for text classification. IJCAI-99 workshop on machine learning for information filtering*; 1999.
32. United States Department of Agriculture. *National nutrient database*. 2014 [February 5, 2014]; Available from: https://ndb.nal.usda.gov/ndb/.
33. Bandura A. *Social learning theory Englewood Cliffs*. New Jersey: Prentice-Hall; 1977.
34. Cavazos-Rehg PA, Krauss MJ, Sowles SJ, Bierut LJ. "Hey everyone, I'm Drunk." an evaluation of drinking-related twitter chatter. *J Stud Alcohol Drugs* 2015;**76**:635–43.
35. Stoddard SA, Bauermeister JA, Gordon-Messer D, Johns M, Zimmerman MA. Permissive norms and young adults' alcohol and marijuana use: the role of online communities. *J Stud Alcohol Drugs* 2012;**73**:968–75.
36. Ahern J, Galea S, Hubbard A, Midanik L, Syme SL. "Culture of drinking" and individual problems with alcohol use. *Am J Epidemiol* 2008;**167**:1041–9.
37. Volk RJ, Cantor SB, Steinbauer JR, Cass AR. Alcohol use disorders, consumption patterns, and health-related quality of life of primary care patients. *Alcohol Clin Exp Res* 1997;**21**:899–905.
38. Snow WM, Murray R, Ekuma O, Tyas SL, Barnes GE. Alcohol use and cardiovascular health outcomes: a comparison across age and gender in the Winnipeg Health and Drinking Survey Cohort. *Age Ageing* 2009;**38**:206–12.
39. German JB, Walzem RL. The health benefits of wine. *Annu Rev Nutr* 2000;**20**:561–93.
40. Wilsnack RW, Vogeltanz ND, Wilsnack SC, Harris TR. Gender differences in alcohol consumption and adverse drinking consequences: cross-cultural patterns. *Addiction* 2000;**95**:251–65.
41. Eustace LW, Kang DH, Coombs D. Fetal alcohol syndrome: a growing concern for health care professionals. *J Obstet Gynecol Neonatal Nurs* 2003;**32**:215–21.
42. Lieber CS. Alcohol, liver, and nutrition. *J Am Coll Nutr* 1991;**10**:602–32.
43. Rehm J, Mathers C, Popova S, Thavorncharoensap M, Teerawattananon Y, Patra J. Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. *Lancet* 2009;**373**:2223–33.
44. Higdon JV, Frei B. Coffee and health: a review of recent human research. *Crit Rev Food Sci Nutr* 2006;**46**:101–23.
45. Cooper R, Morré DJ, Morré DM. Medicinal benefits of green tea: Part I. Review of noncancer health benefits. *J Altern Complement Med* 2005;**11**:521–8.
46. Montgomery J. Café culture and the city: the role of pavement cafés in urban public social life. *J Urban Des* 1997;**2**:83–102.
47. Baum F, Palmer C. 'Opportunity structures': urban landscape, social capital and health promotion in Australia. *Health Promot Int* 2002;**17**:351–61.
48. Ziersch AM, Baum FE, MacDougall C, Putland C. Neighbourhood life and social capital: the implications for health. *Soc Sci Med* 2005;**60**:71–86.
49. Browning CR, Cagney KA. Neighborhood structural disadvantage, collective efficacy, and self-rated physical health in an urban setting. *J Health Soc Behav* 2002;**43**:383–99.
50. Nguyen QC, Kath S, Meng H-W, Li D, Smith KR, VanDerslice JA, et al. Leveraging geotagged Twitter data to examine neighborhood happiness, diet, and physical activity. *Appl Geogr* 2016;**73**:77–88.
51. Nagar R, Yuan Q, Freifeld CC, Santillana M, Nojima A, Chunara R, 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**:e236.
52. Duggan M. *Mobile messaging and social media – 2015*. Pew Research Center; 2015. Available at: http://www.pewinternet.org/2015/08/19/mobile-messaging-and-social-media-2015/.
53. Eichstaedt JC, Schwartz HA, Kern ML, Park G, Labarthe DR, Merchant RM, et al. Psychological language on twitter predicts county-level heart disease mortality. *Psychol Sci* 2015;**26**:159–69.
54. Schmidt NM, Tchetgen Tchetgen EJ, Ehntholt A, Almeida J, Nguyen QC, Molnar BE, et al. Does neighborhood collective efficacy for families change over time? The Boston neighborhood survey. *J Community Psychol* 2014;**42**:61–79.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.puhe.2017.03.013.