

Original Paper

# The Potential of Smartphone Apps in Informing Protobacco and Antitobacco Messaging Efforts Among Underserved Communities: Longitudinal Observational Study

Edmund WJ Lee<sup>1,2,3</sup>, BA, MA, PhD; Mesfin Awoke Bekalu<sup>1,2</sup>, PhD; Rachel McCloud<sup>1,2</sup>, MPH, PhD; Donna Vallone<sup>4,5,6</sup>, MPH, PhD; Monisha Arya<sup>7,8</sup>, MPH, MD; Nathaniel Osgood<sup>9</sup>, PhD; Xiaoyan Li<sup>9</sup>, MSc; Sara Minsky<sup>1</sup>, MPH; Kasisomayajula Viswanath<sup>1,2</sup>, PhD

<sup>1</sup>Dana-Farber Cancer Institute, Boston, MA, United States

<sup>2</sup>Harvard TH Chan School of Public Health, Boston, MA, United States

<sup>3</sup>Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore, Singapore

<sup>4</sup>Schroeder Institute, Truth Initiative, Washington, DC, United States

<sup>5</sup>College of Global Public Health, New York University, New York, NY, United States

<sup>6</sup>Department of Health, Behavior and Society, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD, United States

<sup>7</sup>Baylor College of Medicine, Houston, TX, United States

<sup>8</sup>Center for Innovation in Quality, Effectiveness and Safety, Michael E DeBakey VA Medical Center, Houston, TX, United States

<sup>9</sup>Department of Computer Science, College of Arts and Science, University of Saskatchewan, Saskatoon, SK, Canada

**Corresponding Author:**

Edmund WJ Lee, BA, MA, PhD

Dana-Farber Cancer Institute

375 Longwood Avenue

Boston, MA, 02215

United States

Phone: 1 6178587988

Email: [Edmund.Lee@dfci.harvard.edu](mailto:Edmund.Lee@dfci.harvard.edu)

## Abstract

**Background:** People from underserved communities such as those from lower socioeconomic positions or racial and ethnic minority groups are often disproportionately targeted by the tobacco industry, through the relatively high levels of tobacco retail outlets (TROs) located in their neighborhood or protobacco marketing and promotional strategies. It is difficult to capture the smoking behaviors of individuals in actual locations as well as the extent of exposure to tobacco promotional efforts. With the high ownership of smartphones in the United States—when used alongside data sources on TRO locations—apps could potentially improve tobacco control efforts. Health apps could be used to assess individual-level exposure to tobacco marketing, particularly in relation to the locations of TROs as well as locations where they were most likely to smoke. To date, it remains unclear how health apps could be used practically by health promotion organizations to better reach underserved communities in their tobacco control efforts.

**Objective:** This study aimed to demonstrate how smartphone apps could augment existing data on locations of TROs within underserved communities in Massachusetts and Texas to help inform tobacco control efforts.

**Methods:** Data for this study were collected from 2 sources: (1) geolocations of TROs from the North American Industry Classification System 2016 and (2) 95 participants (aged 18 to 34 years) from underserved communities who resided in Massachusetts and Texas and took part in an 8-week study using location tracking on their smartphones. We analyzed the data using spatial autocorrelation, optimized hot spot analysis, and fitted power-law distribution to identify the TROs that attracted the most human traffic using mobility data.

**Results:** Participants reported encountering protobacco messages mostly from store signs and displays and antitobacco messages predominantly through television. In Massachusetts, clusters of TROs (Dorchester Center and Jamaica Plain) and reported smoking behaviors (Dorchester Center, Roxbury Crossing, Lawrence) were found in economically disadvantaged neighborhoods. Despite the widespread distribution of TROs throughout the communities, participants overwhelmingly visited a relatively small number of TROs in Roxbury and Methuen. In Texas, clusters of TROs (Spring, Jersey Village, Bunker Hill Village, Sugar Land, and

Missouri City) were found primarily in Houston, whereas clusters of reported smoking behaviors were concentrated in West University Place, Aldine, Jersey Village, Spring, and Baytown.

**Conclusions:** Smartphone apps could be used to pair geolocation data with self-reported smoking behavior in order to gain a better understanding of how tobacco product marketing and promotion influence smoking behavior within vulnerable communities. Public health officials could take advantage of smartphone data collection capabilities to implement targeted tobacco control efforts in these strategic locations to reach underserved communities in their built environment.

(*J Med Internet Res* 2020;22(7):e17451) doi: [10.2196/17451](https://doi.org/10.2196/17451)

## KEYWORDS

mobile health; mobile phone; tobacco use; big data; spatial analysis; data science

## Introduction

### Background

Tobacco use is a major risk factor for lung cancer and premature morbidity [1] and also one of the leading preventable causes of death in the United States [2,3]. Approximately 480,000 deaths (1 in 5 deaths) in the United States annually could be attributed to smoking or tobacco consumption [4]. Although tobacco use is a prevalent public health problem in the United States, research has shown that the health and economic burden of tobacco use disproportionately affect underserved communities and that these communities have not benefitted from tobacco control efforts as much as others [5]. Research has also shown that the density of tobacco retail outlets (TROs) is higher in low-income neighborhoods [6] as well as in communities with higher percentages of ethnic minorities [7,8]. In addition, people from lower socioeconomic positions (SEPs) are often the target of the tobacco industry's advertising, including the places of their residence [9]. The presence of TROs, together with a disproportionate exposure to protobacco messages, is associated with smoking behaviors and may attenuate the attempts of smokers to quit by allowing easy access to tobacco products as well as encouraging impulse purchases by providing environmental cues for smoking [10].

Although many studies have examined how factors such as the proximity of a residence to TROs and exposure to protobacco messages are related to smoking attitudes and initiation, smoking prevalence, and even hospital admissions [3,11-14], few studies have investigated the exposure and awareness of antitobacco messaging within underserved communities, particularly in relation to TROs and across media platforms. For instance, what are the media platforms where underserved communities are exposed to anti- or protobacco messages? Are people strategically exposed to antitobacco messages near locations such as TROs or places they are most likely to smoke or other places? To date, research on the extent of such antitobacco and protobacco message exposure is limited by its reliance on aggregated cross-sectional self-reported survey data.

### The Potential of Smartphone Apps in Informing Protobacco and Antitobacco Messaging Efforts

With the ubiquity of smartphone ownership, particularly within vulnerable communities, data from smartphone apps provide a potential data collection mechanism for informing health policy makers where to focus antitobacco messaging efforts [15,16], especially when they are used to complement traditional data

sources such as the location of TROs. After all, smartphone penetration in the United States is high, with about 81% of the population owning a smartphone; the smartphone ownership figures are also high for underserved communities, such as people living in rural areas (71%), those making less than US \$30,000 annually (71%), and in minority communities (approximately 79% to 80%) [17]. Health promotion organizations and policy makers could take advantage of the contextual information provided by smartphones to identify strategic areas to help ensure adequate exposure to antitobacco messages [18].

Smartphones enable researchers to *passively* engage in data collection at-scale within people's naturalistic environments. For instance, by enabling geolocation tracking with the explicit consent of users, smartphone apps can collect temporally ordered information on the precise locations they visited or the paths they have taken, without being intrusive. The ability to track the behaviors of those in underserved communities in situ is a huge advantage over traditional survey methods that rely on self-reported recall [19]. Methods based on recall are limited in that participants may not remember all locations they have visited accurately, or they may omit details for the sake of social desirability (eg, underreporting of places they may perceive as undesirable). Geolocation tracking can provide valuable information for health organizations in helping to strategically target antitobacco messaging efforts by accurately identifying where protobacco messaging is being encountered. For example, the mobility or path data that outline how individuals move through their communities would be useful for identifying the most popular TROs and other place-based platforms where messages are being aired.

