



Practice of Epidemiology

Examination of How Neighborhood Definition Influences Measurements of Youths' Access to Tobacco Retailers: A Methodological Note on Spatial Misclassification

Dustin T. Duncan*, Ichiro Kawachi, S.V. Subramanian, Jared Aldstadt, Steven J. Melly, and David R. Williams

* Correspondence to Dr. Dustin T. Duncan, New York University School of Medicine, Department of Population Health, 227 East 30th Street, Room 621, New York, NY 10016 (e-mail: dustin.duncan@nyumc.org).

Initially submitted June 1, 2013; accepted for publication September 24, 2013.

Measurements of neighborhood exposures likely vary depending on the definition of “neighborhood” selected. This study examined the extent to which neighborhood definition influences findings regarding spatial accessibility to tobacco retailers among youth. We defined spatial accessibility to tobacco retailers (i.e., tobacco retail density, closest tobacco retailer, and average distance to the closest 5 tobacco retailers) on the basis of circular and network buffers of 400 m and 800 m, census block groups, and census tracts by using residential addresses from the 2008 Boston Youth Survey Geospatial Dataset ($n = 1,292$). Friedman tests (to compare overall differences in neighborhood definitions) were applied. There were differences in measurements of youths' access to tobacco retailers according to the selected neighborhood definitions, and these were marked for the 2 spatial proximity measures (both $P < 0.01$ for all differences). For example, the median average distance to the closest 5 tobacco retailers was 381.50 m when using specific home addresses, 414.00 m when using census block groups, and 482.50 m when using census tracts, illustrating how neighborhood definition influences the measurement of spatial accessibility to tobacco retailers. These analyses suggest that, whenever possible, egocentric neighborhood definitions should be used. The use of larger administrative neighborhood definitions can bias exposure estimates for proximity measures.

exposure science; modifiable areal unit problem; neighborhood; spatial scale; spatial zone; tobacco retailers; uncertain geographic context problem

In 1923, sociologist Roderick McKenzie wrote, “Probably no other term is used so loosely or with such changing content as the term neighborhood, and very few concepts are more difficult to define” (1, pp. 334–335). It is therefore not surprising that the neighborhood definitions used in spatial epidemiology vary widely (2, 3). As such, there are many approaches to delineating the spatial extent of a neighborhood from which measures of the built environment are derived. Because the neighborhood definition (defined here as a unit around an individuals' home) in a study is applied to all study respondents, the findings can be erroneous if the “wrong” definition is selected. It is also important to note that the measurements of neighborhood exposures likely vary depending on the definition of neighborhood selected. This problem is well known among geographers and is referred

to as the “modifiable areal unit problem” (4–6). The modifiable areal unit problem has both a scale component and a zone component (4–6). A related (but conceptually distinct) methodological problem is the “uncertain geographic context problem,” which, in part, articulates that a problem in neighborhood health research is the spatial uncertainty in the actual areas that exert contextual influences on the individuals being studied (7, 8). The modifiable areal unit problem and the uncertain geographic context problem are indeed 2 important and different problems in spatial epidemiology, and addressing 1 does not necessarily mean addressing the other (7, 8).

Nondifferential misclassification of the neighborhood-level exposure, when the same neighborhood unit is selected for all participants, would occur to the same degree in each

outcome group, and therefore could result in an underestimation of a true association (9). Indeed, simulation research (10, 11) suggests that defining neighborhoods with too large a spatial unit will result in underestimation of the neighborhood effect (i.e., the effect of a particular neighborhood characteristic on a health outcome). The examination of contextual exposures by using administrative areas, including administrative centroids, could be useful if that is the finest spatial information available, but could be problematic. It is plausible that the exact location of an address could be quite far (e.g., 1,000 m) from the centroid of a spatially aggregated unit such as a US census tract or county, highlighting potential spatial misclassification in health and related research. Although a few studies demonstrate that there can be large distances between specific home addresses and proxy administrative definitions of their neighborhoods (such as census tracts) (12, 13), few have quantified whether the neighborhood exposure effect estimate varies by the neighborhood definition, especially by using statistical methods that rely on hypothesis tests (14–22). The vast majority of studies that have examined whether the neighborhood exposure estimate varies by neighborhood definition have used simple descriptive statistics (e.g., means, medians) (15–22). Some studies have examined correlations between the neighborhood exposure estimates among different neighborhood definitions (14, 18, 19, 22). These statistical methods, although useful and important, are crude and do not indicate which of the neighborhood definitions are different from each other. It is also important to note that most studies examine neighborhood definitions across administrative boundaries only or across egocentric neighborhood definitions only (10). Few studies have simultaneously examined differences between administrative boundaries and egocentric neighborhood definitions (10).

