AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

OXFORD

Research and Applications

Uncovering the relationship between food-related discussion on Twitter and neighborhood characteristics

VG Vinod Vydiswaran [iD] ,[1,2] Daniel M Romero [iD] ,[2,3,4] Xinyan Zhao [iD] ,[2] Deahan Yu [iD] ,[2] Iris Gomez-Lopez [iD] ,[5] Jin Xiu Lu,[2] Bradley E Iott [iD] ,[2,6] Ana Baylin [iD] ,[7,8] Erica C Jansen [iD] ,[7] Philippa Clarke [iD] ,[5,8] Veronica J Berrocal [iD] ,[9] Robert Goodspeed [iD] ,[10] and Tiffany C Veinot [iD] [2,11]

[1]Department of Learning Health Sciences, University of Michigan, Ann Arbor, Michigan, USA, [2]School of Information, University of Michigan, Ann Arbor, Michigan, USA, [3]Center for the Study of Complex Systems, University of Michigan, Ann Arbor, Michigan, USA, [4]Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan, USA, [5]Institute for Social Research, University of Michigan, Ann Arbor, Michigan, USA, [6]Department of Health Management and Policy, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA, [7]Department of Nutritional Sciences, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA, [8]Department of Epidemiology, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA, [9]Department of Statistics, Donald Bren School of Information and Computer Science, University of California, Irvine, California, USA, [10]Urban and Regional Planning Program, Taubman College of Architecture and Urban Planning, University of Michigan, Ann Arbor, Michigan, USA, and [11]Department of Health Behavior and Health Education, School of Public Health, University of Michigan, Ann Arbor, Michigan, USA

Corresponding Author: VG Vinod Vydiswaran, PhD, Department of Learning Health Sciences, Medical School, University of Michigan, 1161F NIB, 300 N Ingalls St, Ann Arbor, MI 48109, USA; vgvinodv@umich.edu

ABSTRACT

Objective: Initiatives to reduce neighborhood-based health disparities require access to meaningful, timely, and local information regarding health behavior and its determinants. We examined the validity of Twitter as a source of information for neighborhood-level analysis of dietary choices and attitudes.

Materials and Methods: We analyzed the "healthiness" quotient and sentiment in food-related tweets at the census tract level, and associated them with neighborhood characteristics and health outcomes. We analyzed keywords driving the differences in food healthiness between the most and least-affluent tracts, and qualitatively analyzed contents of a random sample of tweets.

Results: Significant, albeit weak, correlations existed between healthiness and sentiment in food-related tweets and tract-level measures of affluence, disadvantage, race, age, U.S. density, and mortality from conditions associated with obesity. Analyses of keywords driving the differences in food healthiness revealed foods high in saturated fat (eg, pizza, bacon, fries) were mentioned more frequently in less-affluent tracts. Food-related discussion referred to activities (eating, drinking, cooking), locations where food was consumed, and positive (affection, cravings, enjoyment) and negative attitudes (dislike, personal struggles, complaints).

Discussion: Tweet-based healthiness scores largely correlated with offline phenomena in the expected directions. Social media offer less resource-intensive data collection methods than traditional surveys do. Twitter may assist in informing local health programs that focus on drivers of food consumption and could inform interventions focused on attitudes and the food environment.

Conclusions: Twitter provided weak but significant signals concerning food-related behavior and attitudes at the neighborhood level, suggesting its potential usefulness for informing local health disparity reduction efforts.

INTRODUCTION

Background and significance

Negative health outcomes are unequally spatially distributed in the United States. Neighborhood characteristics such as poverty, access to grocery stores, and fast food restaurant density are associated with dietary patterns[1] that influence health outcomes such as obesity, diabetes, kidney failure, and cardiovascular disease.[2] Initiatives to reduce neighborhood-based disparities in food-related behavior and outcomes require access to meaningful, timely, and actionable information regarding that behavior and its determinants at a local level. However, because diets and attitudes are difficult to measure at scale, these phenomena have not yet been discerned at the neighborhood level. In this article, we ask whether social media data can be used to assess dietary patterns and attitudes at the granularity of a neighborhood.

Although social media data have proven useful for characterizing health-related processes,[3,4] we aim to assess discussions related to a specific topic (ie, diet) at a highly localized level (ie, a census tract, roughly equivalent to a neighborhood). Furthermore, we aim to mine social media data at large enough scale that our sample is representative of social media users in the neighborhood. We study Twitter use to characterize neighborhood-level food-related behavior and attitudes and analyze the sentiment expressed in food-related tweets. While different reasons could explain negative and positive sentiment related to food, our goal is to highlight the emotional components of the attitudes of those who tweet about food within the context of the characteristics of neighborhoods from which they tweet and the healthiness of those foods. Emotions are components of attitudes[5,6] and attitudes have been shown to be a correlate of health behavior.[7] Further, research shows that attitudes tend to be transmitted between people and shared by groups.[8] We extend this research and our prior work[9] by studying how tweets contribute to the sharing of food-related attitudes related by studying sentiments expressed in tweets related to both healthy and unhealthy foods.

Related work

Health disparity elimination is a key goal in the U.S. Healthy People 2020 plan.[10] Yet, a recent report[11] notes gaps in the country's health information infrastructure, including a lack of accepted community health indicators.[11] Social media have been used for numerous public health applications,[3,4] showing promise for addressing these gaps. In particular, characterizing local health behavior can assist healthcare and public health professionals and local nonprofits in designing food-related interventions such as educational programs and community gardens and aid state and local governments in designing policies regarding land use, zoning, transportation, food access in public facilities, and incentives to influence dietary behaviors.

Food-related tweets can be geolocated and correlated with spatially referenced health outcomes. For example, the caloric value of foods mentioned in tweets is associated with state-level obesity rates.[12] Additionally, higher obesity prevalence was demonstrated in a clinical sample of people living in zip codes where tweets referred to higher-calorie foods.[13] Yet, little work considers food-related social media posts at the neighborhood scale. State and zip code residents' characteristics vary considerably, and zip codes may not be spatially contiguous.[14] Census tracts are smaller, containing roughly 4000 people, and more demographically homogeneous,[15] and thus reveal more about local food environments. For example, the status of food deserts—urban areas where the closest grocery store is >0.5 miles away or rural areas where it is >10 miles away—is measured at the census tract level.[16] Our research is among the first to correlate tract-level Twitter measures with tract-level health outcomes: mortality from obesity-linked conditions.

Social media–based methods for assessing food-related discussion have primarily focused on caloric estimation and macronutrients. Researchers have identified census tract-level variance in Twitter mentions of calories in 3 U.S. cities associated with tract demographics.[17] Another tract-level study used social media data to identify greater mentions of foods high in fat, cholesterol and sugar in food desert tracts and characterize such tracts' "linguistic signatures."[18] This work reveals potential tract-level variance in food-related discussions. However, previous social media-based measures have lacked key dietary aspects such as sodium intake, fiber, and added sugar.[19] Our work offers more systematic assessment methods.

Additionally, social media have not yet been leveraged to assess changeable factors that may affect eating behavior. For example, social media could reveal local attitudes regarding foods and eating; information-based interventions often target such attitudes, which are modifiable and influence behavior.[20] Furthermore, increasing numbers of interventions attempt to improve neighborhood food supplies through initiatives such as farmer's markets and community gardens.[21] We extend prior work by investigating tweets that reveal attitudes concerning food and locations of acquisition and consumption.

