RESEARCH ARTICLE

CrossMark

Individual Mobility and Uncertain Geographic Context: Real-time Versus Neighborhood Approximated Exposure to Retail Tobacco Outlets Across the US

Thomas R. Kirchner, et al. *[full author details at the end of the article]*

Abstract

There is growing interest in the way exposure to neighborhood risk and protective factors affects the health of residents. Although multiple approaches have been reported, empirical methods for contrasting the spatial uncertainty of exposure estimates are not well established. The objective of this paper was to contrast real-time versus neighborhood approximated exposure to the landscape of tobacco outlets across the contiguous US. A nationwide density surface of tobacco retail outlet locations was generated using kernel density estimation (KDE). This surface was linked to participants' ($N_p = 363$) inferred residential location, as well as to their real-time geographic locations, recorded every 10 min over 180 days. Real-time exposure was estimated as the hourly product of radius of gyration and average tobacco outlet density ($N_{hour} = 304$, 164 h). Ordinal logit modeling was used to assess the distribution of real-time exposure estimates as a function of each participant's residential exposure. Overall, 61.3% of real-time, hourly exposures were of relatively low intensity, and after controlling for temporal and seasonal variation, 72.8% of the variance among these low-level exposures was accounted for by residence in one of the two lowest residential exposure quintiles. Most moderate to high intensity exposures (38.7% of all real-time, hourly exposures) were no more likely to have been contributed by subjects from any single residential exposure cluster than another. Altogether, 55.2% of the variance in real-time exposures was *not* explained by participants' residential exposure cluster. Calculating hourly exposure estimates made it possible to directly contrast real-time observations with static residential exposure estimates. Results document the substantial degree that real-time exposures can be misclassified by residential approximations, especially in residential areas characterized by moderate to high retail density levels.

Keywords Spatial uncertainty · Exposure science · Human mobility · Urban computing · Retail environment

1 Introduction

Geographic location provides a relational connection between individuals, their health behaviors, and a rapidly expanding array of data on neighborhood structure and

composition [1–3]. Once hierarchically integrated, it becomes possible to systematically investigate the impact of a wide range of neighborhood socio-ecological covariates on indicators of individual health and well-being [4–7]. The geographic landscape of retail tobacco products provides an excellent example [8–19]. A considerable body of empirical evidence demonstrates the ways point-of-sale tobacco marketing influence tobacco users' product preferences as well as decisions to initiate or refrain from use [8, 9, 11, 20–26], and that targeted point-of-sale tobacco (POST) marketing tactics sustain tobacco-related health disparities [12, 27–32]. Studies have also focused on the lowest displayed pack price, promotions, placement for leading brands, and flavor descriptors. More recent work has found that cigarette packs cost less at retailers located near public schools than those near private schools [11], and that outlet proximity to schools and parks is linked to advertising practices and illicit youth sales [12].

Despite a proliferation of studies that seek to identify and study mechanisms linking neighborhood environments to the health-related decision-making and behavior of residents, [2, 33–36] a number of unresolved conceptual issues persist, not least the definition of the ecological units of analysis themselves. Because there are no naturally occurring units of neighborhood, the neighborhood areal units employed by researchers are often arbitrary or linked to confounding geographic circumstances, [37] undermining their utility for geographically-explicit health research. [2] The ideal neighborhood unit would represent the "true causally relevant spatial context," [1] however, so long as the true spatial context remains undefined, we are left with what Kwan has termed the 'uncertain geographic context problem' (UGCoP). [38] Inferential errors brought about by the use of arbitrary methods for delineating neighborhoods likely contribute to the misclassification of effects and inconsistent evidence for socio-ecological relationships and health behavior across the research literature. [38–41]

Residential location remains the most readily available geographic linkage for integrating individual and neighborhood data. Conventionally, researchers join residential addresses directly to the administrative zones they fall within (e.g., US Census polygons) or draw new 'ego-centric' buffer zones around those addresses [4]. These residential areal units act as spatially and temporally invariant containers for various neighborhood risk and protective factors, each of which can then be linked to residents themselves. Data on individual mobility patterns improve upon the assumptions that administrative zones or ego-centric buffers make about the propensity for individuals to come into contact with environmental risk factors. Mobility data capture actual travel patterns, including each person's ever-accumulating set of origins and destinations, allowing for precise estimation of exposure to environmental risk factors over time. Dynamic conceptualizations of neighborhood exposures may enrich our understanding of health and place by better reflecting the true causally relevant spatial context of exposures to environmental risk in the population [6, 7, 42–44].

This paper presents a direct comparison of real time versus neighborhood approximated (residential address; [4]) exposure to a well-known environmental risk factor, the spatial distribution of tobacco outlets across the contiguous United States [19, 45–47]. This comparison provides an empirical framework for contrasting relative uncertainty in spatial exposures and estimating the likelihood of their misclassification. Residential and real-time exposures were computed for a sample of 363 people, each of whom voluntarily recorded their location every 10 min for 180 days using their cellular phone. Analyses use a common density-based exposure metric to examine the extent to

which variation in exposure levels observed in real-time could be explained by exposure estimates approximated solely from residential location. By partitioning variation in exposure due to routine patterns of mobility, we specifically tackle the first of Chen and Kwan's (2015) three key dimensions for research on retail access: contextual uncertainty due to real-time mobility patterns.

2 Methods

For clarity of presentation the Methods are organized into three sections: 2.1 Data Preparation, 2.2 Tobacco Outlet Exposure Estimation, and 2.3 Statistical Modeling. Section 2.2 describes the calculation of home and real-time (hourly) exposure estimates, and Section 2.3 describes the statistical analysis framework, which is a special case of multinomial logistic regression, providing a framework for interpretation of the model estimation process. Agresti [48] shows the direct correspondence between the log-linear and logit modeling frameworks, both in mathematical and inferential terms, and we present an application of this approach here.

