Article

Patterns

Diet during the COVID-19 pandemic: An analysis of Twitter data

Highlights

- We used Twitter data to quantify self-reported diet trends during the COVID-19 pandemic
- Healthy food consumption increased during the pandemic; alcohol consumption decreased
- Proximity to grocery stores and more time at home were associated with healthier diet
- Proximity to liquor stores corresponded with increased alcohol consumption

Authors

Mark A. Hernandez, Shagun Modi, Kanisha Mittal, Pallavi Dwivedi, Quynh C. Nguyen, Nina L. Cesare, Elaine O. Nsoesie

Correspondence

mark.hernandez@ll.mit.edu

In brief

This study finds that staying at home and access to grocery stores may have positively impacted individuals' diets during the COVID-19 pandemic. However, access to liquor stores and staying at home may have facilitated increased alcohol consumption among those in proximity to liquor retailers.

Hernandez et al., 2022, Patterns *3*, 100547
August 12, 2022
https://doi.org/10.1016/j.patter.2022.100547

CellPress

Patterns

CellPress
OPEN ACCESS

Article

Diet during the COVID-19 pandemic: An analysis of Twitter data

Mark A. Hernandez,[1,2,5,*] Shagun Modi,[1] Kanisha Mittal,[1] Pallavi Dwivedi,[3] Quynh C. Nguyen,[3] Nina L. Cesare,[1,4] and Elaine O. Nsoesie[1,4]
[1]Boston University School of Public Health, Boston, MA 02118, USA
[2]Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, MA 02421, USA
[3]Department of Epidemiology and Biostatistics, University of Maryland School of Public Health, College Park, MD 20740, USA
[4]These authors contributed equally
[5]Lead contact
*Correspondence: mark.hernandez@ll.mit.edu
https://doi.org/10.1016/j.patter.2022.100547

THE BIGGER PICTURE The COVID-19 pandemic upended many aspects of daily life, including how we eat and drink. Restaurant closures and retail restrictions likely impacted individuals' consumption habits, but longitudinal surveys that monitor nutrition and/or alcohol intake are costly to administer and are prone to response bias. In this study, we use digital trace data from Twitter to track population-level patterns in nutritional intake. Linking geotagged tweets to data measuring US county characteristics and built environment, this study finds that increased time at home and access to grocery stores during the pandemic may have promoted healthy-food consumption. This study also suggests that access to alcohol retail establishments may have led to more drinking. These findings validate the importance of the built environment to health behaviors while highlighting how social media data may be used to assess the impact of public health crises.

1 2 3 4 5 **Development/Pre-production:** Data science output has been rolled out/validated across multiple domains/problems

SUMMARY

In this study, we measured the association between county characteristics and changes in healthy-food, fast-food, and alcohol tweets during the COVID-19 pandemic in the United States. Our analytic dataset consisted of 1,282,316 geotagged tweets that referenced food consumption posted before (63.2%) and during (36.8%) the pandemic and included all US states. We found the share of healthy-food tweets increased by 20.5% during the pandemic compared with pre-pandemic, while fast-food and alcohol tweets decreased by 9.4% and 11.4%, respectively. We also observed that time spent at home and more grocery stores per capita were associated with increased odds of healthy-food tweets and decreased odds of fast-food tweets. More liquor stores per capita was associated with increased odds of alcohol tweets. Our results highlight the potential impact of the pandemic on nutrition and alcohol consumption and the association between the built environment and health behaviors.

INTRODUCTION

The COVID-19 pandemic disrupted how individuals interacted with food establishments in the United States. Stay-at-home orders and closures across the restaurant and bar industries changed how people accessed and consumed food and alcohol. Several studies have shown shifts in population-level diet and alcohol consumption in the wake of the pandemic.[1–7] Three trends have emerged from the existing literature. First, studies suggest that the consumption of fast food decreased during the pandemic in the United States.[1–3] Second, studies suggest that consumption of alcohol increased during the pandemic in the United States.[2–5] Third, findings regarding the changes in healthy food consumption were inconsistent or limited.[1,6]

However, a major limitation of most studies is a reliance on self-reported past food and alcohol consumption from traditional surveys, which might be impacted by recall bias.[8–11] Social media data can provide similar and other advantages to traditional survey data. For example, data from Twitter can comprise a large sample covering the full geographic extent of the United States.