MATERIALS AND METHODS

Regional focus

To investigate the use of Twitter to characterize neighborhood-level food-related behavior and attitudes, we focused on Metropolitan Detroit, also known as the Detroit-Warren-Ann Arbor Combined Statistical Area, which includes 10 counties with 1591 census tracts. Its variability in tract-level food access makes it exemplary for examining the food tweet–neighborhood characteristic relationship. For example, 532 (33.4%) census tracts in the region are food deserts. Additionally, tracts vary in proportions receiving supplemental nutrition assistance program benefits, with a median of 14% and range of 0%-80% of households.[16] Metropolitan Detroit is also home to many community-based programs to improve access to healthy food that could benefit from neighborhood-based information about dietary patterns and attitudes.[22–27]

Data sources

Twitter application programming interfaces (APIs) were used to gather geotagged tweets from Metropolitan Detroit. The location query–based collection was enhanced using the Twitter Gardenhose stream, which provides a 10% random sample of the entire Twitter collection. Tweet authors were identified from tweets gathered through these methods, and their account timelines, sequential lists of their previous tweets up to a limit of 3200, geo-tagged or not, were crawled via the Twitter API and added to the collection. Data

Box 1 Sources used to collect food keywords.

Source	Types of food words
U.S. Department of Agriculture Database	Beef products; beverages; cereal grains and pasta; dairy and egg products; fast foods; fats and oils; finfish and shellfish products; fruit and fruit juices; lamb, veal, and game products; legumes and legume products; meals, entrees, and side dishes; nut and seed products; pork and poultry products; sausages and luncheon meat; sweets; vegetable and vegetable products
Wikipedia	Cookie brands, pastries, candies, popcorn brands, branded snack foods, frozen dessert brands, soda and soft drinks, cakes, ice cream brands, doughnut shops, juice and juice drinks, chocolate bar brands, breakfast cereals, potato chip brands, crackers, deep fried foods, cheeses, processed meat, lunch meat, sausages, duck as food, seafood, comfort food, brand name food products, brand name soft drink products, whole grains, cooking techniques, soul foods and dishes, American Chinese cuisine, American foods, quick breads, baked goods, custard desserts, pudding, dried foods, candy bars, beverages, brand name soft drinks, Mexican dishes, coffee houses, restaurant chains in the United States
Literature	Most popular fast food restaurants in the United States

Box 2 Examples of paraphrased tweets included and excluded as related to food.

Category	Paraphrased tweets
Direct mentions; included	• Tweeter drank 3 glasses of milk in succession and expresses hope this would help them grow. • Tweeter declares love for both another person and hot pockets. • Tweeter refers to going to Qdoba alone. • Tweeter states their mother is making red beans, rice, and chicken.
Indirect mentions; excluded	• Tweeter posts a photo of the sky at an apple orchard. • Tweeter expresses dismay that the lunchroom at work is cold. • Tweeter observes many advertisements about McDonald's on television.
Excluded	• Tweeter wants to go home and feed a pet fish, as well as play a video game. • Tweeter states that they do not have a "beef" with people at another school. • Tweeter wants a sword to cut fruit with friends.

ing); −1 indicates unhealthy, high in at least 1 bad component and possibly containing 1 or more good components; 1 indicates healthy, no bad components and high in at least 1 good component (eg., healthy fats [unsaturated or omega-3], fiber, micronutrients, little processing); and 2 indicates definitively healthy, high in at least 2 good components and no bad ones. A second nutrition researcher (E.C.J.) independently classified a randomly selected sample of food words (n = 798, 20%) into these 4 categories. The Cohen's kappa for interrater reliability was 0.75 (95% confidence interval, 0.72-0.78]), indicating high agreement and demonstrating the robustness of our scores.

Sentiment vocabulary for food-related tweets

To study food-related attitudes (defined as evaluating an entity with a degree of positivity or negativity),[28] we assessed sentiment expressed in tweets that mentioned food. As described in Vydiswaran et al,[9] we expanded the positive and negative emotion sense category words from the Linguistic Inquiry and Word Count dictionary,[29] with food-specific sentiment words commonly used in social media from popular web pages, Yelp reviews, and a smaller set of tweets.

Finding relevant food-related tweets

We defined a relevant food-related tweet as one that "conveys information about the dietary choices that Twitter users make, including the specific foods they desire, foods they eat, how those foods are prepared, and where and when food is obtained and consumed." Additionally, given our neighborhood focus, tweets characterizing the "food environment" or availability of specific food in an area were included. "Availability" may refer to food's physical presence, price, freshness, and nutritional quality; a person's distance from and proximity to food establishments; locations of food stores, services, and other venues where food may be obtained (eg, food banks, homes, res-

collection, conducted in early 2016, yielded 21.19 million tweets from 2014 to 2016, authored by 120 748 unique tweeters.

Food vocabulary and keyword healthiness score

Tweets were mined for food-related terms. A vocabulary of 3928 terms was compiled from multiple online sources (see[9] for details). Box 1 summarizes the categories included. One author, a public health nutrition researcher (A.B.), assigned a "healthiness" score to each keyword using a 4-point system: −2 indicates definitively unhealthy, high in 2 or more bad dietary components (eg, unhealthy fats [saturated or trans], added sugar, sodium, substantial process-
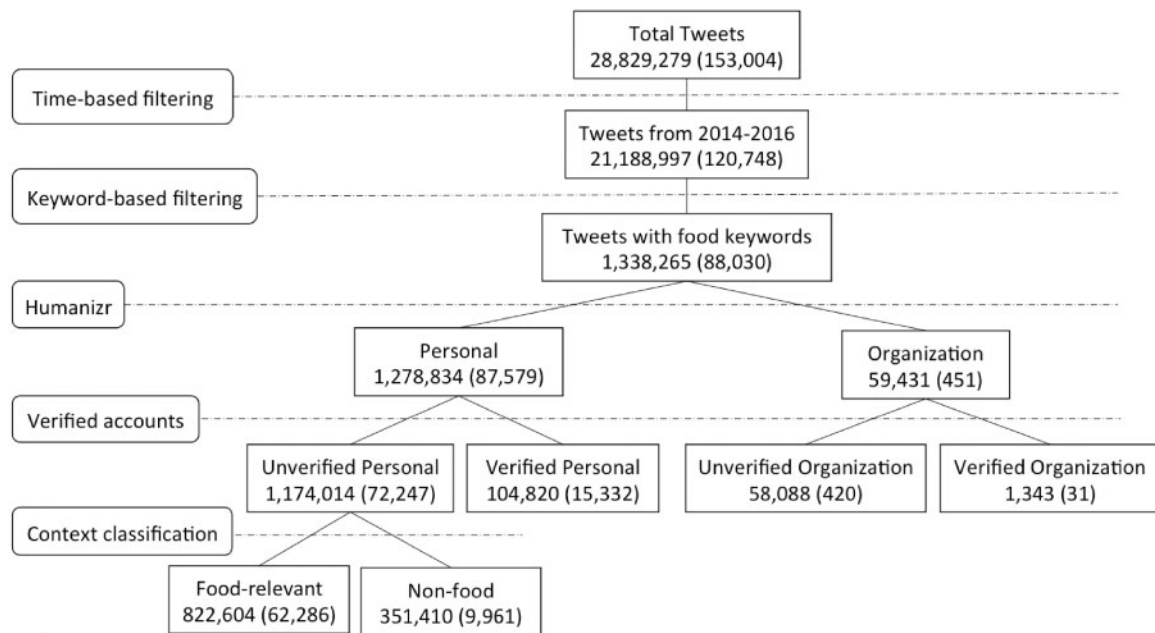
Figure 1. Number of tweets, with number of users in parentheses, filtered through multiple steps of analysis.

taurants, or foraging); a system that provides access (including delivery and "carry outs"); a state of food security or insecurity. Tweets directly or indirectly revealing any of these were included. Box 2 shows examples of types of included and excluded tweets.