2.1 Data Preparation

2.1.1 Longitudinal Human Mobility Data

Longitudinal human mobility data was continuously recorded via the OpenPaths (https://openpaths.cc) platform, launched by the New York Times Company Research and Development Lab in May 2011 [49]. Participants who provided consent to participate in this research and downloaded the OpenPaths application for either iOS or Android were then able to use the application to continuously capture their current geographic location—i.e., latitude and longitude, wirelessly uploaded according to a 10-min sampling rate. When participants remained stationary (less than 5 m location change between successive coordinates), the OpenPaths application suspended data collection to save battery, ensuring that OpenPaths data represents periods of at least minimal mobile activity rather than extended periods of rest, such as while people sleep at night or sit for long periods at their desks. While stationary, the application continued to "listen" for a greater than 5-m location change and reinitiated continuous tracking whenever such a change was detected.

    The raw longitudinal mobility dataset used for the present analysis contained 8,458,902 observations collected from 859 individuals worldwide from 01/01/2012 to 06/06/2015. Each observation included a unique user ID, date, time, latitude, and longitude. The raw mobility data were initially clipped to the United States using state outline polygons published by United States Census Bureau, yielding a US cohort of 744 individuals with 4,647,152 observations. Data from the lower 48 states and Washington, DC were then clipped to correspond with the tobacco density layer (Section 2.2), yielding 729 individuals with 4,556,619 observations. Participants tracked their location over a mean of 267 days (Median = 178; SD = 278), recording an average of 6251 location coordinates (Median = 3192; SD = 10,289). To standardize the amount of data contributed by each individual while maximizing the amount available for analysis, each participant's daily data records were left-truncated, retaining

their most recent 180 days of tracking data. Individuals with fewer than 180 days of data were not included in the present analyses, ensuring all between-person contrasts were based on an equivalent 180-day sampling frame. The final longitudinal mobility dataset thus included $N_p = 363$ individuals and 1,418,151 real-time observations over the 180-day study period.

2.1.2 Estimating Tobacco Outlet Density

A nationwide density surface of tobacco retail outlet locations was generated using kernel density estimation (KDE). This non-parametric method extrapolates from spatially distributed point data by estimating their continuous density using spatial density functions known as kernels, each of which has a specified circular radius size known as the bandwidth [50]. Gaussian kernels with a fixed 8047 m (5 mile) bandwidth were used to generate the final density surface, from which density estimates could be extracted with a resolution of 250 m.

The empirical basis for this probability density surface was a national dataset of tobacco retail outlets, identified by North American Industry Classification Systems (NAICS) codes. Developed by the Office of Management and Budget, NAICS is the standard used by Federal statistical agencies to classify businesses based on their primary activity [51]. In 2012, geocoded data was obtained from D&B Hoovers. The following retail categories and corresponding NAICS codes were included: beer, wine, and liquor stores (NAICS: 445310); supermarkets and other grocery stores (NAICS: 44511); convenience stores (NAICS: 44512); pharmacies and drug stores (NAICS: 446110); gasoline stations with convenience stores (NAICS: 44711); other gasoline stations (NAICS: 44719); department stores (NAICS: 452111); discount department stores (NAICS: 452112); and tobacco stores (NAICS: 453991). For pharmacies and department stores, we individually reviewed all chains with 50 or more locations to determine if they sold tobacco and excluded them accordingly [52]. Based on this analysis, we also excluded all other department stores and pharmacies as they likely do not sell tobacco. In addition, we excluded major department chains and grocery stores that, based on their store's policy, do not sell tobacco products (i.e., Target, Whole Foods, Trader Joe's, Wegmans). We also excluded pharmacies and drug stores in the 55 Massachusetts and 2 California municipalities that have banned the sale of tobacco products within these establishments. The final dataset comprised $N = 269{,}781$ retail outlets.

2.2 Tobacco Outlet Exposure Estimation

2.2.1 Residential Exposure

Residential locations were extracted from participants' OpenPaths mobility data using a two-step procedure based on Toole et al. [53], and originally given by Zheng and Xie [54]. In the first step, meaningful locations—stay events—were extracted from mobility data for each participant for each nighttime period (8 pm–7 am). A second step then combines each participant's set of nighttime stay events covering the entire 180 day analysis period to a set of aggregate stay points. The aggregate stay point that comprises the most nighttime stay events for each participant was selected as the maximally

unbiased estimator of residential location. Residential retail exposures were then generated by spatial joining each participant's inferred residential location to the density surface of tobacco retail outlet locations (Mean = 9.50, Median = 3.81, SD = 12.25). Exposures were square root transformed owing to right skew, and classified into five groups based on exposure intensity.

2.2.2 Real-Time Exposure

Participants' OpenPaths mobility data was used to compute hourly radius of gyration (R_g), measured in meters [28]. R_g estimates the size and spread of a participant's personal activity space for a given hour. R_g is defined by the standard deviation between locations and their center of mass:

$$R_g = \sqrt{\frac{1}{N} \sum_{k=1}^{N} (r_k - r_{\mathrm{mean}})^2}$$

where N is the total number of location coordinates collected from each individual per hour, and r_{mean} is the individual center of mass, or the mean longitude and latitude of all N locations. The great circle distance in meters between a specific location and the center of mass ($r_k - r_{\mathrm{mean}}$) was calculated using Vincenty's formulae [55]. The dataset was then aggregated over 465,279 hourly observations of R_g. Given the present focus on routine day-to-day mobility patterns, long-range travel was excluded by dropping observations with R_g larger than 160 km ($N_{\mathrm{drop}} = 945$), the maximum distance a car can travel in 1 h. Hours with zero movement ($R_g = 0$) were then excluded, yielding 363 individuals with a total 304,164 hourly observations (65.4% of the hourly data).