This type of smartphone data collection might also provide opportunities for health organizations to partner with underserved communities for *participatory science* efforts [20]. Although collecting data from the underserved communities can be extremely difficult [21], the use of smartphones could circumvent this problem as it adds a minimum logistical burden given the role of smartphones in their lives. The data collection could employ ecological momentary assessment (EMA) techniques, which assess particular events in the lives of subjects at periodic intervals, such as smoking behavior or exposure to specific types of messages, which are automatically prompted. These data, together with geolocation of smoking-related behaviors, could be used to map smoking *hot spots* [22]—defined as locations where there are non-random observed patterns of clustering—which are areas where there is a

statistically significant clustering of respondents who report smoking in the same area.

These types of data collection efforts can facilitate the proactive reporting of exposure to tobacco messages that have high temporal specificity and are capable of capturing details of even ephemeral exposures (eg, a photo of an advertisement on a rotating billboard or as part of a video at a gas station and radio advertisements). Insights into where and how anti- and protobacco messages are reaching those in underserved communities can assist tobacco control practitioners and policy makers in helping reduce the disproportionate burden of tobacco use within vulnerable communities.

### Objectives of the Study

This study aimed to examine how smartphone app data collection could complement existing data sources to help inform tobacco control efforts for underserved communities. There are three specific objectives of this paper. First, we sought to identify if there were concentrations of TROs in Massachusetts and Texas. Second, using both passive (ie, geolocations) and active (ie, self-reports) data, we aimed to identify (1) the most popular TROs, denoted by a small number of TROs that attracted the most human traffic; (2) the areas in which participants were most likely to smoke; and (3) the locations where the participants reported exposure to tobacco messages and where the concentrations of pro- and antitobacco messages were. Third, we drew suggestions for tobacco control based on our results.

## Methods

### Study Design and Recruitment

To address the objectives of our study, we conducted a small-scale feasibility test using a smartphone app in underserved communities in Massachusetts and Texas. We have chosen to conduct the study in these 2 states, given the diversity in tobacco control policy implementation, where Massachusetts had stricter tobacco laws as compared with Texas [23,24]. Ethics approval was obtained from the respective institutional review boards (IRBs) of Harvard University, Baylor College of Medicine, and the University of Saskatchewan after extensive review, which ensured that adequate layers of protection were in place for our participants. Upon receiving the IRB approvals, we recruited 95 participants (smokers and nonsmokers) aged 18 to 34 years who resided in different cities within Massachusetts and Texas to participate in our 8-week smartphone tobacco tracking study.

Participants were required to meet the following criteria: (1) existing Android smartphone users with a data plan (although we covered the costs of their plan for the duration of the study) or would be willing to change their primary phone to a study-compatible phone, (2) consented to download a location-tracking smartphone app called *Ethica* and to keep their location-tracking feature switched on for the duration of the study, and (3) were willing to complete a pretest at the start of the study and a posttest at the end of the study as well as respond to EMAs that would be pushed to them. *Ethica* is a smartphone app designed to collect sensor-based data (eg,

geolocations, accelerometry, and electrodermal activity) as well as contextual self-reports (eg, EMAs).

Once the participants downloaded *Ethica* (assisted by study staff) and registered using their email and a password, the study staff helped ensure that the location-tracking feature on their phones was enabled. The participants were then asked to complete the pretest via the app on their smartphone. This pretest contained questions pertaining to demographics, smoking status, number of cigarettes smoked in the past 30 days, and other related smoking attitudinal and behavioral questions. The location-tracking app began collecting geolocation data once registration was complete. At the end of the study, the participants completed a similar posttest.

### Profile of Participants

Among the 95 participants (49 females, 42 males, 4 nonresponse), 51 were from Massachusetts and 43 were from Texas, with 1 nonresponse. Of all the 95 participants, when asked if they were of Hispanic, Latino, or Spanish origin, 54 (57%) reported they were “not of Hispanic, Latino, or Spanish origin”; 7 (7%) were “Mexican, Mexican American, Chicano”; 10 (11%) were “Puerto Rican”; 12 (13%) were “Dominican”; 1 (1%) was “Cuban”; 6 (6%) were “another Hispanic, Latino, or Spanish origin (eg, Guatemalan, Salvadoran, Honduran, Nicaraguan, Panamanian, Colombian, Venezuelan, Peruvian)”; and 5 (5%) with no response. In terms of race, 37/95 (39%) of our participants identified as “Black or African American”, 33/95 (35%) as “White”, and the rest identified themselves as a combination of different ethnic groups (eg, “American Indian or Alaska Native”, “Asian”, “Native Hawaiian”, or “other Pacific Islander”).

The median total combined household income was between US \$20,000 and US \$29,000 (from 1 [US \$10,000] to 9 [ $\geq$ US \$75,000]; median 3.00 [US \$20,000 and US \$29,000]; SD 2.11), and the median education status was having *some college* (1 [completed grade school or less] to 8 [completed graduate or professional school after college]; median 5.00 [some college]; SD 1.32). In total, 53 participants self-identified as smokers, whereas 41 were nonsmokers, with 1 nonresponse. The participants were also asked to report the total number of cigarettes they smoked in the past 30 days (mean 11.3, SD 13.4).

### Data Management and Processing

#### *Geolocations of Tobacco Retail Outlets*

The geolocations of TROs were obtained from the North American Industry Classification System (NAICS). Developed by the Office of Management and Budget, NAICS is the federal standard business classification system based on the primary activities of businesses. We identified and extracted records for 252 TROs in Massachusetts and 1422 in Texas, based on the NAICS classification of *Tobacco Stores*. For this study, we chose to focus our analysis on TROs that were solely cigars, cigarettes, and tobacco dealers and retailers or smoke shops (NAICS8 code: Tobacco Stores) and excluded retailers whose primary descriptions were not in the area of tobacco sales (eg, beer, wine, and liquor stores; convenience stores; and gasoline stations with convenience stores), even though they might sell tobacco products. The reason for excluding these stores was

that people might pass by or linger at these places because of reasons (eg, when shopping for groceries) other than tobacco purchase and consumption.

### Participants' Geolocations

Ethica collected approximately 31 million time-stamped geolocations from all the participants recorded in millisecond intervals. To increase the reliability of the data (eg, as there were multiple geolocations of individuals recorded when they were stationary), we collapsed participants' geolocations into multiple 10-min time intervals and extracted the most accurate and representative longitude and latitude locations for each interval. In total, there were 279,840 geolocations of participants from Massachusetts and 227,991 geolocations for participants in Texas.

### Geolocations of Smoking Behaviors

Questions regarding smoking behaviors were randomly administered 4 times a day to smokers via Ethica (between 8 AM and 9 PM on weekdays and between 10 AM and 9 PM on weekends). The geolocations of participants when they were smoking were captured from their responses to the question "Have you smoked in the past hour," in which they were asked to select from the following responses: (1) I smoked a cigarette in the past hour, (2) I smoked an electronic cigarette (e-cigarette) in the past hour, (3) I used another tobacco product in the past hour, (4) I am smoking a cigarette right now, (5) I am smoking an e-cigarette right now, (6) I am using another tobacco product right now, and (7) I have not smoked. To obtain the geolocations of the participants when they were smoking, we extracted the longitude and latitude of smokers at the time if they indicated that they were smoking a cigarette or an e-cigarette or using another tobacco product *right now*. A total of 10,393 smoking geolocations in Massachusetts and 10,187 in Texas were recorded.