The objective of this study, therefore, was to evaluate the influence of commonly used administrative and egocentric neighborhood definitions (i.e., census tract, census block group, and egocentric buffers of various spatial scales and zones) on measurements of spatial accessibility to tobacco retailers by using addresses from the 2008 Boston Youth Survey Geospatial Dataset (a population-based sample of youth in Boston, Massachusetts) and by using various statistical methods (including those that rely on hypothesis tests). We focused on youth because the neighborhood context might be salient to them (i.e., they likely have restricted mobility because they often do not drive). Spatial accessibility to tobacco retailers has been linked to tobacco use among youth (23–28), which is the leading risk factor for preventable death in the United States, including death from lung cancer and other fatal cancers (29).

MATERIALS AND METHODS

Boston Youth Survey Geospatial Dataset

The address data used in this study are from the 2008 Boston Youth Survey Geospatial Dataset, which included 9th- to 12th-grade students in the Boston Public Schools who participated in the 2008 survey and provided the nearest cross-streets to their residential addresses ($n = 1,292$) (30–32).

Schools that served adults, students transitioning back to school after incarceration, suspended students, and students with severe disabilities were ineligible. In 2008, a total of 22 (of 32) eligible public high schools in Boston participated in the Boston Youth Survey. Participating and nonparticipating eligible schools did not have statistically significant differences in key characteristics, including racial/ethnic composition of students and student mobility rates. To generate the sample, we obtained a list of unique classrooms within each participating school, stratified by grade. Classrooms were randomly selected for survey administration. Every student within the selected classrooms was invited to participate. Selection of classrooms continued until approximately 100–125 students had been sampled per school. Of the students enrolled in the classrooms selected for participation ($n = 2,725$), 1,878 completed a survey (68.9% response rate). The majority of nonparticipants (85.5%) were absent from school on the day of survey administration. All geocoded participants resided in Boston. Students' addresses were geocoded to the nearest intersections to protect confidentiality. More detailed information on sampling and geocoding is described elsewhere (30–32).

Spatial accessibility to tobacco retailers

Geographical data on tobacco-selling retail outlets were obtained and geocoded from the Cigarette and Tobacco Excise Unit of the Commonwealth of Massachusetts' Department of Revenue for July 1, 2006, to September 30, 2008. These data pertain to retailers who had tobacco licenses in Massachusetts during this time period and were restricted to the city of Boston for the present study. From 2006 to 2008, there were 787 licensed tobacco retailers in Boston.

We used ArcGIS, version 10.1, software (Environmental Systems Research Institute, Redlands, California) to calculate density of and distance to tobacco retailers for each youth. The spatial density measurements were expressed as the number of stores per km². We also measured distance to the closest tobacco retailer (in m), as well as the average distance to the closest 5 tobacco retailers (in m) based on the youth's address. For the distance calculations, we used network distances. Our approach of including both the distance to the closest retailer and the average distance to the closest 5 retailers is consistent with the methodology of existing research on complexities in spatial distance measurements (14, 33).

Neighborhood definitions: egocentric buffers and census geography

Egocentric neighborhood definitions (also referred to as "egocentric buffers" and "egohoods") define a neighborhood as a radius around a particular location, such as a home (20, 22, 34, 35). In this study, we calculated spatial accessibility to tobacco retailers for each youth on the basis of the following 4 egocentric buffers: a 400-m circular buffer, a 400-m street-network buffer, an 800-m circular buffer, and an 800-m street-network buffer. We used distances of 400 m and 800 m to define the neighborhoods because these relatively short distances constitute a proximal neighborhood