Real-time exposure was conceptualized as the product of a participant's hourly movement (R_g) and their hourly aggregate exposure to retail outlets (Mean = 22,620, Median = 7.01, SD = 144,191). Each real-time mobility coordinate contributed by the participants was joined to a tobacco outlet density value extracted from the KDE surface. Hourly exposure levels are, thus, the product of each participant's R_g within each hour under observation and the average tobacco outlet density value across the set of mobility coordinates recorded within the same hour. This exposure variable approximates the number of tobacco outlets that surrounded each participant within each hour of the study and, as expected from a count variable of this kind, the observed distribution of exposure values was heavily skewed to the right. To improve correspondence with the standard assumptions of a categorical count-based data analysis framework, both residential and hourly exposure values were square root transformed and determined to closely follow a negative binomial distribution. This produced real valued outputs in the range 0–27, which were divided into 27 clusters by binning with a one unit spacing. Each cluster effectively labels all participant hours within a particular range of exposure intensity, with subsequent clusters representing the increasing intensity of participants' hourly exposures in this sample.

To model temporal patterns of exposure, time-related variables based on social convention were derived for each observation. These were time-of-day (24-h clock),

day-of-week (weekday versus weekend), and season. *Time of day* was defined by four 6-h windows: 3:00–9:00 as "early," 9:00–15:00 as "day," 15:00–21:00 as "evening," and 21:00–3:00 as "late," treating observations recorded between midnight and 3 AM as part of the preceding day [56]. *Day-of-week* was coded binary, indicating whether each observation fell on a weekend, defined as falling after 17:00 Friday through 17:00 Sunday. *Season* was also defined as categorical, with December–February as winter, March–May as spring, June–August as summer, and September–November as fall.

2.3 Statistical Analyses

2.3.1 Log-Linear Model Selection

The interactive association between real-time exposures (27 clusters) and residential (5 clusters) exposures was stratified across time-of-day (4 categories), day-of-week (2 categories), and season (4 categories), using a set of multivariate contingency tables that populated a 5-dimensional matrix with a total of 27×5×4×2×4 = 4320 cells. Patterns of association within this large multivariate contingency table were analyzed with generalized categorical data analysis techniques [48]. Specifically, we employed an exponential, log-linear modeling framework. Log-linear models convert the multiplicative relations among joint and marginal counts in a contingency table to additive, linear associations by transforming the counts to logarithms [48]. Hierarchically nested model comparison techniques were used to iteratively identify the most parsimonious combination of factors required to explain the observed data. Systematic comparison of hierarchically nested log-linear models produced a likelihood ratio test statistic presented in the text. The saturated model represents the log frequencies for the cell index ($_{h,w,t,s,e}$) of all non-ordinal combinations of both real-time exposure and residential retail exposure (home), time-of-day, weekend, and season:

$$ln\left(\mu^{hwtse}\right) = \lambda + \lambda_h^H + \lambda_w^W + \lambda_t^T + \lambda_s^S + \lambda_e^E + \lambda_{hw}^{HW} + \cdots + \lambda_{hwt}^{HWT} + \cdots + \lambda_{hwte}^{HWTE}$$
$$+ \cdots + \lambda_{hwtse}^{HWTSE}, h$$
$$= 1, \ldots, 5, w = 1, 2, t = 1, \ldots, 4, s = 1, \ldots, 4, e = 1, \ldots, 27.$$

Where H is home exposure, T is time-of-day, W is weekend, S is season, and E is real-time exposure. h, t, w, s, and e are categories within H, T, W, S, and E. μ^{hwtse} represents the expected cell frequencies in the five-dimensional contingency table. λ is the constant. λ_h^H denotes the row effect of λ_i^H, $i \in h$. λ_w^W, λ_t^T, λ_s^S, and λ_e^E also represent row effects. λ_{hw}^{HW} is the interaction term λ_{ij}^{HW} between H and W, where $i \in h$, $j \in w$. $\lambda_{hwt}^{HWT}, \ldots, \lambda_{hwte}^{HWTE}, \ldots, \lambda_{hwtse}^{HWTSE}$ are higher dimensional interaction terms. For easier read, letter symbols are used in Table 1 to represent the highest interaction model terms.

2.3.2 Log-linear Model Interpretation: Ordinal Logit Modeling

Ordinal logit models were used to interpret observed associations within a log-linear framework, utilizing well-established logistic model reporting and interpretation

Table 1 Step-down contrasts of "best-in-class" models

Model	(Model Terms)	Deviance	df	p value	G^2	Δ df	$p(\Delta$df; 0.01)[a]
1.1	(W, T, H, S, E)	321,168.90	302,853	0.00***	1057.6	54	<0.001[b]
1.2	(WT, WH, WS, WE, TH, TS, TE, HS, HE, SE)	320,111.30	302,907	0.00***	−108,897.9	495	<0.001[b]
1.3	(HTE, WTE, HWT, HWE, HES, HTS, TSE, WSE, WTS, HWS)	429,009.20	303,402	0.00***	169,139.32	286	< 0.001[b]
1.4	(WTSE, TSHE)	259,869.88	303,688	1.00	−12,713.60	16	<0.001[b]
1.5	(WTHE, TSHE)	272,583.48	303,704	1.00	158,014.6	127	<0.001[b]
1.6	(WTHE, WSHE, TSHE)	114,568.88	303,831	1.00	50,677.26	77	<0.001[b]
1.7	(WTHE, WTSE, WSHE, TSHE)	64,396.25	303,902	1.00	15,677.26	36	<0.001[b]
1.8	(WTHE, WTSE, WTSH, WSHE, TSHE)	48,718.99	303,938	1.00	48,718.99	303,938	1.00
1.9	(WTSHE)	0.00	0	1.00	–	–	–

W weekend, *T* time of day, *S* season, *H* residential exposure cluster, *E* real-time exposure level

***Significant lack of fit at 99% confidence level

[a] p value of G^2 comparing to the model below

[b] Hierarchically nested models are significantly different at 99.9% confidence level

standards [48]. In this paper, log-linear estimation and selection was used to identify the most parsimonious model, best fitting the observed data, and then ordinal logit modeling was used to examine our primary aim: examining the extent to which variation in exposure to tobacco retail outlets observed in real-time could be explained by exposure estimates approximated as a function of residential location and time (time-of-day, day-of-week, season-of-year). Following Agresti (2012), this was accomplished by setting the residential exposure variable (home) from within the log-linear model as an ordinal dependent variable. Within this framework, each ordinal residential exposure level represented a "cluster" of participants, and model results estimate the probability—i.e., the log-odds ratio or logit—that each real-time exposure observation was contributed by a participant from each of the residential density clusters.