Training a food-related tweet classifier

We designed a machine learning–based framework to distinguish food-related tweets from those using potentially food-related keywords in a nonfood sense. We refer readers to prior work[9] for a detailed description and evaluation of the tweet classifier. A hybrid model consisting of a food word filter followed by a support vector machine model achieved the highest performance (F1 of 0.858) among the tested classification algorithms. To estimate laypersons' attitudes toward food-related behavior, we removed tweets from businesses identified using the Humanizr tool,[30] and verified accounts from well-known personalities. The final classifier was applied to all remaining tweets containing at least 1 food keyword. Figure 1 shows the number of tweets filtered through this process.

"Healthiness" and sentiment scores for food-related tweets

A food-related tweet may mention both healthy and unhealthy items. Tweets were assigned a healthiness score based on the healthiness scores of the food keywords described previously, computed in 3 parts: the healthy score, the unhealthy score, and the net healthiness score. The (un)healthy score is the number of (un)healthy food words in the tweet scaled by the level of (un)healthiness per the food keyword list. Positive net healthiness scores indicate more mentions of healthy than unhealthy foods.

Similarly, the sentiment expressed is measured by counting food-related sentiment words, identified using the sentiment vocabulary. For each tweet, 3 sentiment scores are computed—a positive and a negative sentiment score are computed by normalizing the number of positive (negative) sentiment words with the total number of

tokens in the tweet. The overall sentiment score is the difference of positive and negative sentiment scores. A higher overall score means greater positive sentiment. For example, the tweet "I love bacon! !!!!" is 4 tokens long and has 1 positive and no negative sentiment words;. So, the positive score is $^1/_4 = 0.25$, negative score is 0, and overall sentiment score is 0.25.

Tract-level tweet-based measures

Tweet-based healthiness scores were aggregated at the user level based on the number of healthy and unhealthy food words in tweets a user authored (normalized by total number of words the user tweeted). User-level scores were aggregated at the census tract level as the average healthiness measure for the tract.

A similar normalization was performed on the sentiment measures. To account for the imbalance in the overall sentiment expressed in a given locality, the score was adjusted for baseline sentiment in a tract. Baseline scores were computed over 1000 randomly selected tweets per tract, and subtracted from the tweet sentiment scores.

Neighborhood characteristics measures

Neighborhood affluence and disadvantage scores

Following published measures,[31] aggregate measures of neighborhood affluence and disadvantage were created using demographic data from the American Community Survey (ACS) 2011-2015 estimates at the census tract level.[32] Based on prior work,[33] the neighborhood affluence score was created via factor analysis using (1) proportion with incomes >\$75 000 and (2) proportion with educational levels of an associate degree or higher (84.5% of the variance explained). Also building on prior research,[33,34] the neighborhood disadvantage score was calculated using the same method using the variables (1) proportion living in households with supplemental security income in the past 12 months and (2) proportion living below the federal poverty level (85.8% of the variance explained). These variables were transformed and standardized using z scores.

Fast food density

The fast food density measure assesses the spatial accessibility of fast food restaurants. We created a kernel density measure using ArcGIS's zonal statistics tool[35] with a radius setting of 1320 feet. Locations were gathered from the 2015 Reference USA Database.[36] Fast food vendors were selected based on their primary industrial classification codes (eg, North American Industry Classification System code "722211: Limited-service restaurants" and Standard Industrial Classification code "581222: Pizza stores"). Additionally, a list of U.S. chains compiled based on past research[37,38] and industry sources,[39–41] was used to query the Reference USA Database. The variable was transformed into a z score.

Percent young adult

ACS also lists percentages of people in certain age groups at the census tract level. Owing to the possible relationship between food-related tweets and the population using Twitter, we computed a measure for the proportion between 18 and 29 years of age in 2011-2015, corresponding to young adulthood. The variable was transformed into a z score.

Percent African American

ACS surveys numbers of people of specific races or ethnicities in each census tract. Owing to obesity-related health disparities among African Americans,[42] who experience pronounced disparities related to mortality for cardiovascular disease, diabetes, and kidney disease in the study region,[43] we created a tract-level measure for the proportion of the tract in 2011-2015 that were African American.

Mortality by cause of death

Using geocoded Michigan Department of Vital Statistics data concerning all deaths in the state from 2010 to 2014, we calculated tract-level mortality rates based on cause of death using International Classification of Diseases codes. We focused on causes known to be correlated with obesity and cardiometabolic syndrome, which are also related to diet.[44–50] These included diabetes (International Classification of Diseases codes: E10-E14), kidney failure (N17-N19), cardiovascular diseases such as hypertensive or ischemic heart disease (I11, I13, I20-I25), heart failure (I50), and cerebrovascular disease or stroke (I60-I69).[51] Previous literature has consistently found correlations between dietary patterns and mortality for these conditions.[48,52–63] We calculated mortality rates for each by dividing numbers of deaths by total tract population for that 5-year period.

Data analyses

We analyzed tweets generated from 2014 to 2016 to best match the time periods of the neighborhood characteristics measures. To address sparsity of food-related tweets in some census tracts, we restricted our sample to 80% of tracts with the largest numbers of tweets (n = 1273).

We conducted regression analyses of associations between food-related tweet healthiness and sentiment scores and neighborhood characteristics measures, and tested the following hypotheses: (1) Less affluent neighborhoods have more positive sentiment toward food. (2) Neighborhoods with more African Americans have more positive sentiment toward food. These hypotheses are based on prior research suggesting that attitudes provide partial explanations for poorer diets,[5,6] and that more positive sentiment about food correlates with neighborhood characteristics that have been elsewhere as-

sociated with poorer diets: fast food density, disadvantage, and percentage African American.[7,8]

For the sentiment associations, we conducted the analysis over all food-related tweets, and separately over healthy tweets (with healthy score greater than zero), and unhealthy tweets (with unhealthy score greater than zero). After testing for conformity to linear regression assumptions, we ran bivariate and multivariate regressions for tract-level healthiness scores (healthy, unhealthy, and net healthiness), sentiment scores (positive, negative, and overall), and causes of mortality as dependent variables. Independent variables were the 5 neighborhood measures (affluence, disadvantage, percentage African American, percentage young adult, and fast food density). Finally, total number of tweets in the census tract was included as a control variable to avoid the tweet volume driving the results. Figure 2 shows the correlation coefficients between the dependent and independent variables.