environment for youth (36, 37). We used both circular and street-network buffers because past research studying neighborhood effects on health has used both, and each has distinct properties (22, 34). In line with previous research (16, 22, 31, 32, 34, 38), line-based network buffers were specifically used in this study. The organizing geography for circular buffers is a circle radius around a location or address, whereas street-network buffers use the street network as the organizing geography (i.e., street-network buffers use a width around line segments that follow the street network). Researchers frequently use circular buffers in spatial epidemiology, perhaps because they are functionally easier to create and less computationally demanding, but we note that street-network buffers, especially line-based network buffers, are more relevant to human geography because they take into account the street geography and certain physical barriers, such as rivers. Consequently, it is not surprising that research suggests that street-network buffers are more predictive of physical activity than are circular buffers (34). The circular buffers were created by using the ArcGIS 10 Buffer tool (Environmental Systems Research Institute). The street-network buffers were created from StreetMap streets (which came from ArcGIS 10 Data and Maps), excluding highways and ramps, by using the ArcGIS Network Analyst Extension (Environmental Systems Research Institute). The street-network buffers consisted of 50-m buffers around street center lines that extend along the network from the geocoded residential addresses. To calculate the density of tobacco retailers by using administrative neighborhood definitions, we used the 2010 US Census boundaries for the census block groups and census tracts. Both census block groups (average population = 1,000 residents) and census tracts (average population = 4,000 residents) are geographical units used by the US Census Bureau (Spaldings, Maryland). A census block group is the smallest geographical unit for which the US Census Bureau publishes sample data. Census tracts are small subdivisions of a county. To calculate distance to tobacco retailers by using administrative neighborhood definition variables, we used the census areas' internal points, which were calculated by the US Census Bureau (http://www.census.gov/geo/www/2010census/gtc/gtc_area_attr.html). Usually, the internal point is at or near the geographical center of the unit. However, for some nonconvex geographical units, the calculated geographical center may be located outside the boundaries of the unit. In this circumstance, the internal point is identified as a point inside the entity boundaries nearest to the calculated geographical center. For simplicity, we refer to the internal points as centroids, which they are in many cases.

Analysis

First, we evaluated the positional error when using administrative units (i.e., census tracts and census block groups). Specifically, we calculated the difference in location (Euclidean distance) from the youths' intersection addresses and the centroids of their census block groups and census tracts by using longitude (*X*) and latitude (*Y*) coordinates of geocoded home addresses and census block group and census tract internal points in the Massachusetts State Plane projection (in m).

Second, we computed descriptive statistics for the density and distance measurements for each of the different neighborhood definitions (i.e., 400-m and 800-m circular and street-network buffers, census block groups, and census tracts). To evaluate the relative difference between the spatial tobacco retailer variables across neighborhood definitions, we calculated Spearman's correlation coefficients. Spearman's correlation is more robust than Pearson correlation in the face of nonlinear relationships and variables with a skewed distribution. Correlations between measurements of youths' accessibility to tobacco retailers across neighborhood definitions were calculated. Then, the Friedman test was applied to compare overall differences in neighborhood definitions in measurements of youths' spatial accessibility to tobacco retailers. The Friedman test is a nonparametric randomized block analysis of variance (similar to the parametric repeated measures analysis of variance) (39). The test statistic of the Friedman's test is a χ^2 with [(number of repeated measures) - 1] degrees of freedom (39). Post hoc analysis for the Friedman's test was performed when the null hypothesis was rejected. This allowed us to discover which of the groups (i.e., neighborhood definitions) were responsible for the reason that the null hypothesis was rejected. Analyses were performed for our various neighborhood definitions for all density and distance measurements. Statistical analyses to address the study objectives were conducted by using the R statistical package, version 2.15 (R Foundation for Statistical Computing, Vienna, Austria).

RESULTS

The median difference between the census tract internal points and the home addresses for the sampled youths was 347 (range, 18–2,256) m, and the median difference between the census block group internal points and the home addresses was 205 (range, 2–2,256) m.

Tables 1–3 show descriptive statistics on measurements of youths' spatial accessibility to tobacco retailers for the

Table 1. Density of Tobacco Retailers by Neighborhood Definition (*n* = 1,292), 2008 Boston Youth Survey Geospatial Dataset, Boston, Massachusetts

Neighborhood Definition	No. of Retail Tobacco Stores per km ²		
	Mean (SD)	Median (IQR)	Range
400-m Circular buffer	8.06 (6.23)	8.00 (7.90)	0–49.70
400-m Network buffer	10.54 (8.43)	9.40 (12.20)	0–69.80
800-m Circular buffer	6.98 (4.29)	6.50 (5.00)	0–47.70
800-m Network buffer	9.28 (5.95)	8.70 (5.73)	0–62.20
Census block group ^a	8.67 (11.78)	5.80 (11.50)	0–129.30
Census tract ^a	8.17 (7.79)	6.90 (8.30)	0–61.80

Abbreviations: IQR, interquartile range; SD, standard deviation.

^a For the census block group and census tract distance measurements, we used the census area internal points, which were calculated by the US Census Bureau (Spaldings, Maryland). Usually, the internal point is at or near the geographical center of the unit.