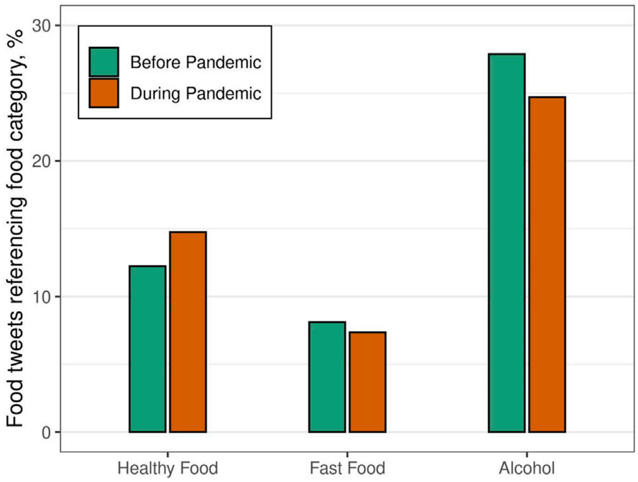

Figure 1. Changes in healthy-food, fast-food, and alcohol tweets during the COVID-19 pandemic in the United States

Geotagged food tweets were extracted from 1% of random tweets collected during the pre-pandemic period (May 15, 2019 to January 31, 2020) and the pandemic period (May 15, 2020 to January 31, 2021) using the Twitter streaming application programming interface. n = 811,050 geotagged food tweets during pre-pandemic period. n = 471,266 geotagged food tweets during pandemic period.

Also, tweets are collected continuously before and after events of interest. Tweets also contain voluntary information about diet and alcohol consumption, allowing for the natural observation of attitudes and behaviors related to food and alcohol consumption. Several studies have shown the feasibility and utility of leveraging social media to track population trends in diet and alcohol consumption.[12–17]

The objective of our study was to leverage Twitter data to address the gaps in understanding changes in diet and alcohol consumption during the COVID-19 pandemic. We examined geotagged tweets in the United States that referenced food terms collected during two periods: before and during the pandemic (May 15, 2019 to January 31, 2020 and May 15, 2020 to January 31, 2021) to quantify how references to the consumption of healthy food, fast food, and alcohol changed during the pandemic compared with before the pandemic. We also linked the geotagged tweets to the US county of origin to assess the relationship between county characteristics and changes in the share of healthy-food, fast-food, and alcohol tweets. This study adds to our understanding of how public health interventions and other changes that occurred during the pandemic could have impacted health behaviors, specifically, the consumption of healthy food, fast food, and alcohol. It also shows associations between county built environment characteristics and health behaviors and how this relationship may have changed during the pandemic.

RESULTS

Data summary

Our initial dataset consisted of 11,445,868 tweets referencing food terms during the pre-pandemic period (6,270,184 tweets from May 15, 2019 to January 31, 2020) and pandemic period (5,175,684 tweets from May 15, 2020 to January 31, 2021). Of the initial dataset, 7,429,601 tweets (64.9%) remained after filtering out tweets that did not reference food consumption (pre-pandemic: 4,111,454; pandemic: 3,318,147). Of the food-consumption tweets, 1,282,316 tweets (17.3%) were geotagged (pre-pandemic: 811,050; pandemic: 471,266). Of the 471,266 geotagged food-consumption tweets during the pandemic, 458,419 (97.3%) were matched to a county with a sufficient number of tweets and a valid Google mobility score.

Changes in healthy-food, fast-food, and alcohol tweets during the COVID-19 pandemic

The share of food tweets referencing healthy food was 20.5% higher (2.5 percentage point increase) during the pandemic period (14.7%) compared with the pre-pandemic period (12.2%) (Figure 1). In contrast, the fast-food tweets were 9.4% lower (0.8 percentage point decrease) during the pandemic period (7.4%) compared with the pre-pandemic period (8.1%). Similarly, alcohol tweets were 11.4% lower (3.2 percentage point decrease) during the pandemic period (24.7%) compared with the pre-pandemic period (27.9%).

Figure 2 shows relative changes in healthy-food, fast-food, and alcohol tweets by state (see Figure S1 for total number of geotagged tweets for each state). The share of healthy-food tweets increased in 49 states during the pandemic, with the largest relative increase in Wyoming (+62.1%), Vermont (+57.4%), and Washington (+46.5%). Only Massachusetts and Montana experienced a relative decrease in the shares of healthy-food tweets (−9.3% and −3.4%, respectively). In contrast, the share of fast-food tweets decreased in most states during the pandemic, with the largest relative decreases occurring in Rhode Island (−69.4%) and Wyoming (−68.0%). However, fast-food tweets increased in 15 states, with the largest increases occurring in Vermont (+130.2%), New Hampshire (+95.8%), and Idaho (+86.7%). The share of alcohol tweets decreased in most states during the pandemic, with the largest relative decreases occurring in Alaska (−39.7%), Hawaii (−38.7%), and Vermont (−37.6%). Only six states experienced an increase in alcohol tweets, with the largest relative increases occurring in South Dakota (+30.6%) and North Dakota (+13.2%).