For variable selection, bivariate regressions were performed individually with 1 independent and 1 dependent variable. If the P value was <.15, the independent variable was included. A multivariate regression was performed with all included variables. In most multivariate regression analyses, both neighborhood affluence and disadvantage score indices were included. While the 2 are negatively correlated (Pearson's $\rho = -.845$, $P < .001$), they are not exact opposites. Both indices relate to income, but capture different aspects of economic well-being. The factors in the affluence measure—individual income and highest education level achieved—have been associated with diet quality.[64] Similarly, poverty and access to Supplemental Security Income (SSI) benefits—the factors in the disadvantaged measure—are known to be associated with obesity and dietary quality.[65,66]

Content analysis of tweets

We analyzed the top words driving differences between the most and least affluent tracts and performed content analysis to derive themes in food-related discussions.[67] We selected a stratified sample of 1759 tweets from the most and least affluent tracts that included at least 1 of the top 20 keywords driving differences in healthiness of food mentions between those tracts. After excluding retweets, promotional tweets, and tweets of unknown content (eg, only an image), 1537 remaining tweets (87.4%) were manually coded by one author (T.C.V.) using qualitative content analysis[68] to inductively identify themes. Newly identified themes were added until saturation was achieved.

Ethical considerations

Analyses using mortality data were approved by the University of Michigan Institutional Review Board. Twitter-based analyses using publicly available data do not conform to the university's definition of human subjects research. Nevertheless, because ethical concerns have been raised regarding directly quoting social media content,[69,70] we paraphrase rather than reproduce tweets verbatim.

RESULTS

In all, 822 604 tweets classified as food-related were authored by 62 286 laypersons in Metropolitan Detroit (also see Figure 1). The range per census tract was 102 to 16 330, with a mean of 367.6 ± 641.375 (interquartile range: 169-386).
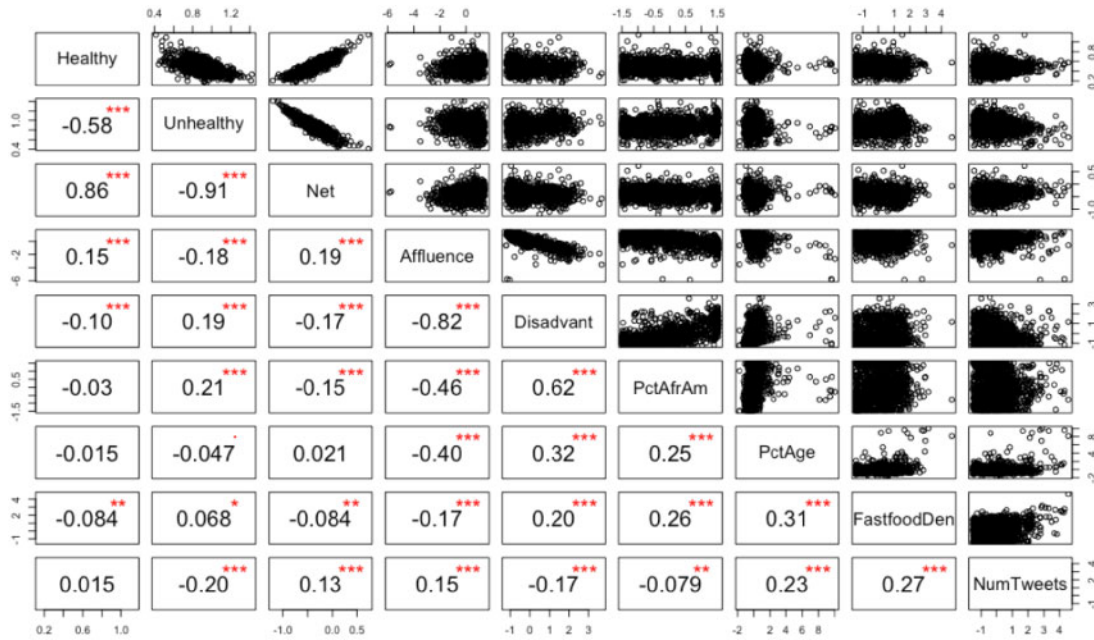
Figure 2. Correlations between independent and dependent variables in this study. The healthy, unhealthy, and net healthiness scores (first 3 variables) are dependent variables; the rest are predictors.

Table 1. Bivariate and multivariate regression analysis of the 3 healthiness scores against individual neighborhood characteristics measures.

Independent variable	Bivariate analysis			Multivariate analysis		
	Healthy	Unhealthy	Net	Healthy	Unhealthy	Net
Affluence index	.020 (.024)[c]	−.028 (.031)[c]	.048 (.035)[c]	.029[c]	−.021[b]	.041[c]
Disadvantage index	−.013 (.010)[c]	.029 (.036)[c]	−.042 (.028)[c]	.012	−.006	.012
% African American	−.004 (.001)	.032 (.046)[c]	−.036 (.022)[c]	—	.025[c]	−.017
% young adult	−.002 (.0002)	−.006 (.002)	.005 (.0004)	—	−.016[c]	—
Fast food density	−.010 (.007)[b]	.010 (.005)[a]	−.019 (.007)[b]	−.008[a]	.013[b]	−.019[b]
Number of tweets	.002 (.0002)	−.029 (.041)[c]	.031 (.018)[c]	—	−.025[c]	.031[c]
R^2				.030	.099	.057

For the bivariate regression analysis, the numbers show regression coefficients, with R^2 in parentheses. Variables included in the multivariate analysis are affluence index, disadvantage index, % African American, % young adult, fast food density, and number of tweets in the census tract.

[a]$P < .05$. [b]$P < .01$. [c]$P < .001$.

Regression analysis for healthiness scores

The bivariate regression analysis for healthiness scores (Table 1) shows that the affluence and disadvantage measures were significant for all 3 measures. The affluence measure was positively correlated with healthy and net healthiness scores, and negatively correlated with unhealthy scores. Conversely, the disadvantage measure was negatively correlated with healthy and net healthiness scores and positively correlated with unhealthy scores.

A higher proportion of African Americans in a tract was significantly associated with higher unhealthy score and lower net healthiness score. The multivariate analysis shows that the affluence score was consistently one of the most significant factors for the 3 healthiness scores—associated with higher healthy and net healthiness scores and lower unhealthy scores. Additionally, the contextual variable for U.S. density was significantly associated with both higher unhealthy scores and lower net healthiness scores. The best regression model for the net healthiness score ($R^2 = .057$) consists of afflu-

ence, fast food density, and the factor controlling for the number of tweets in a tract.

Regression analysis for sentiment scores

Table 2 shows the multivariate regression analysis for the sentiment scores. The best regression model for the overall sentiment score over all tweets ($R^2 = .174$) consists of affluence, disadvantage, race, and age—with positive association with higher disadvantage score and percent African American, and negative association with affluence score and age. Affluence, disadvantage, and race remain significant in the regression models for healthy and unhealthy tweets in the same direction as all tweets, while age is significant only for unhealthy tweets, again in the same direction as all tweets.