Table 2. Youth's Distance to Closest Tobacco Retailers by Neighborhood Definition ($n = 1,292$), 2008 Boston Youth Survey Geospatial Dataset, Boston, Massachusetts

Neighborhood Definition	Distance to Closest Tobacco Retailer, m		
	Mean (SD)	Median (IQR)	Range
Address	241.53 (240.97)	168.00 (259.00)	0–1,893.00
Census block group ^a	341.04 (237.95)	279.00 (219.00)	1.00–1,717.00
Census tract ^a	384.09 (258.59)	352.90 (287.95)	6.40–1,692.00

Abbreviations: IQR, interquartile range; SD, standard deviation.

^a For the census block group and census tract distance measurements, we used the census area internal points, which were calculated by the US Census Bureau (Spaldings, Maryland). Usually, the internal point is at or near the geographical center of the unit.

different neighborhood definitions. The median tobacco retailer densities (in stores per km²) were 8.00 for the 400-m circular buffer, 9.40 for the 400-m network buffer, 6.50 for the 800-m circular buffer, 8.70 for the 800-m network buffer, 5.80 for the census block group, and 6.90 for the census tract (Table 1). The medians of the distance to the closest tobacco retailers were 168.00 m for the youths' specific home addresses, 279.00 m for census block groups, and 352.90 m for census tracts (Table 2). The medians of the average distances to the closest 5 tobacco retailers were 381.50 m for the youths' specific home addresses, 414.00 m for census block groups, and 482.50 m for census tracts (Table 3).

Spearman correlation coefficients of youths' spatial accessibility to tobacco retailers are presented in Tables 4–6. For the tobacco retailer density measurement, there were significant and moderate-to-strong correlations across neighborhood definitions (from 0.39 to 0.90) (all $P < 0.001$). For the closest tobacco retailer measurement, correlations were lower (from 0.30 to 0.47) (all $P < 0.001$). The correlations for the distance to the closest 5 tobacco retailers ranged from 0.53 to 0.69 (all $P < 0.001$).

Results from the Friedman test are presented in Table 7. Overall, the estimates from all 3 measurements of youths'

Table 3. Average Distance to Closest 5 Tobacco Retailers by Neighborhood Definition, 2008 Boston Youth Survey Geospatial Dataset ($n = 1,292$)

Neighborhood Definition	Distance to Closest 5 Tobacco Retailers, m		
	Mean (SD)	Median (IQR)	Range
Address	448.93 (271.63)	381.50 (290.00)	35.00–2,095.00
Census block group ^a	479.34 (268.73)	414.00 (283.00)	86.00–2,041.00
Census tract ^a	554.27 (278.11)	482.50 (291.70)	195.80–2,097.00

Abbreviations: IQR, interquartile range; SD, standard deviation.

^a For the census block group and census tract distance measurements, we used the census area internal points, which were calculated by the US Census Bureau (Spaldings, Maryland). Usually, the internal point is at or near the geographical center of the unit.

spatial accessibility to tobacco retailers varied for each neighborhood definition (all $P < 0.001$). Although there were neighborhood definition differences in measurements of youths' spatial accessibility to tobacco retailers according to all 3 selected measurements overall, these differences were marked for the 2 spatial proximity measurements (both $P < 0.01$ for all differences). For the tobacco retail density measurement, although the overall comparison was statistically significant ($P < 0.0001$), as were the majority of the specific comparisons, the following differences were not significant: census block group versus 400-m circular buffer, census tract versus 400-m circular buffer, 800-m network buffer versus 400-m network buffer, and census block group versus 800-m circular buffer. Figure 1, which represents an individual youth's residential location, shows the location of the youth's home address, the various buffers used in this study for that address, and the corresponding census block group, and census tract.

DISCUSSION

In spatial epidemiologic investigations, multiple neighborhood definitions are rarely considered in the same study, and little research has considered the influence of neighborhood definition as it relates to spatial misclassification, especially the biased estimation of neighborhood-level exposures. However, bias may be introduced by incorrectly defining neighborhoods (i.e., selecting the "wrong" neighborhood definition). In this investigation, we examined the influence of neighborhood definition on data measuring exposure to tobacco retailers by using addresses from a population-based sample of Boston youth. Our results demonstrate that neighborhood definitions may influence neighborhood-level exposure estimates, especially when using distance-based measurements. Correlations between neighborhood definitions were lowest for the closest tobacco retailers measurement. This suggests that the closest tobacco retailers measurement should not be used when using proxy neighborhood definitions. As Figure 1 shows, the results will also be influenced by the actual locations of the participants in the sample. If more individuals in the sample are located closer to census-unit boundaries, larger differences between the individual-based measurements and measurements based on census units may be observed. Because our study area is a dense urban setting, it makes sense that there is not much difference between using the 400-m and 800-m network buffers. Also, it makes sense that home addresses are closer to tobacco retailers than are census centroids, assuming that home addresses are randomly distributed in residential areas, and census centroids could be in residential or nonresidential areas, whereas tobacco retailers are more likely to be located near residential areas.