Association between food tweets and county characteristics

For every percentage point increase in time spent at home (in places of residence) compared with the pre-pandemic baseline, the odds of a food tweet referencing healthy food during the pandemic increased (odds ratio [OR] = 1.019 [95% confidence interval (CI): 1.011, 1.027]) (Figure 3). Also, for every additional grocery store per 10,000 inhabitants in a county, the odds of a food tweet referencing healthy food during the pandemic increased (OR = 1.019 [95% CI: 1.007, 1.032]).

Furthermore, for every percentage point increase in time spent at home (in places of residence), the odds of a food tweet referencing fast food decreased (OR = 0.967 [95% CI: 0.956, 0.977]) (Figure 3). For every additional restaurant per 10,000 inhabitants in a county, the odds of a food tweet referencing fast food during the pandemic decreased (OR = 0.993 [95% CI: 0.991, 0.995]), while every additional grocery store per 10,000

CellPress
OPEN ACCESS

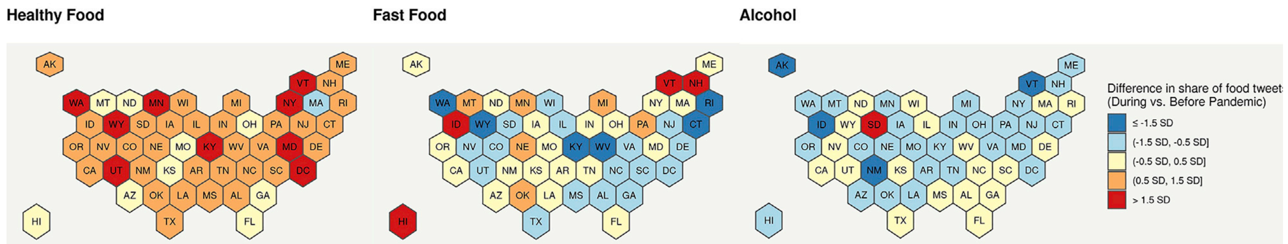

Figure 2. State-level changes in healthy-food, fast-food, and alcohol tweets during the COVID-19 pandemic in the United States
Geotagged food tweets were extracted from 1% of random tweets collected during the pre-pandemic period (May 15, 2019 to January 31, 2020) and the pandemic period (May 15, 2020 to January 31, 2021) using the Twitter streaming application programming interface. States associated with each tweet were determined by conducting a spatial join between the geotagged location of the tweets and state boundaries imported from the 2015–2019 American Community Survey. n = 811,050 geotagged food tweets during pre-pandemic period. n = 471,266 geotagged food tweets during pandemic period. Maps represent relative change during the pandemic in units of standard deviations (SDs) away from the pre-pandemic share of food tweets.

also decreased the odds of a food tweet referencing fast food (OR = 0.967 [95% CI: 0.949, 0.986]).

Liquor stores were associated with an increased odds of a tweet referencing alcohol during the pandemic (OR = 1.066 [95% CI: 1.046, 1.087]), while bars were associated with a decreased odds in a tweet referencing alcohol (OR = 0.989 [95% CI: 0.978, 1.000]) (Figure 3).

ORs for all explanatory variables in each of the three models are reported in the supplemental material (see Table S2).

DISCUSSION

We found that the proportion of tweets referencing healthy food increased by 20.5% during the pandemic period compared with the pre-pandemic period, while the proportion referencing fast food and alcohol decreased by 9.4% and 11.4%, respectively. In addition, more grocery stores per capita and more time spent at home (aggregated at the county level) coincided with an increased odds of healthy-food tweets and a decreased odds of fast-food tweets. More liquor stores per capita was also associated with an increased odds of alcohol tweets.

County-level changes in mobility during the pandemic, specifically more time spent at home, was associated with healthier food habits among Twitter users. For those able to stay at home during the pandemic, more time at home may have coincided with less exposure to fast food chains. In addition, more time spent at home may have afforded people more opportunities to prepare meals consisting of healthy food ingredients. This explanation is supported by our finding that more grocery stores per capita were also associated with the increased odds in healthy-food tweets and decreased odds in fast-food tweets. Our findings align with a previous Twitter analysis that highlighted the relationship between healthy food activity and the density of grocery stores.[18] An established body of literature highlights how food deserts—areas lacking grocery stores—limits access to healthy food.[19] Thus, counties with many grocery stores may have had more healthy food options during the pandemic, while counties lacking grocery stores did not. These findings highlight the potential influence of both healthy food environments and time spent at home on healthier food habits during the pandemic.