Analyzing the individual neighborhood characteristics measure and accounting for other independent variables, affluence index was inversely correlated with overall sentiment score

Table 2. Multivariate regression analysis of overall sentiment score for healthy, unhealthy, and all tweets against neighborhood measures

Independent variable	Healthy tweets	Unhealthy tweets	All tweets
Affluence index	−.020[a]	−.022[c]	−.021[c]
Disadvantage index	.019[a]	.017[a]	.017[b]
% African American	.023[c]	.028[c]	.027[c]
% young adult	–	−.015[c]	−.016[c]
Fast food density	.003	−.003	.001
Number of tweets	–	.003	.002
R^2	.101	.144	.174

[a]$P < .05$. [b]$P < .01$. [c]$P < .001$.

(significant at $P < .001$), indicating that less affluent neighborhoods tend to have more positive sentiment toward food. Further, percent African American was positively correlated with the overall sentiment toward food, and toward both healthy and unhealthy food (all significant at $P < .001$), indicating that after controlling for affluence, neighborhoods with more African Americans have more positive sentiment toward food. The direction of the correlation and the relative significance of the correlation analysis do not change even when either the disadvantage or affluence measure is excluded. Adjusting for baseline tract-level sentiment, we found higher levels of positive sentiment in more affluent tracts and lower levels in disadvantaged neighborhoods (not shown).

Regression analysis of causes of mortality

Table 3 shows the multivariate regression analysis against 5 obesity-related causes of mortality: diabetes, kidney failure, heart failure, stroke, and hypertensive or ischemic heart disease. After adjusting for other neighborhood characteristics and correcting for multiple hypothesis testing using Holm-Bonferroni method to control the familywise error rate,[71] there was a significant correlation between Twitter-based net healthiness scores and the mortality rate from heart failure ($P < .05$), and the relationship with kidney failure mortality approached significance ($P < .1$). Mortality rates from diabetes and other obesity-related cardiovascular conditions, including stroke and hypertensive or ischemic heart disease, had no significant relationships with Twitter-based measures.

Content analysis of tweets
Food words driving differences between tracts
To characterize food-specific disparities among neighborhoods, we compared relative keyword frequencies in the top and bottom quintiles (20%) of tracts on affluence score. Box 3 lists the top 15 keywords driving differences in the net healthiness scores in the most and least affluent tracts. Keywords more common in less affluent tracts were higher in saturated fat than were those in more affluent tracts, and more affluent tracts included more mentions of fruits and vegetables.

Key themes in food-related tweets
Table 4 summarizes the key themes, proportions of theme-relevant tweets, and paraphrased examples. Nine food-related themes were identified and grouped into 3 main themes: (1) activities such as preparing and consuming food; (2) positive and negative attitudes, such as affection or dislike toward specific food or restaurants, and strug-

gling with food; and (3) food-vending locations. The most frequent themes were mentions of eating or drinking (behavior; 28.9% tweets), affection toward food or food establishments (positive attitude; 11.3%), and restaurants where food was consumed (location; 9.5%).

DISCUSSION

Offline research shows associations between neighborhood characteristics and dietary patterns, with greater socioeconomic disadvantage and fast food density (a proxy for exposure) being associated with consuming more saturated fat[72] and fewer fruits and vegetables.[73] Twitter discussion of healthy and unhealthy foods was correlated with precisely those neighborhood measures. Moreover, our analyses of keywords driving differences between healthiness scores showed more mentions of fruits and vegetables in more affluent tracts and more mention of foods high in saturated fats in less affluent tracts, again mirroring prior research. Results were further validated through qualitative analysis showing that large proportions of tweets mentioned food-related behavior, and locations where food and drink were consumed (eg, Starbucks, McDonalds), from which dietary aspects can be inferred.

Food tweet healthiness was significantly associated with the "hard" health outcome of mortality associated with one obesity-linked condition (heart failure) and marginally associated with a second (kidney failure). While these analyses are preliminary and based on a small sample of tracts, the latter result is suggestive of a trend. However, associations were not found for several other obesity- and cardiometabolic disease–related causes of death. Providing some resonance with this pattern, negative heart failure, diabetes, and kidney failure outcomes have been related to the neighborhood food environment,[58–60] whereas evidence for other cardiovascular diseases and stroke is more ambiguous.[61–63] Although there is a need for further research, perhaps in larger geographic areas, these results align with and extend prior findings on associations between health conditions and food-related tweets at state and zip code scales.[12,13] While larger than census tracts, zip codes are constructed primarily for mail delivery and their residents' characteristics vary significantly. Census tracts, being smaller and more spatially contiguous and demographically homogeneous, reveal more about local food environments. Further, using smaller spatial units can minimize aggregation problems such as the ecological fallacy or the modifiable areal unit problem.[74] Thus, for our study, census tracts provide the right spatial granularity.

While social media data are potential sources for exposure-based population health management studies, boundaries imposed by census tracts may differ from residents' experiences. An increasing body of spatial statistical literature aims to define spatial boundaries based on data. Approaches such as "Bayesian wombling" analyze data at a fine spatial aggregation level (eg, census tracts or census blocks) and probabilistically determine whether boundaries among contiguous areal units are warranted by the data.[75–77] Such research is beyond the scope of our work.

Our analysis revealed significant associations between sentiment concerning healthy and unhealthy food. Overall sentiment in healthy and unhealthy food tweets is positively correlated with neighborhood characteristics that have been associated with poorer diets: disadvantage, and percent African American (significant at $P < .05$ for disadvantage and $P < .001$ for percent African American percentage for both healthy and unhealthy foods). This supports existing research on attitudes toward food as a correlate of

Table 3. Multivariate regression analysis of 5 obesity-related causes of mortality against the net healthiness score and neighborhood measures

Independent variables	Diabetes	Kidney failure	Heart failure	Stroke	Hypertensive or Ischemic heart disease
Affluence index	-7.4×10^{-4d}	-3.6×10^{-4d}	—	-7×10^{-4c}	$-.003^{d}$
Disadvantage index	-4.6×10^{-4d}	-3.0×10^{-4d}	-1.0×10^{-4b}	—	$-.001^{d}$
% African American	2.1×10^{-4d}	1.9×10^{-4d}	—		$.001^{d}$
% young adult	—	—	-2.1×10^{-4d}	$-.001^{d}$	$-.001^{d}$
Fast food density	9.5×10^{-5c}	5.8×10^{-5c}	2.0×10^{-4d}	5×10^{-4b}	$.001^{d}$
Number of tweets	1.5×10^{-4d}	7.5×10^{-5d}	1.8×10^{-4d}	—	-1×10^{-4}
Net healthiness score	1.8×10^{-4}	-1.7×10^{-4a}	-3.9×10^{-4b}	—	8×10^{-4}
R^2	.151	.102	.061	.0151	.2339

$^{a}P < .1.$ $^{b}P < .05.$ $^{c}P < .01.$ $^{d}P < .001.$

Box 3 Top 15 keywords driving the most differences in the net healthiness scores.

Top keywords from the most affluent tracts:
starbucks, coffee, sushi, pumpkin, cherry, vegan, tea, apple, chipotle, oil, coney island, orange, donuts, turkey, chocolate

Top keywords from the least affluent tracts:
pizza, grill, taco, cake, mcdonalds, fries, honey, cream, bacon, ice cream, steak, fried, subway, fish, cookie

eating behavior. Adjusting for other factors related to poorer diet, the sentiment score for unhealthy tweets is also negatively correlated with percentage of young adults (significant at $P < .001$). This suggests that, adjusting for other factors, residents in tracts with higher proportions of young adults express less positivity about unhealthy foods. The sentiment scores are not significantly correlated with the fast food density measure. In contrast, tweets from more affluent tracts expressed less positive sentiment about food than about other topics. While this may in part be explained by overall differences in sentiment between tracts of different affluence levels, Twitter data reveal an intriguing hypothesis: food may be an isolated source of enjoyment in otherwise difficult lives.