### Geolocations of Tobacco Message Exposure

Through Ethica, participants were able to take or upload photos of tobacco messages and advertisements they came across in their communities (eg, billboards, TROs) or on the internet. After this, they were prompted to answer an EMA survey where the participants were given the options to identify the messages as either antitobacco or protobacco and to report when they saw the message (where 1=I see it right now; 2=in the past hour; 3=in the past 1-5 hours; 4=more than 5 hours ago). The latitude and longitude of the photos and EMA surveys were logged using Ethica.

### Statistical Analysis

Data were imported into ArcMap 10.6.1 for mapping and statistical analyses, where we conducted *spatial autocorrelation* and *optimized hot spot analysis* as well as *power-law analysis* in R studio (version 1.8383) to address all the study objectives. The global spatial autocorrelation was used to test for the presence of spatial variation in a given dataset [25], specifically in examining the correlation among data points that are close to one another and to determine if there is a nonrandom spatial clustering among data points that were in close proximity [26]. The global Moran index is a statistic that indicates the presence of statistically significant spatial clustering, which produces a number between  $-1$  and  $+1$ . A negative value indicates the presence of negative spatial autocorrelation, which is the tendency for dissimilar values to be located together. On the other hand, positive values indicate the presence of positive spatial autocorrelation, where data points with similar values are clustered together [25]. If the presence of spatial clustering was detected, we then conducted the optimized hot spot analysis to examine the locations of the *hot spots* and *cold spots* of TROs, smoking, and antitobacco messages.

The power-law analysis aimed to test if there was an observable power-law distribution in the data. A power-law distribution is also known as a heavy tail distribution, where smaller values on the  $x$ -axis correspond to large values in the  $y$ -axis. In other words, in the context of this study, if a power-law distribution is observable, a small number of TROs would attract the most human traffic. This may suggest that certain TROs are more popular, or centrally located, such that people are more likely to pass by as compared with TROs located in obscure locations.

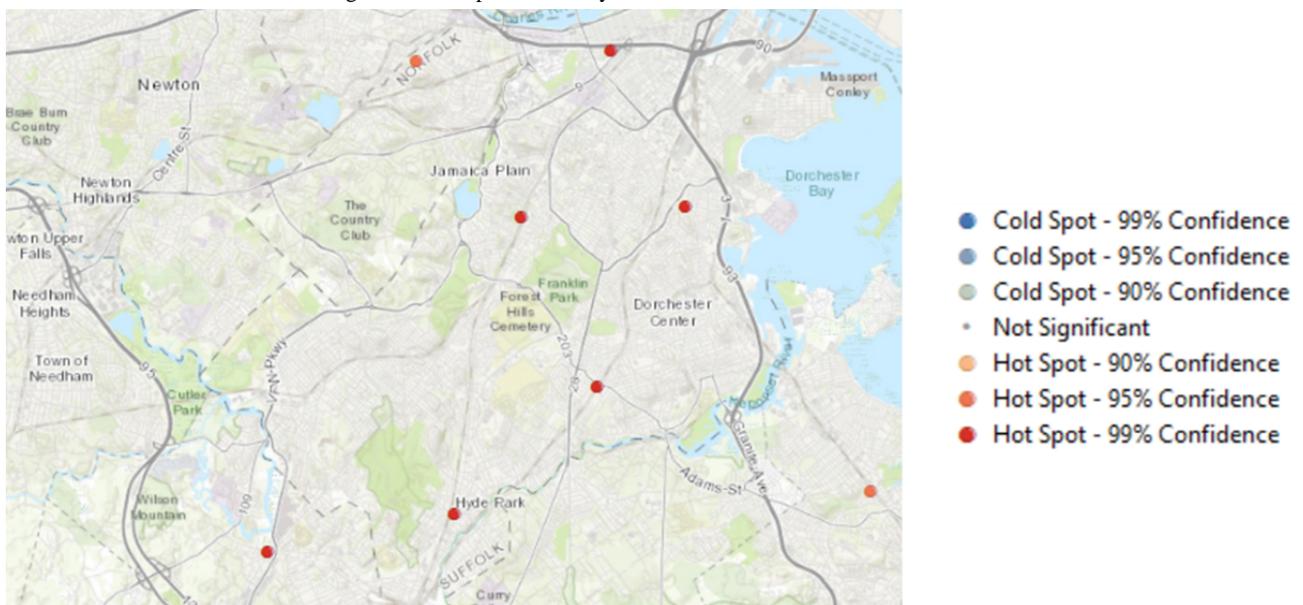
### Results

Our first objective was to identify if there were concentrations of TROs (ie, TRO hot spots) in Massachusetts and Texas (see Figures 1-3). To do so, we conducted spatial autocorrelation on the geolocations of TROs and determined if there was a statistically significant spatial clustering of TROs. The results suggest that there was evidence of clustering of TROs in both Massachusetts (global Moran index=0.79;  $z=8.04$ ;  $P<.001$ ) and Texas (global Moran index=0.69;  $z=5.85$ ;  $P<.001$ ). Next, optimized hot spot analysis in ArcMap showed that there was a statistically significant clustering of TROs in the city of Boston, with the most significant clusters found in Dorchester Center, Jamaica Plain, and Hyde Park ( $z\geq 3.50$ ;  $P<.001$ ). The TRO hot spots in Texas were found in Houston, and they were in places such as Spring, Jersey Village, Bunker Hill Village, Sugar Land, and Missouri City.

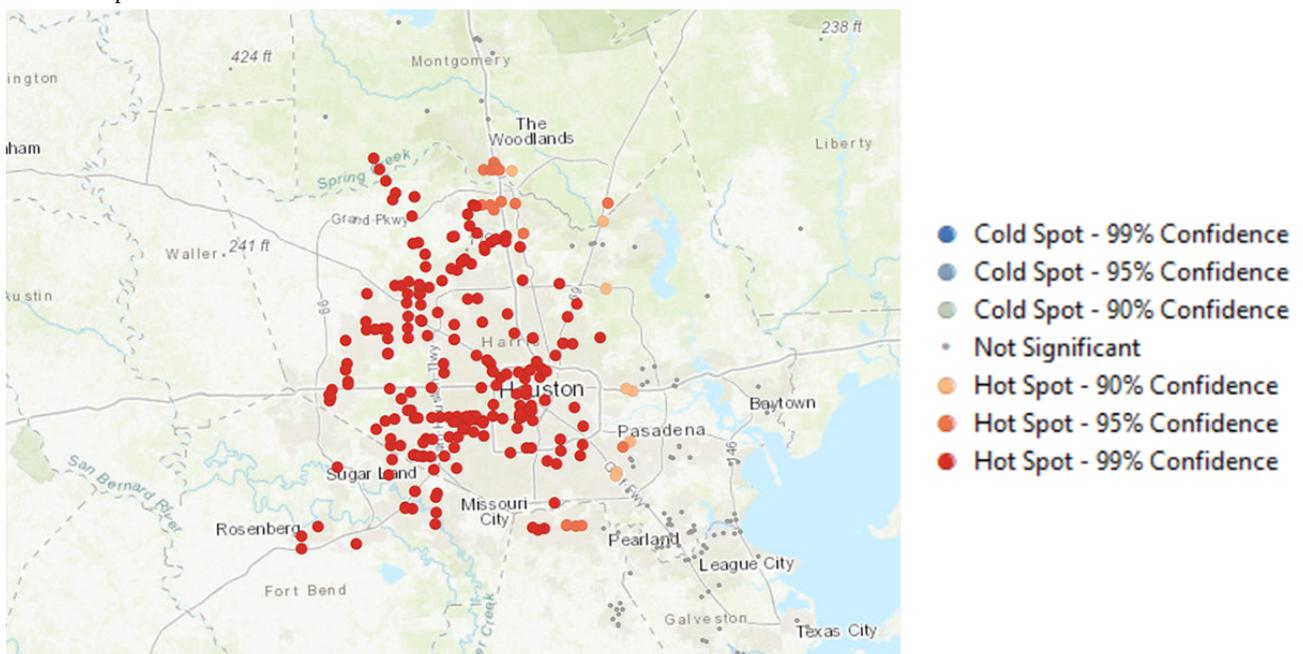
**Figure 1.** Hot spots of tobacco retail outlets in the state of Massachusetts.