Similarly, our finding that alcohol tweets decreased slightly during the pandemic counters previous research showing increases in alcohol consumption.[2–5] This inconsistency may highlight the disconnect between alcohol consumption and posting about alcohol consumption on social media. The latter may occur more frequently in certain contexts of drinking alcohol that may have changed during the pandemic. For example, drinking to cope with pandemic-related isolation may be a context in which someone does not share alcohol consumption as freely compared with other contexts. Another explanation is that, while overall alcohol tweets may have decreased during the pandemic, certain county characteristics may have facilitated an increase in alcohol tweets. This explanation is supported by our findings that the number of liquor stores per capita—but not bars per capita—were associated with an increased odds of alcohol tweets during the pandemic. The findings also align with previous studies showing a positive association between liquor store access and binge drinking.[20,21] This finding highlights the potential influence of the built environment—particularly the presence of liquor stores that were kept open in most states during the pandemic—on alcohol consumption.

Tertiary findings indicate that there was a significant decrease in fast-food tweets in areas with more restaurants per capita and a significant decrease in alcohol tweets in areas with more bars per capita. In addition, our finding that fast-food tweets decreased overall during the pandemic was consistent with previous research showing a decline in fast food consumption in the United States during the pandemic.[1–3] A potential explanation for these relationships is that counties with a high density of restaurants and bars would likely have high rates of fast-food tweets and alcohol tweets before the pandemic. When the pandemic disrupted the food and drink industries, the largest relative decreases in fast-food tweets and alcohol tweets would likely occur in counties with many fast food restaurants and bars.

This study adds to previous research demonstrating that health behavior data derived from social media platforms can be used to detect small-area trends across periods of time. While small area estimation using data from behavioral surveys can be challenging due to limitations in sample size, the volume and veracity of Twitter data allow researchers to identify noteworthy shifts in attitudes and prevalence for specific areas. This is particularly useful during times of public health crisis, when traditional forms of data collection may be stalled. While