Ours is one of the few studies to examine the effect of neighborhood definitions on exposure estimation and spatial misclassification, especially in such a comprehensive way. Although few studies are directly comparable to ours, there is some existing research on the topic. One of the few studies that compare estimates across geographical units descriptively showed that estimates of built-environment features (e.g., population density, land-use mix, and park density)

Table 4. Spearman Correlation Coefficients^a of Density of Tobacco Retailers, 2008 Boston Youth Survey Geospatial Dataset, Boston, Massachusetts

Neighborhood Definition	Neighborhood Definition					
	400-m Circular Buffer	400-m Network Buffer	800-m Circular Buffer	800-m Network Buffer	Census Block Group	Census Tract
400-m Circular buffer	1.00	0.90	0.66	0.71	0.48	0.53
400-m Network buffer		1.00	0.56	0.66	0.48	0.52
800-m Circular buffer			1.00	0.86	0.39	0.57
800-m Network buffer				1.00	0.41	0.58
Census block group					1.00	0.58
Census tract						1.00

^a All $P < 0.001$.

varied by the geographical unit (e.g., census tract or borough) (17). Another descriptive study found that the mean number of food stores (e.g., local fast food, convenience stores) varied by buffer specification (i.e., 400 m vs. 800 m (15)), whereas another descriptive study found differences in various built-environment variables (e.g., parks and community design features) across 3 network buffers (i.e., 400 m, 800 m, and 1,600 m) (16). Forsyth et al. (20) descriptively examined and found differences for food- and physical activity-built environments by using buffers of diverse spatial scales and zones. Thornton et al. (21) descriptively evaluated the effect of various estimates of supermarket access, including by using Euclidean and network buffers. Sparks et al. (18) descriptively evaluated the effect of various census-based neighborhood definitions on supermarket access, as well as evaluating correlations, which were moderate to strong. Boruff et al. (22) found that neighborhood definition (i.e., buffer size and zone) influenced the spatial exposure estimates of various built-environment features (e.g., percent commercial, percent recreational and parks). This study also examined correlations across buffers and found moderate to strong correlations (22). Similarly, in another study, correlations for food-environment variables across egocentric buffers were generally strong (19). Aparicio et al. (14) found

high correlations between different measurements of accessibility (e.g., minimum network distance and average distance to the 5 closest health services) across spatial units (i.e., census tracts, dissemination areas, blocks) in the Montreal, Canada, metropolitan census area, but they also found evidence of aggregation errors.

Study implications

Results from this study demonstrate that neighborhood definition influences measurements of youths' spatial accessibility to tobacco retailers, and that the use of census boundaries as neighborhood proxies can be problematic when estimating an individual's exposure. We suggest that neighborhood definitions should be driven by theory, not data. We recognize that different neighborhood definitions may be more or less important depending on the research question. Often, researchers select a neighborhood definition without explicitly addressing why the definition was chosen. Although many studies have used administrative neighborhood definitions (e.g., US census tracts) when evaluating neighborhood features, the use of an individual's specific address rather than a proxy (e.g., administrative neighborhood boundary) can be important, and we note that census boundaries

Table 5. Spearman Correlation Coefficients^a of Youths' Spatial Access to the Closest Tobacco Retailer^b, 2008 Boston Youth Survey Geospatial Dataset, Boston, Massachusetts

Neighborhood Definition	Neighborhood Definition		
	Address	Census Block Group	Census Tract
Address	1.00	0.46	0.30
Census block group		1.00	0.47
Census tract			1.00

^a All $P < 0.001$.

^b For the census block group and census tract distance measurements, we used the census area internal points, which were calculated by the US Census Bureau (Spaldings, Maryland). Usually, the internal point is at or near the geographical center of the unit.