Qualitative analyses demonstrated that positive attitudes included cravings and general evaluation of a food or establishment, which may predict future consumption. They also expose emotional evaluations of food that comprise expressed attitudes in a more fine-grained manner. Negative sentiment revealed issues such as personal struggles with overeating. Therefore, Twitter data may provide information regarding eating behavior and factors that drive it. Moreover, analyses contrasting more and less affluent tracts provide a more local view of issues in communities experiencing food-related health disparities, including neighborhood differences regarding dietary behavior. Thus, food-related tweets could assist local health programs focused on consumption and inform interventions focused on attitudes and the environment. These results also suggest that Twitter could be used to create food consumption–based phenotypes and enrich the field of behavioral phenotyping in nutrition.[78]

Public health research concerning population-level health behaviors and attitudes is typically conducted through surveys via ongoing initiatives such as the Behavioral Risk Factor Surveillance System and National Health and Nutrition Examination Survey, conducted with smaller groups than can be found on social media, and more costly to implement. Larger-scale, continuous, and less resource-intensive data collection methods are desirable. While developing the food keyword healthiness score and training set for tweet classification models required significant manual effort, the resultant classifiers can be applied to a large set of unlabeled tweets. Our results suggest that social media data can provide a reliable signal for dietary patterns and food-related attitudes at the census tract level despite the noisy nature of user-generated text data, the limited fraction of geolocated tweets, and access only to public discussions rather than actual dietary patterns. However, survey data can more accurately determine the demographic characteristics of social media users and overcome likely bias due to differences between social media users and nonusers. While biases in survey-driven estimates can be adjusted with the knowledge of sampling bias, new methods are needed to correct for sampling biases when using social media data.

Limitations

Our study relies on mentions of healthy and unhealthy foods in tweets, and sentiment expressed about them. While mentioning food words alone does not indicate actual consumption of those foods, we believe that it represents information shared in the community around food, which may reflect attitudes, and ultimately, behavior. This measure is also incomplete; it does not reflect the amount of food consumed. Further, many foods are not inherently completely unhealthy or healthy. Although the classification scheme was validated independently by 2 nutrition researchers, with Cohen's kappa for interrater reliability of 0.75, nutritional science is complex and still evolving. The food keyword healthiness rating reflects our understanding of a typical portion of the food in terms of unhealthy fats, added sugar, sodium, and amount of processing but does not account for the actual caloric or nutrient content.

There are potential biases introduced from using social media for neighborhood and community-based studies. Specifically, rural communities and senior populations are typically underrepresented. Further, tweeting behavior varies significantly with access to broadband, across urban-rural divide and on racial dimensions. While we account for some of these factors (eg, percent African American, percent young adult, number of tweets from a tract) and normalized the measures by aggregating at both user and tract levels, this may be insufficient to overcome systematic biases in the social media data. Finally, our multivariate correlation analysis was limited to

Table 4. Content analysis of tweets, showing behaviors and POS and NEG attitudes toward food and food locations, with paraphrased example tweets

Theme	Count	Example tweets
Behavior: • eating or drinking	445 (28.9)	• Tweeter states that they have eaten a lot of bacon, which made for a good morning. • Tweeter checks into a Chipotle restaurant to eat with friends. • Tweeter says that they cannot stop eating Gala apples they like. • Tweeter refers to engaging in personal reflection while eating large quantities of sushi.
Behavior: cooking or preparing food	98 (6.4)	• Tweeter says they are going to make pancakes and bacon. • Tweeter posts a photo of ribs that they have been grilling.
Attitudes: POS: affection for food or food establishment	177 (11.3)	• Tweeter says the word "bacon" repeatedly. • Tweeter states that pizza is an aphrodisiac. • Tweeter enthuses about flavored iced coffee.
Attitudes: POS: craving	101 (6.4)	• Tweeter thinks a plateful of bacon sounds good. • Tweeter announces an urgent craving for sushi.
Attitudes: POS: enjoying food and drink	127 (8.3)	• Tweeter states that bacon makes everything better, and wishes for a bacon emoji. • Tweeter checks into a restaurant and claims it has the best French fries.
Attitudes: NEG: dislike for food or food establishment	26 (1.6)	• Tweeter felt nauseous after smelling food from a restaurant. • Tweeter says a restaurant's coffee is disgusting.
Attitudes: NEG: struggles with food (overeating, discomfort after eating)	19 (1.2)	• Tweeter laments drinking coffee again after trying to quit and declares an addiction. • Tweeter has eaten a hash brown, saying that they are cheating. • Tweeter feels bad and body conscious for eating pizza while a friend works out at the gym.
Locations: • coffee shops	86 (5.5)	• Tweeter complains that the staff working in the coffee shop added whipped cream to their drink.
Locations: restaurants	150 (9.5)	• Tweeter debated going to Taco Bell, went, and ate a number of tacos; had regrets.

Values are n (%).
NEG: negative; POS: positive.

census tracts from metropolitan Detroit (n = 1273). Validation over a larger area is needed.

The bivariate and multivariate analyses revealed significant but weak correlations between tweet-based healthiness and sentiment scores and neighborhood characteristics. Further, the variance in tweet-based scores is not sufficiently explained by the neighborhood characteristics measures alone, suggesting that additional factors such as cultural preferences, linguistic variations, and personal interactions outside of Twitter may not be captured in tweets.

Our procedure for extracting food-related social media content at the census tract level and identifying the healthiness of tweets and associated sentiment can be applied to other social media and other geographical locations. Future research should examine this procedure in larger areas and seek to validate Twitter measures through direct measures of food consumption, perhaps partnering with surveillance-based surveys to investigate representative samples of populations at the census tract level.

CONCLUSION

Results provide an initial step toward utilizing social media to track food-related health behaviors and attitudes at a large scale but at localized geographical levels that reveal differences between communities that, despite proximity, may exhibit significant variation along environmental, socioeconomic, demographic, and cultural dimensions.

FUNDING

AUTHOR CONTRIBUTIONS

CONFLICT OF INTEREST STATEMENT

None declared.

REFERENCES

1. Diez Roux AV, Mair C. Neighborhoods and health. *Ann N Y Acad Sci* 2010; 1186 (1): 125–45.
2. Mozaffarian D. Dietary and policy priorities for cardiovascular disease, diabetes, and obesity. *Circulation* 2016; 133 (2): 187–225.
3. Kendall L, Hartzler A, Klasnja P, *et al*. Descriptive analysis of physical activity conversations on Twitter. In: *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. New York, NY: ACM; 2011: 1555–60.
4. Sarker A, Ginn R, Nikfarjam A, *et al*. Utilizing social media data for pharmacovigilance: a review. *J Biomed Inform* 2015; 54: 202–12.