**Figure 2.** Zoomed-in view of the most significant hot spots in the city of Boston.



**Figure 3.** Hot spots of tobacco retail outlets in the state of Texas.



Our second objective was to draw upon both passive (ie, geolocations) and active (ie, self-reports) data to identify (1) the most popular TROs, (2) the areas in which participants were most likely to smoke, and (3) the locations where participants reported being exposed to tobacco messages and where these concentrations of protobacco and antitobacco messages were.

To examine which were the most popular TROs—if a small number of TROs attracted the most traffic—we tested if a power-law distribution was observable by analyzing geolocations of participants near the TROs. To do so, we created a 100-m buffer around all the TROs in our dataset and performed a spatial join with all the geolocations that intersected within the buffer. The selection of a 100-m buffer was consistent with previous research [27]. We then exported the data to R studio to fit a power-law distribution in accordance with the steps recommended by Clauset et al [28]: (1) construct a discrete power-law distribution object, (2) estimate the  $x_{\min}$  and exponent  $\alpha$  of the power law and assign them to the power-law object, and (3) bootstrap to obtain the  $P$  value for the hypothesis test of if the data followed a power-law distribution. In the Kolmogorov-Smirnov test, the null hypothesis is that observations will follow a specific distribution, whereas the alternative hypothesis specifies that a set of distribution does not follow a specific distribution. As such, to claim that observations follow a power law distribution, the  $P$  value would have to be equal or more than .05 for the null hypothesis to be accepted, thereby indicating the presence of a power-law distribution. The analysis found marginal support for the power-law distribution ( $D=0.12$ ;  $P=.05$ ) for TROs in Massachusetts (Figure 4) but not for Texas. The top TROs that attracted the most human traffic from our sample were in the neighborhood of Roxbury in the city of Boston and Methuen, a city close to Boston.

Next, to identify the areas where participants were most likely to smoke (ie, smoking hot spots), we conducted spatial autocorrelation on geolocations where the participants reported their smoking behavior through the EMAs and determined if there was a statistically significant spatial clustering of smokers who reported smoking (Figure 5). The results suggest that there was significant clustering in both Massachusetts (global Moran index=0.29;  $z=34$ ;  $P<.001$ ) and Texas (global Moran

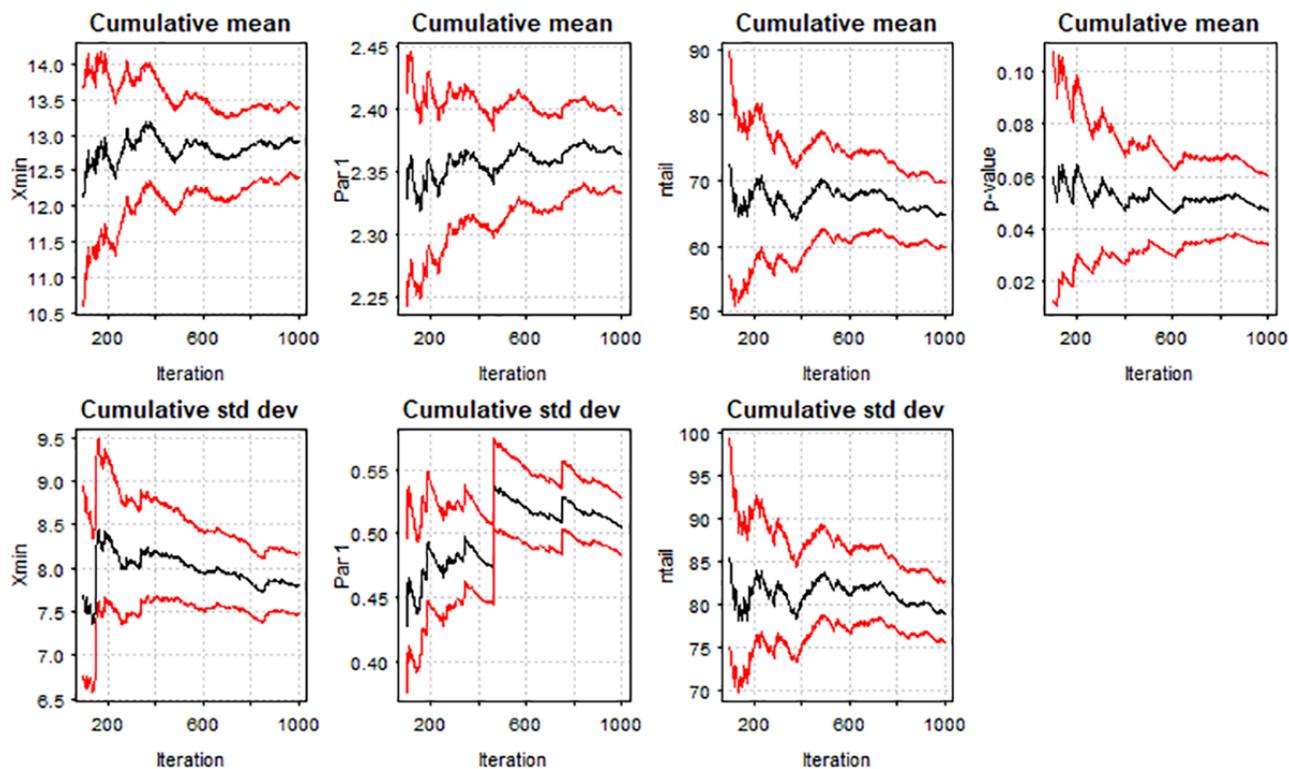
index=0.25;  $z=63.5$ ;  $P<.001$ ). Next, we conducted optimized hot spot analysis, and the results showed that in Massachusetts, the heaviest smokers (based on the number of cigarettes smoked in the past 30 days) tended to report that they smoked in Dorchester Center, Roxbury Crossing, Lawrence, and Peabody ( $z\geq 2.84$ ;  $P<.001$ ). In Texas, the heaviest smokers tended to report smoking in West University Place, Aldine, Jersey Village, Spring, and Baytown ( $z\geq 2.67$ ;  $P<.001$ ).

To identify the locations where the participants reported being exposed to antitobacco and protobacco messages, we examined the photos taken by the participants through the app where they rated if the messages were either antitobacco or protobacco. In Massachusetts, there were 41 antitobacco and 48 protobacco messages reported (see Multimedia Appendix 1). The top 3 most frequent platforms for exposure to antitobacco messages in Massachusetts were on (1) television and others (19.5% each, 8/41 for television and 8/41 for others), (2) store sign or display (7/41, 17.1%), and (3) billboard/bus/train stop advertisements (6/41, 14.6%). The top 3 highest exposures to protobacco messages were on (1) store sign or display (27/48, 56.3%), (2) newspaper or magazine (7/48, 14.6%), and (3) website (5/48, 10.4%).