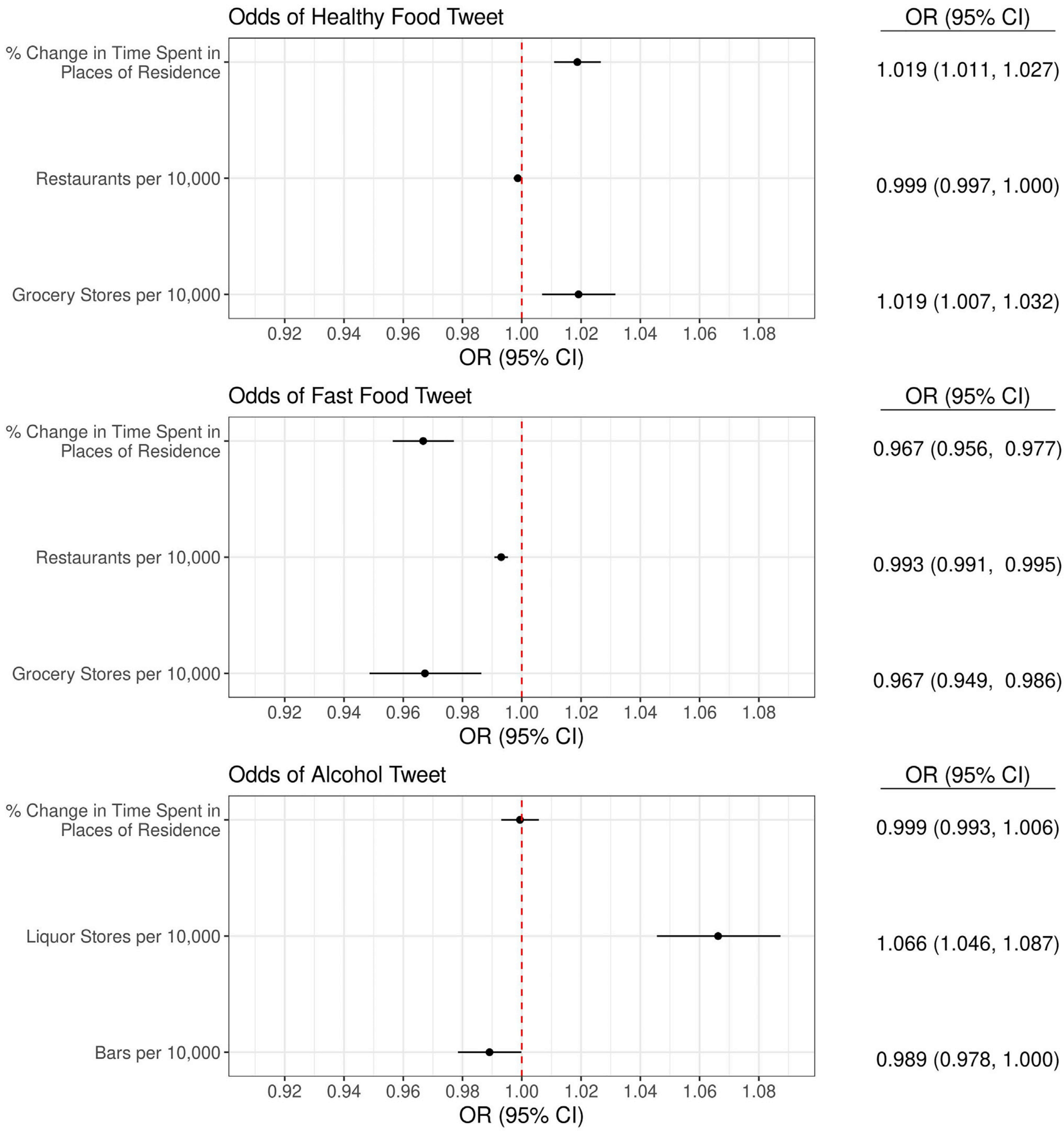

Figure 3. Association between food tweets and county characteristics related to population mobility and built environment

Geotagged food tweets were extracted from 1% of random tweets collected during the pre-pandemic period (May 15, 2019 to January 31, 2020) and the pandemic period (May 15, 2020 to January 31, 2021) using the Twitter streaming application programming interface. The sample used for the regression model consisted of 458,419 geotagged food tweets collected from 1,258 counties during the pandemic period in the United States. Data on population mobility were imported from the Google COVID-19 Community Mobility Reports. Built environment variables were derived from two data sources: the 2017–2018 Community Business Patterns database and the 2015–2019 American Community Survey. All models were adjusted for county-level demographics (race or ethnicity, age, and household income) and the respective pre-pandemic baseline of food tweet data collected from the pre-pandemic period. Odds ratios (ORs) are reported as the measure of association. Error bars indicate the 95% confidence intervals (CIs).

it is up to the researcher to identify possible causes for these trends, our findings align with others in showing the potential impact of public health interventions implemented during the pandemic on health behaviors.

Our study has some limitations. First, the findings from this study are not necessarily generalizable to the general US population, because not everyone uses Twitter and not all users on Twitter allow geolocation of their tweets. Second, tweets did not contain demographic information, thus preventing us from performing a stratified analysis. We addressed this limitation by adjusting for county-level demographic estimates related to race or ethnicity, age, and household income in our regression analysis. Third, because our regression analysis was conducted at the county level, there is risk of ecological fallacy; our models assume that aggregate county-level characteristics accurately represent the behaviors, demographics, and built environment of Twitter users. Future studies could address some of the stated limitations by applying known approaches to improving geolocation of tweets and inferring demographic attributes.[22] Overall, our study provides valuable insights on the changes in diet and alcohol consumption during the COVID-19 pandemic in the United States. Our findings showed an increase in healthy-food tweets and a decrease in fast-food and alcohol tweets during the pandemic compared with before the pandemic among Twitter users. We demonstrated potential links between behavior change and the built environment—particularly the association between healthier diet and access to grocery stores and more time spent at home, as well as the association between alcohol use and access to liquor stores during the pandemic.

EXPERIMENTAL PROCEDURES

Resource availability

Lead contact

Further information and requests for resources and data access will be fulfilled by the lead contact, Mark Hernandez (mark.hernandez@ll.mit.edu).

Materials availability

This study did not utilize any materials aside from the data and code noted below and did not generate new reagents.

Data and code availability

Original tweet text reported in this study cannot be deposited in a public repository because of privacy concerns and Twitter data-sharing policy. Analytic datasets containing tweet IDs, food or alcohol tweet coding, and area-level measures have been deposited in a GitHub repository—publicly available through Zenodo under https://doi.org/10.5281/zenodo.6057126. Tweet IDs can be used to retrieve original tweet data via Twitter's public application programming interface (API).

All original code has been deposited in a GitHub repository—publicly available through Zenodo under https://doi.org/10.5281/zenodo.6057126. Any additional information required to reanalyze the data reported in this paper is available from the lead contact (Mark Hernandez—mark.hernandez@ll.mit.edu) upon request.

Twitter data

We used the Twitter streaming application programming interface to collect a random subset of tweets (approximately 1% of public tweets were provided by Twitter) originating from within the United States. Twitter generates the random sample from all publicly available tweets. We split the data into two time periods: before and during the COVID-19 pandemic (pre-pandemic: May 15, 2019 to January 31, 2020; pandemic: May 15, 2020 to January 31, 2021). In addition to the tweet text, other relevant data included the time when the tweet was sent and latitude and longitude coordinates representing where the tweet originated.

Inclusion criteria and machine learning classification

Our analytic dataset consisted of geotagged tweets that referenced the consumption of healthy food, fast food, or alcohol in the United States during the pre-pandemic and pandemic periods. To construct this dataset, we used four steps to identify, label, and filter tweets based on key inclusion criteria: (1) identified all tweets that included food terms; (2) categorized food tweets that referenced healthy food, fast food, or alcohol; (3) applied machine learning methods to identify food tweets that inferred food consumption; and (4) used a spatial join between geocoded tweets and county boundaries to assign food consumption tweets to US counties.