3 Results

Overall, mean hourly R_g was 0.55 km with a standard deviation of 2.66 km. Minimum and maximum were 0 and 152.94 km. Within each day, R_g was the lowest in the middle of the night and higher across the remainder of the day. Two spikes were observed in R_g in early morning and late afternoon on weekdays, which is likely related to commuting between home and work—while there was generally more variation on weekends. Figure 1 illustrates the daily drop in mobility across the late-night hours, followed by a steep rise across the early morning, and then divergence on Saturday and Sunday, with early Sunday afternoon revealed as the window of greatest mobility on average.
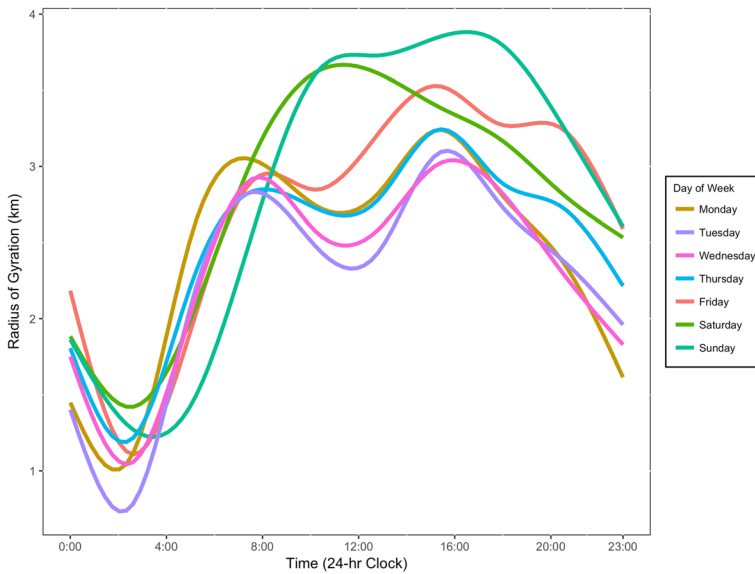
Fig. 1 Radius of Gyration Measured Mobility Patterns by Time of Day across Day of Week

Figure 2 presents generalized additive model smoothed real-time exposure intensity by time of day for each residential exposure cluster. On weekdays, real-time exposure spikes in the early morning and late afternoon across all residential exposure clusters, probably due to increased movement and exposure to retail outlets during the commute to and from work. However, this relationship becomes more prominent as residential exposure cluster increases from low to high. In contrast, weekend exposure was elevated at noon and then remained high until early evening, and was broadly consistent across residential exposure clusters, except in the high exposure cluster which deviates from the patterns observed.

3.1 Model Fitting: Best-in-class Log-linear Model Selection

Table 1 presents an overview of the best-in-class model selection process used to identify the most parsimonious model, defined as the minimal set of parameters required to provide an adequate fit to the observed data. Following hierarchically nested model comparison techniques, Model performance is evaluated by both degree of freedom (df) and likelihood ratio test statistic (G^2). p value represents lack of fit of the models, which means models with p value of 1 fit at 99% confidence level and p value of 0 indicates model not fitting. The initial basis for comparison is Model 1.9, which is the saturated model that corresponds perfectly to the raw data, having zero degrees of freedom, as the number of parameters is equivalent to the total number of cells generated by all interactive combinations of the five factors under study: retail exposure (the conceptual dependent variable), residential exposure (home), weekend, time-of-day (time), and season. Model 1.4 fits while using only 829 cells to model the total 1325 cells under study. It also accurately predicts a five-dimensional contingency table with lower level four-way interactions. This is the most
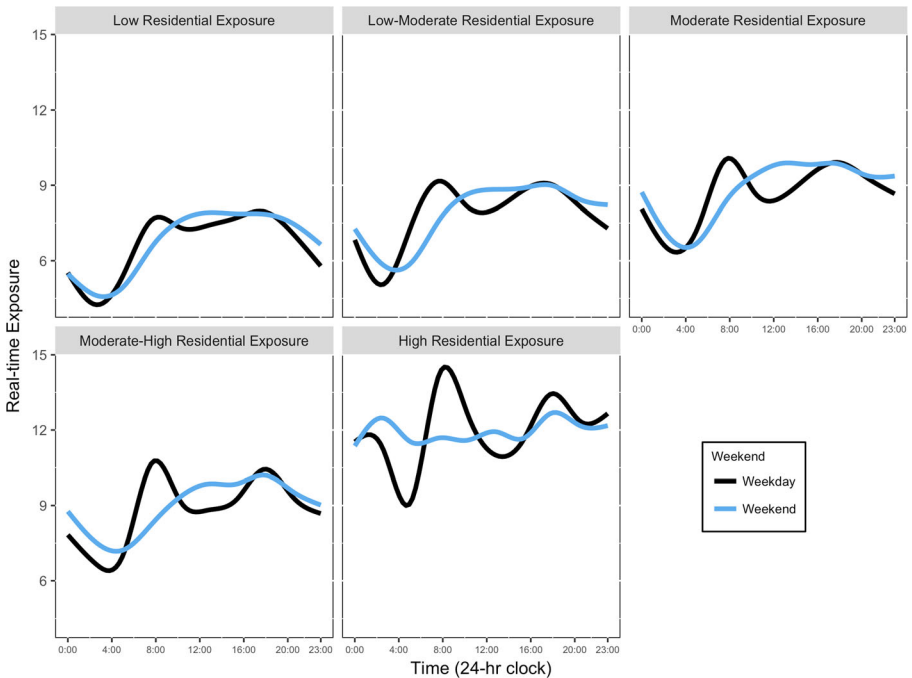
Fig. 2 Weekday and Weekend Real-time Exposure by Time of Day for Each Residential Exposure Cluster

parsimonious model, effectively isolating an informative pattern in the data that then becomes the basis for inference.