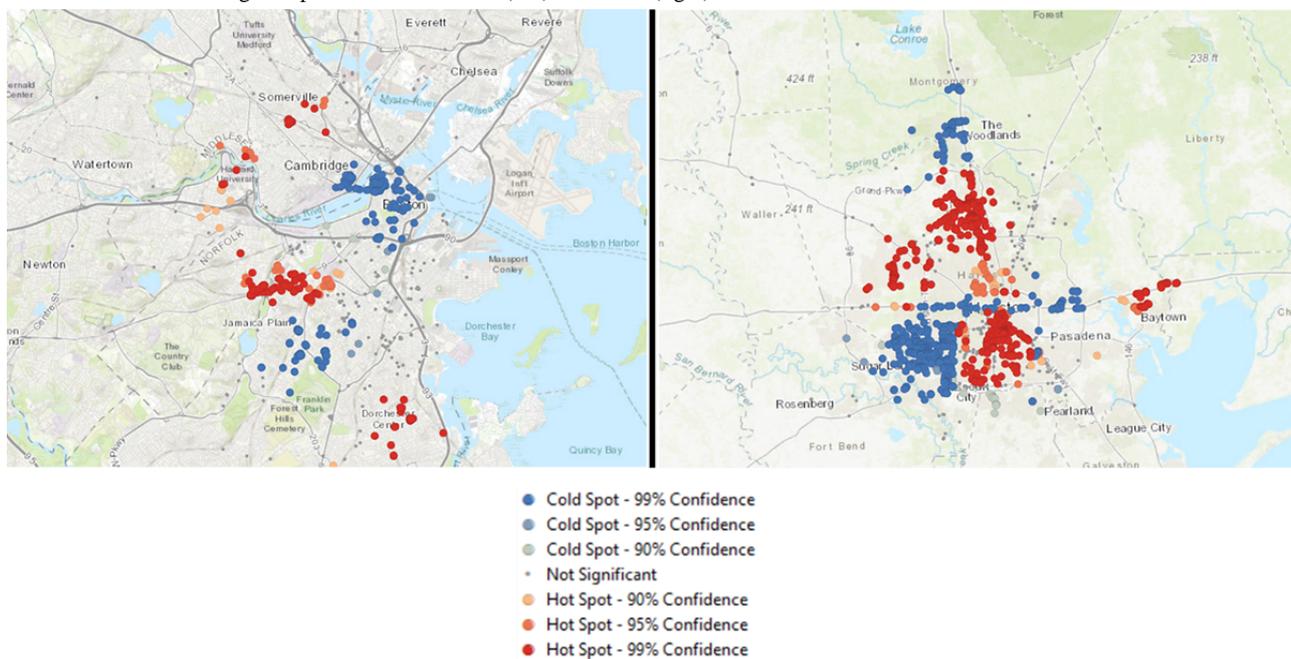
In Texas, there were 63 antitobacco and 43 protobacco messages (Multimedia Appendix 1). The top three highest exposures to antitobacco messages in Texas were (1) others (21/63, 33.3%), (2) television (14/63, 22.2%), and (3) store sign or display (7/63, 11.1%). The top 3 highest exposures to protobacco messages were on (1) store sign or display (25/43, 58.1%), (2) others (7/43, 16.3%), and (3) television (4/43, 9.3%).

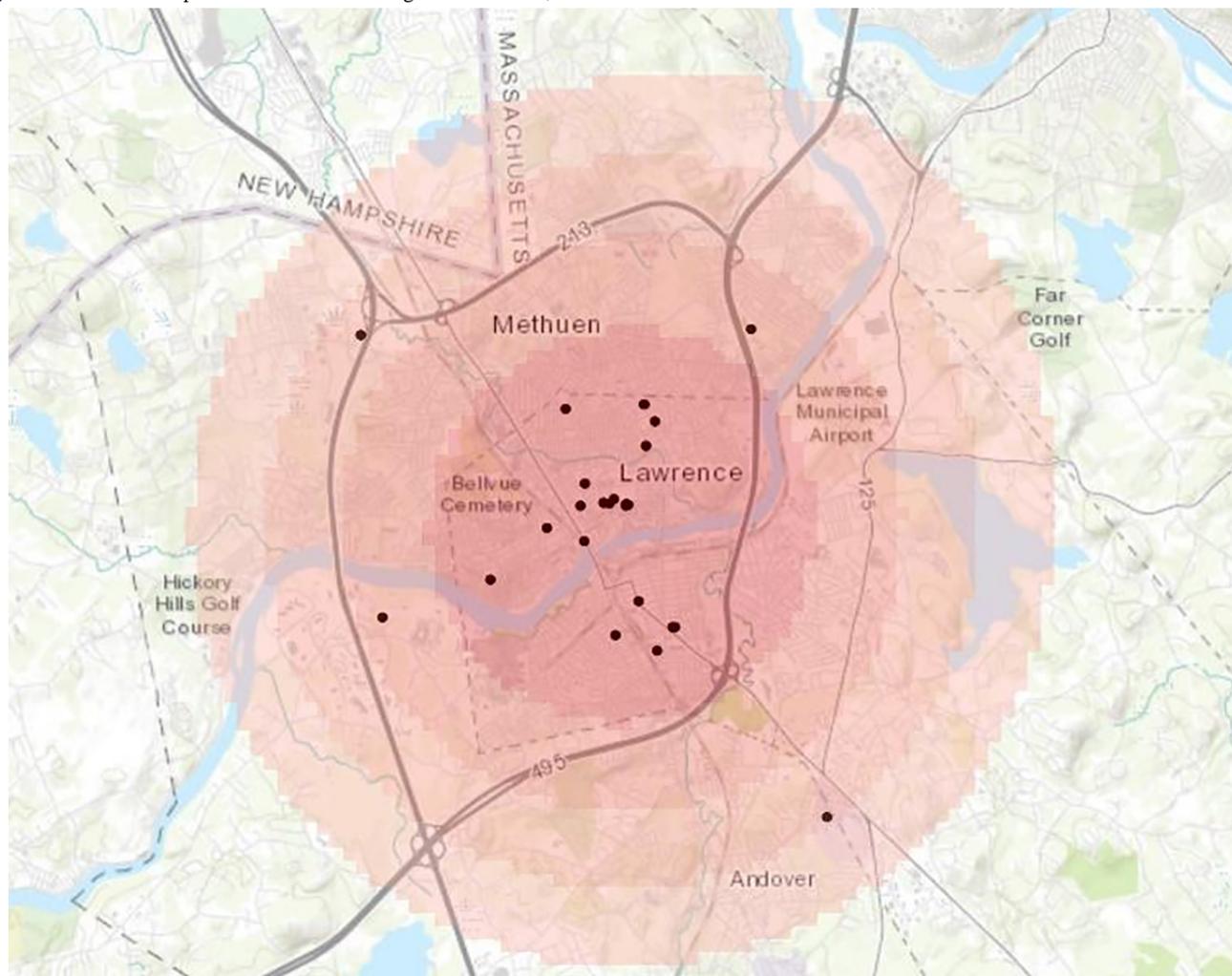
Finally, we aimed to examine if there were spatial clustering of tobacco messages and if such clusters were located near TROs or smoking hot spots. We analyzed the data using spatial autocorrelation, and the results suggested that there was evidence of antitobacco message clustering in Massachusetts (global Moran index=0.28;  $z=1.89$ ;  $P=.06$ ) but not in Texas (global Moran index=-0.12;  $z=0.73$ ;  $P=.07$ ). We then conducted an optimized hot spot analysis for antitobacco messages in Massachusetts, and the results showed that the clustering of antitobacco messages ( $z=3.85$ ;  $P<.001$ ) only occurred in Lawrence in Massachusetts (Figure 6). There was no evidence of protobacco message clusters.