5. Verplanken B, Hofstee G, Janssen HJ. Accessibility of affective versus cognitive components of attitudes. *Eur J Soc Psychol* 1998; 28 (1): 23–35.

6. Ajzen I. Nature and operation of attitudes. *Annu Rev Psychol* 2001; 52 (1): 27–58.

7. Godin G, Kok G. The theory of planned behavior: a review of its applications to health-related behaviors. *Am J Health Promot* 1996; 11 (2): 87–98.

8. Erickson BH. The relational basis of attitudes. The relational basis of attitudes. In Wellman B, Berkowitz SD, eds. *Structural Analysis in the Social Sciences, Vol. 2. Social Structures: A Network Approach*. New York, NY: Cambridge University Press; 1988: 99–121.

9. Vydiswaran VGV, Romero DM, Zhao X, et al. "Bacon bacon bacon": food-related tweets and sentiment in metro Detroit. In: proceedings of the 12th International AAAI Conference on Weblogs and Social Media (ICWSM); 2018: 692–5.

10. U.S. Department of Health and Human Services. Healthy People 2020. ODPHP Publication No. B0132; 2010. https://www.healthypeople.gov/2020/About-Healthy-People Accessed October 12, 2018.

11. Institute of Medicine. *For the Public's Health: The Role of Measurement in Action and Accountability*. Washington, DC: National Academies Press; 2011.

12. Abbar S, Mejova Y, Weber I. You tweet what you eat: studying food consumption through Twitter. In proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI). New York, NY: ACM; 2015: 3197–206.

13. Nguyen QC, Brunisholz KD, Yu W, et al. Twitter-derived neighborhood characteristics associated with obesity and diabetes. *Sci Rep* 2017; 7 (1): 16425.

14. Krieger N, Waterman P, Chen JT, et al. Zip code caveat: bias due to spatiotemporal mismatches between zip codes and US census defined geographic areas-the public health disparities geocoding project. *Am J Public Health* 2002; 92 (7): 1100–2.

15. U.S. Census Bureau. Geographic Terms and Concepts-Census Tract; 2018. https://www.census.gov/geo/reference/gtc/gtc_ct.html Accessed October 12, 2018.

16. U.S. Department of Agriculture Economic Research Service. Food Access Research Atlas; 2017. https://data.nal.usda.gov/dataset/food-access-research-atlas Accessed March 20, 2019.

17. Nguyen QC, Kath S, Meng HW, et al. Leveraging geotagged twitter data to examine neighborhood happiness, diet, and physical activity. *Appl Geogr* 2016; 73: 77–88.

18. De Choudhury M, Sharma S, Kiciman E. Characterizing dietary choices, nutrition, and language in food deserts via social media. In: *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. New York, NY: ACM; 2016: 1157–70.

19. Drewnowski A. Concept of a nutritious food: toward a nutrient density score. *Am J Clin Nutr* 2005; 82 (4): 721–32.

20. Au LE, Whaley S, Rosen NJ, et al. Online and in-person nutrition education improves breakfast knowledge, attitudes, and behaviors: a randomized trial of participants. *J Acad Nutr Diet* 2016; 116 (3): 490–500.

21. McCormack LA, Laska MN, Larson NI, et al. Review of the nutritional implications of farmers' markets and community gardens: a call for evaluation and research efforts. *J Am Diet Assoc* 2010; 110 (3): 399–408.

22. Detroit Food Policy Council. https://detroitfoodpolicycouncil.net/ Accessed December 21, 2018.

23. Washtenaw County Food Policy Council. https://washtenawfoodpolicycouncil.wordpress.com/ Accessed December 21, 2018.

24. Food Policy Council. Energizing Connections for Healthier Oakland (ECHO). https://www.oakgov.com/health/echo/Pages/Food-Policy-Council.aspx Accessed December 21, 2018.

25. Schuch S. Genesee County's SPROUT Coalition encourages healthy eating during National Childhood Obesity Awareness Month. https://www.mlive.com/news/flint/index.ssf/2012/09/genesee_countys_sprout_coaliti.html Accessed December 21, 2018.

26. Food and Health. MSU Extension. Michigan State University. https://www.canr.msu.edu/food_health/ Accessed December 21, 2018.

27. Detroit Food Map. http://detroitfoodmap.com/ Accessed December 21, 2018.

28. Glanz K, Rimer BK, Viswanath K. *Health Behavior: Theory, Research, and Practice*. 5th ed. San Francisco, CA: Jossey-Bass; 2015.

29. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. The development and psychometric properties of LIWC2015. Technical Report; 2015. https://repositories.lib.utexas.edu/bitstream/handle/2152/31333/LIWC2015_LanguageManual.pdf Accessed December 21, 2018.

30. McCorriston J, Jurgens D, Ruths D. Organizations are users too: characterizing and detecting the presence of organizations on Twitter. In: *Proceedings of the 9th International AAAI Conference on Weblogs and Social Media (ICWSM)*. New York, NY: ACM; 2015: 650–3.

31. Clarke P, Morenoff J, Debbink M, et al. Cumulative exposure to neighborhood context: consequences for health transitions over the adult life course. *Res Aging* 2014; 36 (1): 115–42.

32. U.S. Census Bureau. American Community Survey (ACS) 5 year public use microdata sample (PUMS), 2011–2015; 2011. https://www.census.gov/programs-surveys/acs/data/pums.html Accessed October 12, 2018.

33. Weden MM, Carpiano RM, Robert SA. Subjective and objective neighborhood characteristics and adult health. *Soc Sci Med* 2008; 66 (6): 1256–70.

34. Massey DS. The age of extremes: concentrated affluence and poverty in the twenty-first century. *Demography* 1996; 33 (4): 395–412.

35. Environmental Systems Research Institute. *ArcGIS: Release 10.1*. Redlands, CA: Environmental Systems Research Institute; 2012.

36. Reference USA Database. http://www.referenceusa.com/ Accessed January 8, 2019.

37. Dubowitz T, Ghosh-Dastidar B, Eibner C, et al. The women's health initiative: the food environment, neighborhood socioeconomic status, BMI, and blood pressure. *Obesity (Silver Spring)* 2012; 20 (4): 862–71.

38. Inagami S, Cohen DA, Brown AF, et al. Body mass index, neighborhood fast food and restaurant concentration, and car ownership. Appendix A. *J Urban Health* 2009; 86 (5): 683–95.

39. FastFoodSource. Fast Food by Cuisine: Burgers and Sandwiches, Chicken, Mexican and International, Fish and Seafood, and Dessert. http://www.fastfoodsource.com/index.php/list-all-restaurants Accessed December 21, 2018.

40. Fast Food in USA. http://fastfoodinusa.com/category/fast-food-brands/ Accessed December 21, 2018.

41. McConnell A, Bhasin K. Ranked: the most popular fast food restaurants in America. *Business Insider*. http://www.businessinsider.com/the-most-popular-fast-food-restaurants-in-america-2012-7? op=1 Accessed December 21, 2018.

42. Abraham PA, Kazman JB, Zeno SA, et al. Obesity and African Americans: physiologic and behavioral pathways. *ISRN Obesity* 2013; 2013: 314295.