To measure the separate strength and significance of each interaction in Model 1.4, models excluding one of the two interaction terms were evaluated. Table 2 shows the likelihood ratio evaluation of the influence of the excluded term compared to Model 1.4 with all two-term four-way interactions. This method provides a specific test of conditional independence between seasonal effects and real-time exposure levels. These model fits indicate that while the interactions between weekend effects and other factors are important, their associations do not contribute in Model 1.5 as much as the interactions between residential (home) exposure, time, season, and real-time exposure.

3.2 Model Inference: Ordinal Logit Modeling of Residential Density Clusters

To examine real-time tobacco retail outlet exposures among clusters of residential exposure, an ordinal logit model was constructed to be mathematically equivalent to the most parsimonious log-linear model (see Section 2.3.2 Table 1 Model 1.4), itself identified through the model selection process described in Section 3.1:

$$logit[P(H = h|T = t, S = s, E = e)] = \alpha + \beta_t^T + \beta_s^S + \beta_e^E + \beta_{ts}^{TS} + \beta_{te}^{TE} + \beta_{se}^{SE}$$
$$+ \beta_{tse}^{TSE} + \varepsilon, h$$
$$= 1, \ldots, 5, t = 1, \ldots, 4, s = 1, \ldots, 4, e = 1, \ldots, 27.$$

where the log-odds ratio (i.e., *logit*) of residential exposure cluster membership (home) for each hour is modeled as a function of the real-time exposure level (H), time-of-day (T), and season (S) associated with each hour. α is a constant. h, t, s, and e are categories within H, T, S, and E, and ε is the error term. β_t^T, β_s^S, and β_e^E represent the effects of parameter T, S, and E respectively. β_{ts}^{TS}, …, β_{tse}^{TSE} represent the effects of interactions between parameters. For easier read, letter symbols are used in Table 2 to represent the highest interaction model terms. Model predicted residential exposure cluster membership and associated confidence intervals were generated via bootstrapping with 500 random samples of 10,000 observations, drawn with replacement from the empirical data distribution, effectively capturing the uncertainty associated with each parameter estimated by the model. This simulation-based resampling approach allowed for precise discrimination between the different residential exposure clusters (Fig. 3).

Figure 3 presents the model predicted results of this process, illustrating the predicted probability that each of the 304,164 h under study was contributed by a participant from each of the residential exposure clusters. Separation between the bootstrapped 95% confidence intervals corresponds to regions of the distribution of real-time hourly exposures that were significantly explained by one or more of the density clusters. Overall, 61.3% of real-time, hourly exposures were of relatively low intensity, and after controlling for temporal and seasonal variation, 72.8% of the variance among these low-level exposures was accounted for by residence in one of the two lowest residential density quintiles. Residence in one of the two highest residential density quintiles accounted for approximately 50% of the variance among extreme exposure levels, but extreme levels of exposure were rare, constituting about 1% of the data. Altogether 55.2% of the variance in real-time exposures was *not* explained by participants' residential exposure cluster, and most moderate to high intensity real-time exposures (38.7% of all hourly exposures) were no more likely to have been contributed by subjects from any single residential density cluster than another. In sum, OpenPaths participants experienced a heterogeneity in hourly tobacco retail outlet exposures that is only partially explained by their static residential exposures.

Table 2 Likelihood ratio evaluation: removal of four-way terms from maximally parsimonious model

Model	Hypothesis	Log-likelihood	df	G^2	Δ df	$p(\Delta \mathrm{df}; 0.01)$[a]
Maximally parsimonious model						
1.4	(WTSE, TSHE)[ψ]	−1,363,431.41	303,688	–	–	–
Conditional independence						
2.1	WTSE = 0	−2,116,700.547	303,123	1,506,538.272	565	< 0.001[b]
2.2	TSHE = 0	−2,037,315.705	303,532	1,347,768.588	156	< 0.001[b]

W: weekend, *T* time-of-day, *S* season, *H* residential exposure cluster, *E* real-time exposure level

[ψ] Model 1.4 from Table 1

[a] *p* value of G^2 comparing to the Model 1.4

[b] Significantly different at 99.9% confidence level compared with Model 1.4
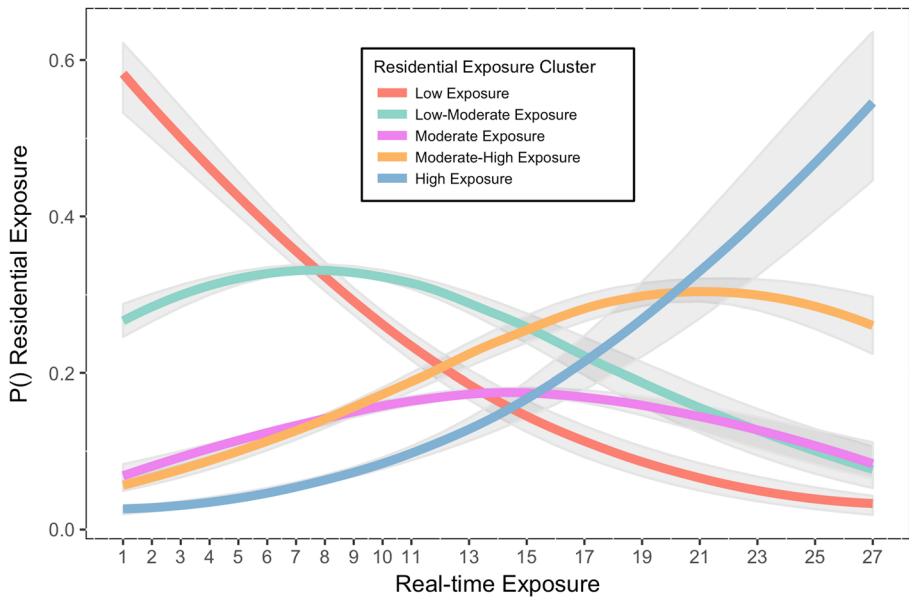
Fig. 3 Real-time Exposure for Each Residential Exposure Cluster