**Figure 4.** Cumulative mean of power-law analysis for traffic of tobacco retail outlets in Massachusetts.



**Figure 5.** Clusters of smoking hot spots in Massachusetts (left) and Texas (right).



**Figure 6.** Clusters of reported antitobacco messages in Lawrence, Massachusetts.

## Discussion

This study showcases how data from smartphone apps could significantly inform tobacco control communication efforts when used to complement existing data sources, such as the geolocations of TROs in our study obtained from NAICS [29]. Using these methods, we were able to identify the locations where our participants were exposed to tobacco messages (anti- and pro-), the specific TROs that attract the highest level of patrons, as well as areas where individuals were most likely to smoke. There are several notable findings from our results. The data showed that physical locations still matter more than online tobacco messages when considering where people were most likely to encounter external cues for tobacco use, such as locations of TROs and areas where protobacco messaging were reported. From NAICS data, there was a concentration of TROs in economically disadvantaged areas within Boston, such as the Dorchester Center and Jamaica Plain [30]. In Texas, our data showed that there was a high concentration of TROs across Houston.

The data from smartphones complement traditional tobacco surveillance data, such as population health surveys [31], in that smartphone data provide context to how access to tobacco products and exposure to marketing and promotional efforts may influence tobacco use behavior within underserved

communities. For example, we found that in both Massachusetts and Texas, participants reported that they predominantly encountered protobacco messages at store signs or displays as compared with web-based sources. This is somewhat surprising considering the increasing concern about the influence of social media posts in the promotion of tobacco, either through user-generated content on social media or through targeted industry web-based advertising efforts [32,33]. In contrast, although our participants reported encountering antitobacco messages through web-based and offline sources, they were most likely to come across antitobacco messages on television. In addition to airing messages on mainstream media, public health officials should consider boosting efforts in placing antitobacco messages around TROs.

Another finding is that in both Massachusetts and Texas, participants reported encountering fewer antitobacco messages in newspapers or magazines as compared with protobacco messages. This is consistent with the findings from a recent study [34], which aimed to determine the extent of exposure to federal court-ordered antismoking advertisements—where tobacco companies were required to pay for these advertisements to correct smoking misinformation [35]—among a nationally representative sample of the adult population in the United States in 2018. The study found that the overall estimated exposure to antismoking advertisements was generally low

(40.6%), with the lowest exposure rates found among people aged 18 to 34 years (37.4%), those who had high school education or less (34.5%), those who earned less than US \$35,000 annually (37.5%), and Hispanic smokers (42.2%). Although it was difficult to definitively pinpoint why our participants reported low exposure to antitobacco advertisements in newspapers, one plausible reason was that young people such as those in our sample may not be using print newspapers and magazines as much as the internet and social media [36,37], and thus, they would be less likely to come across antitobacco messages across traditional media platforms. In addition, research has documented that people from underserved communities were less likely to use newspapers as their primary news sources as compared with individuals from higher SEPs [38].

Second, this type of smartphone data collection allows one to target strategic areas for antitobacco message placement. For instance, in the state of Massachusetts, there was evidence of antitobacco messages only in Lawrence, which traditionally has a higher percentage of adult smokers and TROs per 1000 adults as compared with other parts of Massachusetts [39]. Although this was a positive step, there was a need for broader dissemination of antitobacco messages to reach other areas where popular TRO hot spots were found (Dorchester Center, Jamaica Plain, and Hyde Park), specific TROs (Roxbury and Methuen) with highest human traffic, as well as areas where smoking was concentrated (Dorchester Center, Roxbury Crossing, and Peabody).

Third, it is evident that the use of smartphone data to inform antitobacco messaging efforts for underserved communities is not a magic pill solution, as it would need concurrent *supply side* tobacco control regulations to be most effective. In Houston, the widespread prevalence of TROs remained problematic for targeted antitobacco messaging to be efficacious. In other words, effective and targeted antitobacco messaging in Texas would need to be accompanied by concurrent supply side solutions, such as restricting the number of TROs or increasing tobacco taxes.

Despite the study's significant strengths, there are limitations. First, we relied on a small sample of individuals from underserved communities, and the results would not be generalizable to the overall population. For example, the locations of popular smoking areas could be heavily influenced by the characteristics of our sample. Second, as in all studies that employ smartphone apps, the geolocations were only captured when the smartphones were operational. Third, this

methodology does not guarantee that exposure to all antitobacco messages is captured. Participants might not be able to snap a picture of the antitobacco message on a billboard in time if they were driving or traveling in a car. Finally, we recognized that, similar to many smartphone tracking studies, there are issues pertaining to privacy because of the amount of data collected that may not relate directly to the study's objectives. Considering that we were working with underserved communities that were arguably more vulnerable than the general population, we prioritized the privacy protection of participants from the beginning of this study and took significant steps in communicating with our participants the privacy protection measures we have implemented.

At the policy and system architecture level, Ethica was built to be compliant with the General Data Protection Regulation requirements, which extended data protection for different types of health data collected from individuals [18,40]. In other words, our participants had the right to access and delete their own data. If the participants did not have the technical skills to do so, Ethica would provide technical support as needed. In addition, Ethica allowed the participants to request a copy of their data, and the support staff would provide them with a machine-readable file containing all the data collected about them. On a practical level, Ethica was designed in such a way that participants could snooze their study participation for some time. For instance, there was an incognito function where participants could pause data collection (eg, tracking of their geolocations) at any time they wanted.

Despite these limitations, this study presents a novel way of integrating passive and active data from smartphones with traditional tobacco surveillance information to help inform tobacco control efforts within underserved communities. We recommend that public health researchers continue to explore how to capitalize on big data from smartphones for tobacco control. For instance, future studies could extend our study by recruiting a larger sample of participants from different states and examining how fluctuations in emotions (captured by the EMA) could play a role in influencing tobacco use. Future research could also design smartphone-based interventions examining the optimal locations and time to administer antitobacco messages to people from underserved communities. In conclusion, smartphone data can inform tobacco control efforts in a powerful way, and health organizations and public health researchers should take advantage of this data revolution to strengthen tobacco control efforts to benefit the health of underserved communities [20].

---

## Acknowledgments

This work was funded by the *Truth Initiative* (KV, principal investigator). The authors would like to thank the study coordinators from both Massachusetts (Carmenza Bruff) and Texas (Amanda Zhao and Anna Kimutis) for their efforts in recruitment and data collection as well as Mohammad Hashemian from *Ethica* for app-related technical assistance. In addition, the authors would like to acknowledge Dr Jill Kelly and Dr Matthew Wilson for providing technical guidance on ArcMap through the annual Geographical Information Systems Summer Institute 2019 organized by the Center for Geographic Analysis at Harvard University.

---

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Tobacco messages reported in Massachusetts (Table 1) and Texas (Table 2).

[\[DOCX File , 16 KB-Multimedia Appendix 1\]](#)