43. Michigan Department of Community Health. Health Disparities Reduction and Minority Health Section. Mortality Rates by Race and Ethnicity: Michigan, 2000–2010. https://www.michigan.gov/documents/mdch/Mortality_Rates__Race__Ethnicity_431194_7.pptx Accessed April 4, 2019.

44. Savica V, Bellinghieri G, Kopple JD. The effect of nutrition on blood pressure. *Annu Rev Nutr* 2010; 30: 365–401.

45. Elliott P, Stamler J, Nichols R, et al. Intersalt revisited: further analyses of 24 hour sodium excretion and blood pressure within and across populations. Intersalt Cooperative Research Group. *BMJ* 1996; 312 (7041): 1249–53.

46. Singh K, Betensky RA, Wright A, Curhan GC, Bates DW, Waikar SS. A concept-wide association study of clinical notes to discover new predictors of kidney failure. *Clin J Am Soc Nephrol* 2016; 11 (12): 2150–8.

47. Levy D, Larson MG, Vasan RS, Kannel WB, Ho K. The progression from hypertension to congestive heart failure. *JAMA* 1996; 275 (20): 1557–62.

48. Micha R, Peñalvo JL, Cudhea F, Imamura F, Rehm CD, Mozaffarian D. Association between dietary factors and mortality from heart disease, stroke, and type 2 diabetes in the United States. *JAMA* 2017; 317 (9): 912–24.

49. Murray CJ, Abraham J, Ali MK, et al. The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. *JAMA* 2013; 310 (6): 591–606.

50. Danaei G, Ding EL, Mozaffarian D, *et al*. The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PLoS Med* 2009; 6 (4): e1000058.

51. Flegal KM, Graubard BI, Williamson DF, Gail MH. Cause-specific excess deaths associated with underweight, overweight, and obesity. *JAMA* 2007; 298 (17): 2028–37.

52. Gutiérrez OM, Muntner P, Rizk DV, *et al*. Dietary patterns and risk of death and progression to ESRD in individuals with CKD: a cohort study. *Am J Kidney Dis* 2014; 64 (2): 204–13.

53. Mc Causland FR, Waikar SS, Brunelli SM. Increased dietary sodium is independently associated with greater mortality among prevalent hemodialysis patients. *Kidney Int* 2012; 82 (2): 204–11.

54. Levitan EB, Lewis CE, Tinker LF, *et al*. Mediterranean and DASH diet scores and mortality in women with heart failure: The Women's Health Initiative. *Circ Heart Fail* 2013; 6 (6): 1116–23.

55. Huang X, Jiménez-Moleón JJ, Lindholm B, *et al*. Mediterranean diet, kidney function, and mortality in men with CKD. *Clin J Am Soc Nephrol* 2013; 8 (9): 1548–55.

56. Kelly JT, Palmer SC, Wai SN, *et al*. Healthy dietary patterns and risk of mortality and ESRD in CKD: a meta-analysis of cohort studies. *Clin J Am Soc Nephrol* 2017; 12 (2): 272–9.

57. Saglimbene VM, Wong G, Ruospo M, *et al*. Fruit and vegetable intake and mortality in adults undergoing maintenance hemodialysis. *Clin J Am Soc Nephrol* 2019; 14 (2): 250–60.

58. Morris AA, McAllister P, Grant A, *et al*. Relation of living in a "food desert" to recurrent hospitalizations in patients with heart failure. *Am J Cardiol* 2019; 123 (2): 291–6.

59. Phillips AZ, Rodriguez HP. Adults with diabetes residing in "food swamps" have higher hospitalization rates. *Health Serv Res* 2019; 54 Suppl 1: 217–25.

60. Gutiérrez OM. Contextual poverty, nutrition, and chronic kidney disease. *Adv Chronic Kidney Dis* 2015; 22 (1): 31–8.

61. Kelli HM, Hammadah M, Ahmed H, *et al*. Association between living in food deserts and cardiovascular risk. *Circ Cardiovasc Qual Outcomes* 2017; 10 (9): e003532.

62. Daniel M, Paquet C, Auger N, Zang G, Kestens Y. Association of fast-food restaurant and fruit and vegetable store densities with cardiovascular mortality in a metropolitan population. *Eur J Epidemiol* 2010; 25 (10): 711–9.

63. Suarez JJ, Isakova T, Anderson CA, Boulware LE, Wolf M, Scialla JJ. Food access, chronic kidney disease, and hypertension in the US. *Am J Prev Med* 2015; 49 (6): 912–20.

64. Hiza HAB, Casavale KO, Guenther PM, Davis CA. Diet quality of Americans differs by age, sex, race/ethnicity, income, and education level. *J Acad Nutr Diet* 2013; 113 (2): 297–306. DOI: 10.1016/j.jand.2012.08.011

65. Leung CW, Villamor E. Is participation in food and income assistance programmes associated with obesity in California adults? Results from a state-wide survey. *Public Health Nutr* 2011; 14 (4): 645–52.

66. Andreyeva T, Tripp AS, Schwartz MB. Dietary quality of Americans by Supplemental Nutrition Assistance Program participation status: a systematic review. *Am J Prev Med* 2015; 49 (4): 594–604.

67. Krippendorff KCA. *An Introduction to Its Methodology*. 2nd ed. Thousand Oaks: Sage; 2004.

68. Schreier M. *Qualitative Content Analysis in Practice*. Thousand Oaks: Sage; 2012.

69. Fiesler C, Proferes N. "Participant" perceptions of Twitter research ethics. *Soc Media Soc* 2018; 4 (1):1–14.

70. Ayers JW, Caputi TL, Nebeker C, *et al*. Don't quote me: reverse identification of research participants in social media studies. *npj Digit Med* 2018; 1 (1): 30. https://doi.org/10.1038/s41746-018-0036-2.

71. Holm S. A simple sequentially rejective multiple test procedure. *Scand Stat Theory Appl* 1979; 6 (2): 65–70. http://www.jstor.org/stable/4615733 Accessed March 28, 2019.

72. Moore LV, Roux AVD, Nettleton JA, *et al*. Fast-food consumption, diet quality, and neighborhood exposure to fast food: the multi-ethnic study of atherosclerosis. *Am J Epidemiol* 2009; 170 (1): 29–36.

73. Dubowitz T, Heron M, Bird CE, *et al*. Neighborhood socioeconomic status and fruit and vegetable intake among whites, blacks, and Mexican-Americans in the United States. *Am J Clin Nutr* 2008; 87 (6): 1883–91.

74. Fotheringham AS, Wong D. The modifiable areal unit problem in multi-variate statistical analysis. *Environ Plan A* 1991; 23 (7): 1025–44.

75. Gelfand AE, Banerjee S. Bayesian wombling: finding rapid changes in spatial maps. *WIREs Comput Stat* 2015; 7 (5): 307–15.

76. Lu H, Reilly CS, Banerjee S, Carlin BP. Bayesian areal wombling via adjacency modeling. *Environ Ecol Stat* 2007; 14 (4): 433–52.

77. Ma H, Carlin BP, Banerjee S. Hierarchical and joint site-edge methods for medicare hospice service region boundary analysis. *Biometrics* 2010; 66 (2): 355–64.

78. Burgermaster M, Rodriguez V, Mamykina L. Data-driven psychosocial phenotyping for precision behavioral nutrition. *J Nutr Educ Behav* 2018; 50 (7): S170–1.