4 Discussion

While environmental "exposures" are most commonly thought of as biological – "internal" contact with toxic particles in the environment (e.g., air pollution and infectious pathogens) – there is growing recognition that monitoring of exposures to the broader ecosphere or "eco-exposome" is also important [57–59]. Individual-level geographic location data provide a spatial linkage that makes it possible to estimate the multivariate impact of countervailing societal and environmental systems on individual decision-making and behavior. This gives rise to the possibility of using such information for disease prevention and intervention delivery. It is our position, however, that continuous, real-time location data *need not and should not* be limited to use within real-time, "just-in-time adaptive" interventions. In fact, we believe the present paper provides an example of the way such micro real-time data can be better understood when it is aggregated, because it is only then that we can properly account for the relative significance of the various locations each participant frequents, at least as they pertain to the tobacco point-of-sale landscape.

Traditional estimates of exposure to risk and protective factors in neighborhoods are founded on the idea that the relative concentration of health-related factors around people's homes sufficiently captures and thus can be used to characterize aggregated patterns of environmental exposure within and between neighborhood areas. This paper evaluated the degree to which residential locations approximate actual exposures by comparing empirical observations collected in real-time with static neighborhood estimates that only used information about each participant's residence. Results demonstrate the utility of a continuous geolocation data smoother (i.e., R_g) that makes it possible to generate dynamic, mobility weighted KDE exposure values that retain both

spatial and temporal resolution. Findings suggest that real-time exposures are misclassified by person-level residential exposure estimates to a substantial degree, especially among people residing in areas characterized by moderate to high levels of residential density.

Essentially, results of this work indicate that exposures to moderate and high levels of tobacco outlet density were systematically less-likely than lower density exposures, and that subjects who resided in moderate to high density areas were less likely to experience real-time exposures that were as high as estimated by the observed density around their residential location. This finding suggests that residence-based neighborhood approximations exhibited a tendency to over-estimate exposure levels experienced by residents in the real-world, and somewhat counter-intuitively, that this may have been particularly true within dense urban areas, where despite high-levels of density overall, shorter travel distances among a smaller set of stores dampened observed hourly exposure levels.

This paper advances the literature in a number of ways. Focusing on a reliable, national source of tobacco outlet data allowed us to identify variation in urban dynamics and behavior across different regions of the US. The use of continuous real-time geo-location tracking provides excellent temporal and spatial resolution, which improved sensitivity to detect dynamic patterns in the data. The analysis framework developed here can be used to assess mobility patterns, exposure to points of interest, and associated effects on health behavior. Nevertheless, this study also had methodological limitations that should be considered. This sample is not nationally representative, as participation in the mobility tracking was based on self-selection, and required access to the Internet and a smartphone. Additionally, because no participant demographic information was available, other factors that could potentially affect participants' mobility, such as occupation or income level, could not be measured.

5 Conclusions

Results of this work shed light on the nature of real-time exposures to a spatially distributed environmental risk factor, as compared to a commonly used neighborhood exposure estimate. Future work should leverage methods of this kind to advance our understanding of individual decision-making and behavior change dynamics as a function of environmental conditions. Natural extensions would incorporate other policy and health-relevant risk and protective factors, such as the distribution of food, alcohol, and cannabis products. Research that involves clinical populations attempting to modify habitual health behaviors would be useful, including work with patients working to adhere to dietary restrictions or quit cigarette smoking. It will also be interesting to investigate the basic mechanisms underlying these associations, such as memory and other cognitive processes affected by regular product exposures and associated preferences. Variations across geographic areas and over time may provide insight and identify targets of intervention for public health practitioners, urban planners, and policy makers.

Compliance with Ethical Standards

Conflict of Interest   The authors declare that they have no real or potential conflicts of interest.