## References

1. Fu SH, Jha P, Gupta PC, Kumar R, Dikshit R, Sinha D. Geospatial analysis on the distributions of tobacco smoking and alcohol drinking in India. *PLoS One* 2014;9(7):e102416 [FREE Full text] [doi: [10.1371/journal.pone.0102416](https://doi.org/10.1371/journal.pone.0102416)] [Medline: [25025379](https://pubmed.ncbi.nlm.nih.gov/25025379/)]
2. Centers for Disease Control and Prevention. Health Effects of Cigarette Smoking URL: [https://www.cdc.gov/tobacco/data\\_statistics/fact\\_sheets/health\\_effects/effects\\_cig\\_smoking/index.htm](https://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/effects_cig_smoking/index.htm) [accessed 2020-05-12]
3. Maloney EK, Cappella JN. Does vaping in e-cigarette advertisements affect tobacco smoking urge, intentions, and perceptions in daily, intermittent, and former smokers? *Health Commun* 2016;31(1):129-138. [doi: [10.1080/10410236.2014.993496](https://doi.org/10.1080/10410236.2014.993496)] [Medline: [25758192](https://pubmed.ncbi.nlm.nih.gov/25758192/)]
4. Centers for Disease Control and Prevention. 2019. Current Cigarette Smoking Among Adults in the United States URL: [https://www.cdc.gov/tobacco/data\\_statistics/fact\\_sheets/adult\\_data/cig\\_smoking/index.htm#states](https://www.cdc.gov/tobacco/data_statistics/fact_sheets/adult_data/cig_smoking/index.htm#states) [accessed 2020-04-23]
5. Sakuma KK, Felicitas-Perkins JQ, Blanco L, Fagan P, Pérez-Stable EJ, Pulvers K, et al. Tobacco use disparities by racial/ethnic groups: California compared to the United States. *Prev Med* 2016 Oct;91:224-232 [FREE Full text] [doi: [10.1016/j.ypmed.2016.08.035](https://doi.org/10.1016/j.ypmed.2016.08.035)] [Medline: [27575315](https://pubmed.ncbi.nlm.nih.gov/27575315/)]
6. Tucker-Seeley RD, Bezold CP, James P, Miller M, Wallington SF. Retail pharmacy policy to end the sale of tobacco products: what is the impact on disparity in neighborhood density of tobacco outlets? *Cancer Epidemiol Biomarkers Prev* 2016 Sep;25(9):1305-1310 [FREE Full text] [doi: [10.1158/1055-9965.EPI-15-1234](https://doi.org/10.1158/1055-9965.EPI-15-1234)] [Medline: [27302724](https://pubmed.ncbi.nlm.nih.gov/27302724/)]
7. Lee J, Sun D, Schleicher N, Ribisl K, Luke D, Henriksen L. Inequalities in tobacco outlet density by race, ethnicity and socioeconomic status, 2012, USA: results from the ASPIRE study. *J Epidemiol Community Health* 2017 May;71(5):487-492 [FREE Full text] [doi: [10.1136/jech-2016-208475](https://doi.org/10.1136/jech-2016-208475)] [Medline: [28249990](https://pubmed.ncbi.nlm.nih.gov/28249990/)]
8. Fakunle D, Milam A, Furr-Holden C, Butler J, Thorpe R, LaVeist T. The inequitable distribution of tobacco outlet density: the role of income in two black mid-Atlantic geopolitical areas. *Public Health* 2016 Jul;136:35-40 [FREE Full text] [doi: [10.1016/j.puhe.2016.02.032](https://doi.org/10.1016/j.puhe.2016.02.032)] [Medline: [27076440](https://pubmed.ncbi.nlm.nih.gov/27076440/)]
9. Bekalu MA, Minsky S, Viswanath K. Beliefs about smoking-related lung cancer risk among low socioeconomic individuals: the role of smoking experience and interpersonal communication. *Glob Health Promot* 2019 Sep;26(3):88-93. [doi: [10.1177/1757975917732758](https://doi.org/10.1177/1757975917732758)] [Medline: [29110569](https://pubmed.ncbi.nlm.nih.gov/29110569/)]
10. Cantrell J, Anesetti-Rothermel A, Pearson JL, Xiao H, Vallone D, Kirchner TR. The impact of the tobacco retail outlet environment on adult cessation and differences by neighborhood poverty. *Addiction* 2015 Jan;110(1):152-161 [FREE Full text] [doi: [10.1111/add.12718](https://doi.org/10.1111/add.12718)] [Medline: [25171184](https://pubmed.ncbi.nlm.nih.gov/25171184/)]
11. Barnes R, Foster SA, Pereira G, Villanueva K, Wood L. Is neighbourhood access to tobacco outlets related to smoking behaviour and tobacco-related health outcomes and hospital admissions? *Prev Med* 2016 Jul;88:218-223. [doi: [10.1016/j.ypmed.2016.05.003](https://doi.org/10.1016/j.ypmed.2016.05.003)] [Medline: [27154350](https://pubmed.ncbi.nlm.nih.gov/27154350/)]
12. Shortt NK, Tisch C, Pearce J, Richardson EA, Mitchell R. The density of tobacco retailers in home and school environments and relationship with adolescent smoking behaviours in Scotland. *Tob Control* 2016 Jan;25(1):75-82 [FREE Full text] [doi: [10.1136/tobaccocontrol-2013-051473](https://doi.org/10.1136/tobaccocontrol-2013-051473)] [Medline: [25370699](https://pubmed.ncbi.nlm.nih.gov/25370699/)]
13. Cantrell J, Pearson JL, Anesetti-Rothermel A, Xiao H, Kirchner TR, Vallone D. Tobacco retail outlet density and young adult tobacco initiation. *Nicotine Tob Res* 2016 Feb;18(2):130-137 [FREE Full text] [doi: [10.1093/ntr/ntv036](https://doi.org/10.1093/ntr/ntv036)] [Medline: [25666816](https://pubmed.ncbi.nlm.nih.gov/25666816/)]
14. Loomis BR, Kim AE, Busey AH, Farrelly MC, Willett JG, Juster HR. The density of tobacco retailers and its association with attitudes toward smoking, exposure to point-of-sale tobacco advertising, cigarette purchasing, and smoking among New York youth. *Prev Med* 2012 Nov;55(5):468-474. [doi: [10.1016/j.ypmed.2012.08.014](https://doi.org/10.1016/j.ypmed.2012.08.014)] [Medline: [22960255](https://pubmed.ncbi.nlm.nih.gov/22960255/)]
15. Businelle MS, Ma P, Kendzor DE, Frank SG, Vidrine DJ, Wetter DW. An ecological momentary intervention for smoking cessation: evaluation of feasibility and effectiveness. *J Med Internet Res* 2016 Dec 12;18(12):e321 [FREE Full text] [doi: [10.2196/jmir.6058](https://doi.org/10.2196/jmir.6058)] [Medline: [27956375](https://pubmed.ncbi.nlm.nih.gov/27956375/)]
16. Baskerville NB, Struik LL, Dash D. Crush the crave: development and formative evaluation of a smartphone app for smoking cessation. *JMIR Mhealth Uhealth* 2018 Mar 2;6(3):e52 [FREE Full text] [doi: [10.2196/mhealth.9011](https://doi.org/10.2196/mhealth.9011)] [Medline: [29500157](https://pubmed.ncbi.nlm.nih.gov/29500157/)]
17. Pew Research Center. 2019. Mobile Fact Sheet URL: <https://www.pewinternet.org/fact-sheet/mobile/> [accessed 2020-04-23]
18. Lee EW, Yee AZ. Toward data sense-making in digital health communication research: why theory matters in the age of big data. *Front Commun* 2020 Feb 27;5(11):1-10. [doi: [10.3389/fcomm.2020.00011](https://doi.org/10.3389/fcomm.2020.00011)]
19. Onnela J, Rauch SL. Harnessing smartphone-based digital phenotyping to enhance behavioral and mental health. *Neuropsychopharmacology* 2016 Jun;41(7):1691-1696 [FREE Full text] [doi: [10.1038/npp.2016.7](https://doi.org/10.1038/npp.2016.7)] [Medline: [26818126](https://pubmed.ncbi.nlm.nih.gov/26818126/)]