We identified food tweets using a list of 1,430 food keywords created from the US Department of Agriculture's (USDA) National Nutrient Database and described in previous studies.[15,23] Tweets were filtered by retaining only tweets containing at least one of the food keywords. We then identified references to healthy food, fast food, and alcohol within the broader food dataset. Nutrition information, such as caloric density, was associated with each food item in the USDA database, which was used to help identify "healthy foods." Foods such as vegetables, nuts, and lean proteins (including fish, chicken, and turkey) were classified as "healthy foods." Fried versions of these foods were excluded from the healthy food category. Also tweets that mentioned a fast food restaurant were labeled as "fast food." In addition to this, we independently created a list of 52 keywords to identify references to alcohol (see Table S1).

To identify food tweets that referenced food consumption, we trained and evaluated three machine learning (ML) classifiers: random forest, support vector machines, and logistic regression. The ML classifiers were selected based on findings from previous studies.[12,13,24] The data used in training and evaluating the classifiers consisted of 2,878 labeled tweets indicating whether tweets referenced food consumption or not. Previous studies by co-author Nguyen and colleagues provided 2,379 of the tweets.[15] The remaining data were labeled by three of the co-authors into two classes: tweet referenced food consumption and tweet did not reference food consumption. Each coder worked independently to label the tweets and then the team met to discuss disagreements and arrived at a consensus. The final dataset we used for training and testing consisted of 1,674 tweets that referenced food consumption and 1,204 tweets that did not reference food consumption.

We used 70% of the data for training and 30% for evaluation after standard pre-processing, including the removal of stop words, stemming, lowercasing, and removal of special characters. Hyperparameters for random forest and support vector machines were tuned using tuning functions in R. The hyperparameters tuned for random forest included the number of trees and features selected at each split. For support vector machines, we optimized gamma and the penalty parameter.

Of the three classifiers, random forest performed the best at identifying tweets that referenced food consumption with an F1 score of 0.78 (precision = 0.77; recall = 0.79). In contrast, the F1 score for support vector machines (SVMs) was 0.74 (precision = 0.79; recall = 0.70) and that for logistic regression was 0.74 (precision = 0.77; recall = 0.70). The random forest model was used to classify food-consumption tweets in our full dataset.

Of the tweets that referenced food consumption, we restricted the analytic dataset to tweets that included geotagged locations with latitude and longitude coordinates. We conducted a spatial join between geotagged tweet locations and county boundaries from the 2015–2019 American Community Survey (ACS)[25] to determine the county associated with each tweet. We restricted our county-level analysis to only include tweets from counties with (1) 10 or more food tweets during the pre-pandemic period, (2) 10 or more food tweets during pandemic period, and (3) a valid Google mobility score.[26]

Primary outcomes

We derived three primary outcomes from the food-term variables. These outcomes were dichotomous variables that indicated whether any of the food-related terms in a tweet were categorized as healthy food, fast food, or alcohol. For this paper, we refer to the three primary outcomes as (1) "tweet referencing healthy food" or "healthy-food tweet," (2) "tweet referencing fast food" or "fast-food tweet," and (3) "tweet referencing alcohol" or "alcohol tweet."

Explanatory variables for regression analysis

We integrated county-level data describing sociodemographic characteristics, changes in mobility, and built environment to measure the association between county characteristics and the three primary outcomes. Sociodemographic variables related to age (percentage of younger individuals 10–24 years and percentage of older individuals 65+ years), race and ethnicity (percentage of non-Hispanic white, non-Hispanic Black, and Hispanic), and income (percentage of low-income households with annual household income less than $25,000 and high-income households with annual household income more than $100,000) were obtained from the 2015–2019 ACS.[25]

We obtained data on aggregated changes in time spent in homes from the Google COVID-19 Community Mobility Reports. The variable measured percent change in time spent in places of residence compared with baseline days, where baseline days were defined as the 5-week median for each day of the week during the period prior to the pandemic in the United States (January 3, 2020 to February 6, 2020). We used this variable to approximate changes in time spent at home.

We also included county-level built environment variables using the 2017–2018 County Business Patterns (CBP) database.[27] We constructed variables that measured the number of food establishments per capita by dividing the number of food establishments in each county by the number of inhabitants in the county. This resulted in four food establishment variables: (1) total grocery stores (excluding convenience stores) per 10,000 inhabitants; (2) total restaurants per 10,000 inhabitants; (3) total bars per 10,000 inhabitants; and (4) total liquor stores per 10,000 inhabitants.

Lastly, we included a term to adjust for the county-level baseline trends in food tweets. We calculated the share of healthy-food, fast-food, and alcohol tweets for each county during the pre-pandemic period.