Publisher's Note   Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Diez Roux AV, Mair C (2010) Neighborhoods and health. Ann N Y Acad Sci 1186:125–145. https://doi.org/10.1111/j.1749-6632.2009.05333.x
2. Riva M, Gauvin L, Barnett T (2007) Toward the next generation of research into small area effects on health: a synthesis of multilevel investigations published since July 1998. J Epidemiol Community Health 61:853–861
3. Goodchild M, Janelle D (2004) Spatially integrated social science. Oxford University Press, New York
4. Duncan D, Kawachi I, Subramanian S, Aldstadt J, Melly S, Williams D (2014) Examination of how neighborhood definition influences measurements of youths' access to tobacco retailers: a methodological note on spatial misclassification. Am J Epidemiol 179(3):373–381
5. Kirchner TR, Shiffman S (2016) Spatio-temporal determinants of mental health and well-being: advances in geographically-explicit ecological momentary assessment (GEMA). Soc Psychiatry Psychiatr Epidemiol 51(9):1211–1223. https://doi.org/10.1007/s00127-016-1277-5
6. Matthews SA, Yang T-C (2013) Spatial polygamy and contextual exposures (SPACEs): promoting activity space approaches in research on place and health. Am Behav Sci 57(8):1057–1081
7. Perchoux C, Chaix B, Cummins S, Kestens Y (2013) Conceptualization and measurement of environmental exposure in epidemiology: accounting for activity space related to daily mobility. Health & Place 21:86–93. https://doi.org/10.1016/j.healthplace.2013.01.005
8. Cantrell J, Kreslake JM, Ganz O, Pearson JL, Vallone D, Anesetti-Rothermel A, Xiao H, Kirchner TR (2013) Marketing little cigars and cigarillos: advertising, price, and associations with neighborhood demographics. Am J Public Health 103(10):1902–1909. https://doi.org/10.2105/AJPH.2013.301362
9. Kirchner TR, Cantrell J, Ansetti-Rothermel A, Ganz O, Vallone DM, Abrams DB (2013) Real-time exposure to point-of-sale tobacco marketing predicts daily smoking status and smoking cessation outcomes: a multilevel geo-spatial analysis. Am J Prev Med 45(4):379–385
10. Kirchner TR, Anesetti-Rothermel A, Bennett M, Gao H, Carlos H, Scheuermann TS et al Tobacco outlet density and converted versus native nondaily cigarette use in a national U.S. sample. Tob Control in press
11. Cantrell J, Ganz O, Anesetti-Rothermel A, Harrell P, Kreslake JM, Xiao H, Pearson JL, Vallone D, Kirchner TR (2015) Cigarette price variation around high schools: evidence from Washington DC. Health & Place 31:193–198
12. Kirchner TR, Villanti AC, Cantrell J, Anesetti-Rothermel A, Ganz O, Conway KP, et al (2014) Tobacco retail outlet advertising practices and proximity to schools, parks and public housing affect Synar underage sales violations in Washington, DC. Tob Control, Published Online First: 25 February 2014. https://doi.org/10.1136/tobaccocontrol-2013-051239
13. Ganz O, Cantrell J, Moon-Howard J, Aidala A, Kirchner TR, Vallone D (2015) Electronic cigarette advertising at the point-of-sale: a gap in tobacco control research. Tob Control 24(e1):e110–e1e2
14. Cantrell J, Anesetti-Rothermel A, Pearson JL, Xiao H, Vallone D, Kirchner TR (2015) The impact of the tobacco retail outlet environment on adult cessation and differences by neighborhood poverty. Addiction 110(1):152–161
15. Ilakkuvan V, Tacelosky M, Ivey KC, et al (2014) Cameras for public health surveillance: a methods protocol for crowdsourced annotation of point-of-sale photographs. 3(2):e22. https://doi.org/10.2196/resprot.3277

16. Kirchner TR, Villanti A, Anesetti-Rothermel A, Tacelosky M, Pearson J, Cantrell J et al (2015) National enforcement of the Family Smoking Prevention and Tobacco Control Act at point-of-sale. Tob Regul Sci 1(1):24–35. https://doi.org/10.18001/TRS.1.1.3

17. Cantrell J, Pearson JL, Anesetti-Rothermel A, Xiao H, Kirchner TR, Vallone D (2016) Tobacco retail outlet density and young adult tobacco initiation. Nicotine Tob Res 18(2):130–137

18. Cantrell J, Ganz O, Ilakkuvan V, Tacelosky M, Kreslake J, Moon-Howard J, Aidala A, Vallone D, Anesetti-Rothermel A, Kirchner TR (2015) Implementation of a multimodal mobile system for point-of-sale surveillance: lessons learned from case studies in Washington, DC, and New York City. JMIR Public Health and Surveill 1(2):e20

19. Kirchner TR, Anesetti-Rothermel A, Bennett M, Gao H, Carlos H, Scheuermann TS, Reitzel LR, Ahluwalia JS (2016) Tobacco outlet density and converted versus native non-daily cigarette use in a national US sample. Tob Control 26:85–91. https://doi.org/10.1136/tobaccocontrol-2015-052487

20. Slater S, Chaloupka FJ, Wakefield M, Johnston LD, O'Malley PM (2007) The impact of retail cigarette marketing practices on youth smoking uptake. Arch Pediatr Adolesc Med 161(5):440–445. https://doi.org/10.1001/archpedi.161.5.440

21. Feighery EC, Henriksen L, Wang Y, Schleicher NC, Fortmann SP (2006) An evaluation of four measures of adolescents' exposure to cigarette marketing in stores. Nicotine Tob Res 8(6):751–759. doi: G6730P56K6W02647 [pii]. https://doi.org/10.1080/14622200601004125

22. Carter OB, Mills BW, Donovan RJ (2009) The effect of retail cigarette pack displays on unplanned purchases: results from immediate postpurchase interviews. Tob Control 18(3):218–221

23. Institute PoPA (1992) The point-of-purchase advertising industry fact book. The Point of Purchase Advertising Institute, Englewood

24. Henriksen L, Schleicher NC, Feighery EC, Fortmann SP (2010) A longitudinal study of exposure to retail cigarette advertising and smoking initiation. Pediatrics 126(2):232–238. https://doi.org/10.1542/peds.2009-3021

25. Wakefield M, Germain D, Henriksen L (2008) The effect of retail cigarette pack displays on impulse purchase. Addiction 103(2):322–328

26. Cantrell J, Pearson J, Ansetti-Rothermel A, Haijun X, Vallone D, Kirchner TR Tobacco retail outlet density and young adult tobacco initiation. Nicotine Tob Res in press

27. Siahpush M, Jones PR, Singh GK, Timsina LR, Martin J (2010) The association of tobacco marketing with median income and racial/ethnic characteristics of neighbourhoods in Omaha, Nebraska. Tob Control 19(3):256–258. https://doi.org/10.1136/tc.2009.032185

28. Laws MB, Whitman J, Bowser DM, Krech L (2002) Tobacco availability and point of sale marketing in demographically contrasting districts of Massachusetts. Tob Control 11(Suppl 2):ii71–ii73

29. Paynter J, Edwards R (2009) The impact of tobacco promotion at the point of sale: a systematic review. Nicotine Tob Res 11(1):25–35

30. John R, Cheney MK, Azad MR (2009) Point-of-sale marketing of tobacco products: taking advantage of the socially disadvantaged? J Health Care Poor Underserved 20(2):489–506

31. Lovato C, Watts A, Stead LF (2011) Impact of tobacco advertising and promotion on increasing adolescent smoking behaviours. London: John Wiley & Sons, Ltd. Cochrane Database of Systematic Reviews (10) Art. No.: CD003439. https://doi.org/10.1002/14651858.CD003439.pub2