20. Lee EW, Viswanath K. Big data in context: addressing the twin perils of data absenteeism and chauvinism in the context of health disparities research. *J Med Internet Res* 2020 Jan 7;22(1):e16377 [FREE Full text] [doi: [10.2196/16377](https://doi.org/10.2196/16377)] [Medline: [31909724](https://pubmed.ncbi.nlm.nih.gov/31909724/)]
21. Veinot T, Mitchell H, Ancker J. Good intentions are not enough: how informatics interventions can worsen inequality. *J Am Med Inform Assoc* 2018 Aug 1;25(8):1080-1088. [doi: [10.1093/jamia/ocy052](https://doi.org/10.1093/jamia/ocy052)] [Medline: [29788380](https://pubmed.ncbi.nlm.nih.gov/29788380/)]
22. Harris NL, Goldman E, Gabris C, Nordling J, Minnemeyer S, Ansari S, et al. Using spatial statistics to identify emerging hot spots of forest loss. *Environ Res Lett* 2017 Feb 7;12(2):024012. [doi: [10.1088/1748-9326/aa5a2f](https://doi.org/10.1088/1748-9326/aa5a2f)]
23. Truth Initiative. 2019. Tobacco Use in Texas 2019 URL: <https://truthinitiative.org/research-resources/smoking-region/tobacco-use-texas-2019> [accessed 2020-04-23]
24. Truth Initiative. 2019. Tobacco Use in Massachusetts 2019 URL: <https://truthinitiative.org/research-resources/smoking-region/tobacco-use-massachusetts-2019> [accessed 2020-04-23]
25. Haining R. Spatial autocorrelation. In: Smelser NJ, Baltes PB, editors. *International Encyclopedia of the Social & Behavioral Sciences*. Toronto, Canada: Elsevier; 2015:105-110.
26. Tsai P, Lin M, Chu C, Perng C. Spatial autocorrelation analysis of health care hotspots in Taiwan in 2006. *BMC Public Health* 2009 Dec 14;9(1):464 [FREE Full text] [doi: [10.1186/1471-2458-9-464](https://doi.org/10.1186/1471-2458-9-464)] [Medline: [20003460](https://pubmed.ncbi.nlm.nih.gov/20003460/)]
27. Lipperman-Kreda S, Morrison C, Grube JW, Gaidus A. Youth activity spaces and daily exposure to tobacco outlets. *Health & Place* 2015 Jul;34:30-33. [doi: [10.1016/j.healthplace.2015.03.013](https://doi.org/10.1016/j.healthplace.2015.03.013)] [Medline: [25879915](https://pubmed.ncbi.nlm.nih.gov/25879915/)]
28. Clauset A, Shalizi CR, Newman ME. Power-law distributions in empirical data. *SIAM Rev* 2009 Nov 4;51(4):661-703. [doi: [10.1137/070710111](https://doi.org/10.1137/070710111)]
29. Watkins KL, Regan SD, Nguyen N, Businelle MS, Kendzor DE, Lam C, et al. Advancing cessation research by integrating EMA and geospatial methodologies: associations between tobacco retail outlets and real-time smoking urges during a quit attempt. *Nicotine Tob Res* 2014 May;16(Suppl 2):S93-101 [FREE Full text] [doi: [10.1093/ntr/ntt135](https://doi.org/10.1093/ntr/ntt135)] [Medline: [24057995](https://pubmed.ncbi.nlm.nih.gov/24057995/)]
30. Boston Planning & Development Agency. 2014. Poverty in Boston URL: <http://www.bostonplans.org/getattachment/f1ecaf8a-d529-40b6-a9bc-8b4419587b86> [accessed 2020-04-23]
31. Vereen RN, Westmaas JL, Bontemps-Jones J, Jackson K, Alcaraz KI. Trust of information about tobacco and e-cigarettes from health professionals versus tobacco or electronic cigarette companies: differences by subgroups and implications for tobacco messaging. *Health Commun* 2020 Jan;35(1):89-95. [doi: [10.1080/10410236.2018.1544875](https://doi.org/10.1080/10410236.2018.1544875)] [Medline: [30422690](https://pubmed.ncbi.nlm.nih.gov/30422690/)]
32. Allem J, Dharmapuri L, Leventhal A, Unger J, Cruz TB. Hookah-related posts to Twitter from 2017 to 2018: thematic analysis. *J Med Internet Res* 2018 Nov 19;20(11):e11669 [FREE Full text] [doi: [10.2196/11669](https://doi.org/10.2196/11669)] [Medline: [30455162](https://pubmed.ncbi.nlm.nih.gov/30455162/)]
33. Chu K, Colditz JB, Primack BA, Shensa A, Allem J, Miller E, et al. JUUL: spreading online and offline. *J Adolesc Health* 2018 Nov;63(5):582-586 [FREE Full text] [doi: [10.1016/j.jadohealth.2018.08.002](https://doi.org/10.1016/j.jadohealth.2018.08.002)] [Medline: [30348280](https://pubmed.ncbi.nlm.nih.gov/30348280/)]
34. Chido-Amajuoyi OG, Yu RK, Agaku I, Shete S. Exposure to court-ordered tobacco industry antismoking advertisements among US adults. *JAMA Netw Open* 2019 Jul 3;2(7):e196935 [FREE Full text] [doi: [10.1001/jamanetworkopen.2019.6935](https://doi.org/10.1001/jamanetworkopen.2019.6935)] [Medline: [31298713](https://pubmed.ncbi.nlm.nih.gov/31298713/)]
35. Mathias T. Reuters. 2019. Tobacco Industry Anti-smoking Ads Reached Less Than Half of US Adults URL: <https://www.reuters.com/article/us-health-tobacco-industry/tobacco-industry-anti-smoking-ads-reached-less-than-half-of-u-s-adults-idUSKCN1V91V2> [accessed 2020-04-23]
36. Lee EW, Ho SS, Lwin MO. Explicating problematic social network sites use: a review of concepts, theoretical frameworks, and future directions for communication theorizing. *New Media Soc* 2016 Oct 6;19(2):308-326. [doi: [10.1177/1461444816671891](https://doi.org/10.1177/1461444816671891)]
37. Lee EW, Ho SS, Lwin MO. Extending the social cognitive model—examining the external and personal antecedents of social network sites use among Singaporean adolescents. *Comput Human Behav* 2017 Feb;67:240-251. [doi: [10.1016/j.chb.2016.10.030](https://doi.org/10.1016/j.chb.2016.10.030)]
38. Lee EW, Ho SS. The perceived familiarity gap hypothesis: examining how media attention and reflective integration relate to perceived familiarity with nanotechnology in Singapore. *J Nanopart Res* 2015 May 22;17(5):1-15. [doi: [10.1007/s11051-015-3036-z](https://doi.org/10.1007/s11051-015-3036-z)]
39. Department of Public Health Massachusetts. Tobacco Community Fact Sheet. Make Smoking History. 2019. URL: <https://makesmokinghistory.org/wp-content/themes/makesmokinghistory/tafi/pdf/TAFI-Lawrence.pdf>
40. Ethica. 2020. Is Ethica GDPR Compliant? URL: <https://learn.ethicadata.com/faq/is-ethica-gdpr-compliant/> [accessed 2020-04-23]

## Abbreviations

- e-cigarette:** electronic cigarette
- GED:** General Education Diploma
- IRB:** institutional review board
- NAICS:** North American Industry Classification System
- SEP:** socioeconomic position

**TRO:** tobacco retail outlet

*Edited by G Eysenbach; submitted 13.12.19; peer-reviewed by E Mwashuma, I Contreras; comments to author 11.03.20; revised version received 20.03.20; accepted 21.03.20; published 07.07.20*

*Please cite as:*

*Lee EWJ, Bekalu MA, McCloud R, Vallone D, Arya M, Osgood N, Li X, Minsky S, Viswanath K*

*The Potential of Smartphone Apps in Informing Protobacco and Antitobacco Messaging Efforts Among Underserved Communities: Longitudinal Observational Study*

*J Med Internet Res 2020;22(7):e17451*

URL: <https://www.jmir.org/2020/7/e17451>

doi: [10.2196/17451](https://doi.org/10.2196/17451)

PMID: [32673252](https://pubmed.ncbi.nlm.nih.gov/32673252/)

©Edmund WJ Lee, Mesfin Awoke Bekalu, Rachel McCloud, Donna Vallone, Monisha Arya, Nathaniel Osgood, Xiaoyan Li, Sara Minsky, Kasisomayajula Viswanath. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 07.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.