Statistical analysis

Our statistical analysis was divided into two parts: (1) to describe the before-and-after trends of tweets referring to healthy food, fast food, and alcohol nationally and across states and (2) to measure the association between county-level factors on changes in healthy-food-, fast-food-, and alcohol-related tweets. We conducted our analyses using R statistical software (v.4.0.2) and the following packages: "tidyverse," "tidycensus," "sf," "e1071," and "randomForest."

We first calculated descriptive statistics to assess the share of tweets referencing healthy food, fast food, and alcohol. We then calculated the relative difference (percent change) and absolute difference (percentage point change) in the outcomes during the pandemic period compared with the pre-pandemic period. We calculated these metrics on the entire analytic sample of geotagged food tweets to capture national trends and by state to describe how changes in food tweets differed within the United States. States associated with each tweet were determined by conducting a spatial join between the geotagged tweet locations and state boundaries from the 2015–2019 ACS.[25]

Next, we fit three binomial logistic regression models to the analytic sample of geotagged tweets to measure associations between our explanatory variables and each of the three primary outcomes. We reported ORs as the measure of association. For each model, we set the dependent variables as one of the three primary outcomes indicating whether a food tweet referenced (1) healthy food, (2) fast food, or (3) alcohol. We adjusted for the county-level, pre-pandemic proportion of food tweets referencing each respective food category. For example, the healthy-food model adjusted for the county-level proportion of healthy-food tweets before the pandemic. We adjusted for state fixed effects in all three models. We also included all sociodemographic variables (county share of white, Black, Hispanic, younger, older, low-income, and high-income populations) in all three models. We included the county-level mobility variable for changes in time spent in homes during the pandemic in all three models. Finally, we included built environment variables relevant to each model; we examined grocery stores per capita and restaurants per capita in both the healthy-food model and the fast-food model and bars per capita and liquor stores per capita in the alcohol model.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.patter.2022.100547.

ACKNOWLEDGMENTS

Funding for publication fees was provided from the Boston University Faculty of Computing & Data Sciences (CDS) Faculty Fellows Program.

AUTHOR CONTRIBUTIONS

Conceptualization, E.O.N. and Q.C.N.; resources, E.O.N. and Q.C.N.; data curation, M.A.H., S.M., K.M., P.D., and Q.C.N.; methodology, E.O.N., M.A.H., and N.L.C.; formal analysis, M.A.H.; investigation, M.A.H., S.M., and K.M.; supervision, E.O.N. and N.L.C.; writing – original draft, M.A.H., S.M., and K.M.; writing – review & editing, all authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 2, 2022
Revised: May 19, 2022
Accepted: June 10, 2022
Published: June 15, 2022

REFERENCES

1. Chenarides, L., Grebitus, C., Lusk, J.L., and Printezis, I. (2021). Food consumption behavior during the COVID-19 pandemic. Agribusiness *37*, 44–81. https://doi.org/10.1002/agr.21679.
2. Chen, L., Li, J., Xia, T., Matthews, T.A., Tseng, T.-S., Shi, L., Zhang, D., Chen, Z., Han, X., Li, Y., et al. (2021). Changes of exercise, screen time, fast food consumption, alcohol, and cigarette smoking during the COVID-19 pandemic among adults in the United States. Nutrients *13*, 3359. https://doi.org/10.3390/nu13103359.
3. Bakaloudi, D.R., Jeyakumar, D.T., Jayawardena, R., and Chourdakis, M. (2021). The impact of COVID-19 lockdown on snacking habits, fast-food and alcohol consumption: a systematic review of the evidence. Clin. Nutr. https://doi.org/10.1016/j.clnu.2021.04.020.
4. Pollard, M.S., Tucker, J.S., and Green, H.D., Jr. (2020). Changes in adult alcohol use and consequences during the COVID-19 pandemic in the US. JAMA Netw. Open *3*, e2022942. https://doi.org/10.1001/jamanetworkopen.2020.22942.
5. Grossman, E.R., Benjamin-Neelon, S.E., and Sonnenschein, S. (2020). Alcohol consumption during the COVID-19 pandemic: a cross-sectional survey of us adults. Int J. Environ. Res. Public Health *17*, 9189. https://doi.org/10.3390/ijerph17249189.
6. Janssen, M., Chang, B.P.I., Hristov, H., Pravst, I., Profeta, A., and Millard, J. (2021). Changes in food consumption during the COVID-19 pandemic: analysis of consumer survey data from the first lockdown period in Denmark, Germany, and Slovenia. Front. Nutr. *8*, 635859. https://doi.org/10.3389/fnut.2021.635859.
7. Huang, Y., Huang, D., and Nguyen, Q.C. (2019). Census tract food tweets and chronic disease outcomes in the U.S., 2015–2018. Int J. Environ. Res. Public Health *16*, 975. https://doi.org/10.3390/ijerph16060975.
8. Embree, B.G., and Whitehead, P.C. (1993). Validity and reliability of self-reported drinking behavior: dealing with the problem of response bias. J. Stud. Alcohol *54*, 334–344. https://doi.org/10.15288/jsa.1993.54.334.
9. Davis, C.G., Thake, J., and Vilhena, N. (2010). Social desirability biases in self-reported alcohol consumption and harms. Addict. Behav. *35*, 302–311. https://doi.org/10.1016/j.addbeh.2009.11.001.
10. Johnson, R.K. (2002). Dietary intake–how do we measure what people are really eating? Obes. Res. *10*, 63S–68S.
11. Conway, J.M., Ingwersen, L.A., and Moshfegh, A.J. (2004). Accuracy of dietary recall using the USDA five-step multiple-pass method in men: an observational validation study. J. Am. Diet Assoc. *104*, 595–603. https://doi.org/10.1016/j.jada.2004.01.007.
12. Cesare, N., Dwivedi, P., Nguyen, Q.C., and Nsoesie, E.O. (2019). Use of social media, search queries, and demographic data to assess obesity

