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
Available online 4 May 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

## 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