32. Feighery EC, Schleicher NC, Cruz TB, Unger JB (2008) An examination of trends in amount and type of cigarette advertising and sales promotions in California stores, 2002–2005. Tob Control 17(2):93–98

33. Black J, Macinko J (2014) Neighborhoods and obesity. Nutr Rev 66(1):2–20

34. Hartig T, Mitchell R, de Vries S, Frumkin H (2014) Nature and health. Annu Rev Public Health 35:207–228

35. Larson N, Story M, Nelson M (2009) Neighborhood environments: disparities in access to healthy foods in the U.S. Am J Prev Med 36(1):74–81

36. Pickett K, Pearl M (2001) Multilevel analyses of neighborhood socioeconomic context and health outcomes. J Epidemiol Community Health 55:111–122

37. Openshaw S (1984) The modifiable areal unit problem. GeoBooks, Norwich

38. Kwan M-P (2012) The uncertain geographic context problem. Ann Assoc Am Geogr 102(5):958–968

39. Chen X, Kwan MP (2015) Contextual uncertainties, human mobility, and perceived food environment: the uncertain geographic context problem in food access research. Am J Public Health 105(9):1734–1737. https://doi.org/10.2105/AJPH.2015.302792

40. James P, Berrigan D, Hart JE, et al (2014) Effects of buffer size and shape on associations between the built environment and energy balance. Health & place 27:162–170. https://doi.org/10.1016/j.healthplace.2014.02.003

41. Spielman SE, Yoo EH (2009) The spatial dimensions of neighborhood effects. Soc Sci Med 68(6):1098–1105. https://doi.org/10.1016/j.socscimed.2008.12.048

42. Kwan M-P (2013) Toward temporally integrated geographies of segregation, health, and accessibility. Ann Assoc Am Geogr 103(5):1078–1086

43. Rainham D, McDowell I, Krewski D, Sawada M (2010) Conceptualizing the healthscape: contributions of time geography, location technologies and spatial ecology to place and health research. Soc Sci Med 70(5):668–676. https://doi.org/10.1016/j.socscimed.2009.10.035

44. Cummins S (2007) Commentary: investigating neighbourhood effects on health–avoiding the 'local trap'. Int J Epidemiol 36(2):355–357. https://doi.org/10.1093/ije/dym033

45. Caspi CE, Sorensen G, Subramanian SV, Kawachi I (2014) The local food environment and diet: a systematic review. Health & place 18(5):1172–1187. https://doi.org/10.1016/j.healthplace.2012.05.006

46. Kavanagh AM, Kelly MT, Krnjacki L, Thornton L, Jolley D, Subramanian SV, Turrell G, Bentley RJ (2011) Access to alcohol outlets and harmful alcohol consumption: a multi-level study in Melbourne, Australia. Addiction 106(10):1772–1779. https://doi.org/10.1111/j.1360-0443.2011.03510.x

47. Drewnowski A (2004) Obesity and the food environment: dietary energy density and diet costs. Am J Prev Med 27(3 Suppl):154–162. https://doi.org/10.1016/j.amepre.2004.06.011

48. Agresti A (1990) Categorical data analysis. Wiley, New York

49. House B (2013) OpenPaths: empowering personal geographic data. In: Proceeding of the ISEA; Sydney, Australia

50. Carlos HA, Shi X, Sargent J, Tanski S, Berke EM (2010) Density estimation and adaptive bandwidths: a primer for public health practitioners. Int J Health Geogr 9:39. https://doi.org/10.1186/1476-072X-9-39

51. U.S. Census Bureau (2014) Introduction to NAICS [cited 2014 June 9]. Available from: https://www.census.gov/eos/www/naics/index.html

52. D'Angelo H, Fleischhacker S, Rose SW, Ribisl KM (2014) Field validation of secondary data sources for enumerating retail tobacco outlets in a state without tobacco outlet licensing. Health & place 28:38–44. https://doi.org/10.1016/j.healthplace.2014.03.006

53. Toole JL, Colak S, Sturt B, Alexander LP, Evsukoff A, Gonzalez MC (2015) The path most traveled: travel demand estimation using big data resources. Transp Res C 58:162–177

54. Zheng Y, Capra L, Wolfson O, Yang H (2014) Urban computing: concepts, methodologies, and applications. In: proceedings of the ACM TIST. ACM Press, New York

55. Vincenty T (1975) Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. Surv Rev 23:88–93

56. Kirchner TR, Shiffman S (2013) Ecological momentary assessment. In The Wiley-Blackwell Handbook of Addiction Psychopharmacology (eds J. MacKillop and H. de Wit). https://doi.org/10.1002/9781118384404.ch20

57. Lioy PJ, Smith KR (2013) A discussion of exposure science in the 21st century: a vision and a strategy. Environ Health Perspect 121:405–409

58. Chokshi DA, Farley TA (2014) Health. changing behaviors to prevent noncommunicable diseases. Science 345(6202):1243–1244. https://doi.org/10.1126/science.1259809

59. Cromley EK, McLafferty SL (2012) GIS and public health, 2nd edn. Guilford Press, New York

Affiliations

Thomas R. Kirchner[1,2,3] · Hong Gao[1] · Daniel J. Lewis[4] · Andrew Anesetti-Rothermel[5] · Heather A. Carlos[6] · Brian House[7]

✉ Thomas R. Kirchner
   tom.kirchner@nyu.edu

1   College of Global Public Health, New York University, 715 Broadway, 12th Floor, New York, NY 10003, USA

2   Center for Urban Science and Progress, New York University, New York, NY, USA

3   Department of Population Health, New York University Medical Center, New York, NY, USA

4   Department of Social and Environmental Health Research, London School of Hygiene and Tropical
     Medicine, London, UK

5   Schroeder Institute at Truth Initiative, Washington, DC, USA

6   Norris Cotton Cancer Center, Dartmouth College, Hanover, NH, USA

7   Brown University, Providence, RI, USA