prevalence in the United States. Palgrave Commun. *5*, 106. https://doi.org/10.1057/s41599-019-0314-x.

13. Nguyen, Q.C., McCullough, M., Meng, H.w., Paul, D., Li, D., Kath, S., Loomis, G., Nsoesie, E.O., Wen, M., Smith, K.R., and Li, F. (2017). Geotagged US tweets as predictors of county-level health outcomes, 2015–2016. Am. J. Public Health *107*, 1776–1782. https://doi.org/10.2105/ajph.2017.303993.
14. Meng, H.-W., Kath, S., Li, D., and Nguyen, Q.C. (2017). National substance use patterns on Twitter. PLoS One *12*, e0187691. https://doi.org/10.1371/journal.pone.0187691.
15. Nguyen, Q.C., Li, D., Meng, H.-W., Kath, S., Nsoesie, E., Li, F., and Wen, M. (2016). Building a national neighborhood dataset from geotagged twitter data for indicators of happiness, diet, and physical activity. JMIR Public Health Surveill. *2*, e158. https://doi.org/10.2196/publichealth.5869.
16. Nguyen, Q.C., Kath, S., Meng, H.-W., Li, D., Smith, K.R., VanDerslice, J.A., Wen, M., and Li, F. (2016). Leveraging geotagged Twitter data to examine neighborhood happiness, diet, and physical activity. Appl. Geogr. *73*, 77–88. https://doi.org/10.1016/j.apgeog.2016.06.003.
17. Widener, M.J., and Li, W. (2014). Using geolocated Twitter data to monitor the prevalence of healthy and unhealthy food references across the US. Appl. Geogr. *54*, 189–197. https://doi.org/10.1016/j.apgeog.2014.07.017.
18. Chen, X., and Yang, X. (2014). Does food environment influence food choices? A geographical analysis through "tweets. Appl. Geogr. *51*, 82–89. https://doi.org/10.1016/j.apgeog.2014.04.003.
19. Walker, R.E., Keane, C.R., and Burke, J.G. (2010). Disparities and access to healthy food in the United States: a review of food deserts literature. Health Place *16*, 876–884. https://doi.org/10.1016/j.healthplace.2010.04.013.
20. Pereira, G., Wood, L., Foster, S., and Haggar, F. (2013). Access to alcohol outlets, alcohol consumption and mental health. PLoS One *8*, e53461. https://doi.org/10.1371/journal.pone.0053461.
21. Shimotsu, S.T., Jones-Webb, R.J., MacLehose, R.F., Nelson, T.F., Forster, J.L., and Lytle, L.A. (2013). Neighborhood socioeconomic characteristics, the retail environment, and alcohol consumption: a multilevel analysis. Drug Alcohol Depend. *132*, 449–456. https://doi.org/10.1016/j.drugalcdep.2013.03.010.
22. Cesare, N., Grant, C., Nguyen, Q., Lee, H., and Nsoesie, E.O. (2018). How well can machine learning predict demographics of social media users?. Preprint at arXiv.
23. US Department of Agriculture. (2019). USDA National Nutrient Database for Standard Reference.
24. Cesare, N., Nguyen, Q.C., Grant, C., and Nsoesie, E.O. (2019). Social media captures demographic and regional physical activity. BMJ Open Sport Exerc. Med. *5*, e000567. https://doi.org/10.1136/bmjsem-2019-000567.
25. US Census Bureau (2019). American Community Survey (ACS).
26. Google LLC (2020). Google COVID-19 Community Mobility Reports.
27. US Census Bureau (2018). County Business Patterns (CBP).