Table 6. Spearman Correlation Coefficients^a of the Average Distance to the Closest 5 Tobacco Retailers to Youths' Homes^b, 2008 Boston Youth Survey Geospatial Dataset, Boston, Massachusetts

Neighborhood Definition	Neighborhood Definition		
	Address	Census Block Group	Census Tract
Address	1.00	0.69	0.53
Census block group		1.00	0.62
Census tract			1.00

^a All $P < 0.001$.

^b For the census block group and census tract distance measurements, we used the census area internal points, which were calculated by the US Census Bureau (Spaldings, Maryland). Usually, the internal point is at or near the geographical center of the unit.

Table 7. Comparison of Overall Differences in Youths' Spatial Access to Tobacco Retailers by Neighborhood Definition^a, 2008 Boston Youth Survey Geospatial Dataset, Boston, Massachusetts

Neighborhood Definition by Access Variable	P Value
Tobacco retail density	
Overall	<0.0001
400-m NB versus 400-m CB	<0.0001
800-m CB versus 400-m CB	0.0010
800-m NB versus 400-m CB	<0.0001
Census block group versus 400-m CB	0.1106
Census tract versus 400-m CB	0.9957
800-m CB versus 400-m NB	<0.0001
800-m NB versus 400-m NB	0.8321
Census block group versus 400-m NB	<0.0001
Census tract versus 400-m NB	<0.0001
800-m NB versus 800-m CB	<0.0001
Census block group versus 800-m CB	0.7062
Census tract versus 800-m CB	0.0001
Census block group versus 800-m NB	<0.0001
Census tract versus 800-m NB	<0.0001
Census tract versus census block group	0.0269
Closest tobacco retailers	
Overall	<0.0001
Census block group versus address	<0.0001
Census tract versus address	<0.0001
Census tract versus census block group	0.0070
Mean distance to the closest 5 tobacco retailers	
Overall	<0.0001
Census block group versus address	0.0038
Census tract versus address	<0.0001
Census tract versus census block group	<0.0001

Abbreviations: CB, circular buffer; NB, network buffer.

^a For the census block group and census tract distance measurements, we used the census area internal points, which were calculated by the US Census Bureau (Spauldings, Maryland). Usually, the internal point is at or near the geographical center of the unit.

might not be the most relevant neighborhood definition for understanding one's spatial behavior and exposure. An individual's specific address may be more relevant to young people's social realities and health/wellbeing (2, 3). Additionally, localized buffer-based neighborhood definitions may be preferred to administrative neighborhood definitions, because defining neighborhoods with administrative units may be especially inadequate for individuals living on the margins of those areas (38). We also note that, as opposed to street-network buffers, circular buffers include areas inaccessible when walking, which suggests that the circular buffer (which is frequently used) may be less appropriate for measuring neighborhood features that are accessed by walking (34). Future epidemiologic research should justify the chosen neighborhood definition and perhaps undertake sensitivity analyses. Because distance variables, as compared

to density variables, were more susceptible to misclassification, if one must use an administrative area, perhaps this should be done only when using density variables. Although results from the present study suggest the use of egocentric buffers to define neighborhoods, we also understand that privacy concerns are a potential reason for using larger spatial administrative units. In addition, researchers should also consider other key factors, such as availability of geospatial data and potential spatial error in the geospatial data set.

Study limitations

Limitations of the study should be noted. First, we note that geographic information systems data can have positional errors. Second, we recognize that this research could not account for the sale of tobacco products through outlets other than physical stores, such as online retailers (40–42). We also note that the modifiable areal unit problem is a concern in this study and in all other spatially oriented research (4–6). In this analysis, we included egocentric buffers, census block groups, and census tracts, which are neighborhood definitions used in research in Boston and other locales. We note, though, that neighborhood definitions based on census geography, especially census tracts, are historically the most common neighborhood proxy in research of neighborhoods and health, perhaps because of readily available data at that level (2, 43). Although we selected the most common neighborhood definitions used in public health and spatial epidemiology research, we recognize that other neighborhood definitions exist and have been applied to research in Boston. For example, researchers have used neighborhood definitions based on information from the Boston Public Health Commission (44) and the Boston Redevelopment Authority (45, 46). However, for this study, we did not seek to evaluate these areas, because they are much larger than census tracts (increasing the likelihood of spatial misclassification), and they could not be applied to other geographical regions. Also, the use of larger neighborhoods such as these likely makes sense when one wants to make policy recommendations, because the city uses these neighborhood definitions to allocate resources. The use of smaller neighborhood locations as we have done likely makes sense when research questions are designed to evaluate local factors that influence individuals' behavior (3). Additionally, there are different ways of measuring egocentric neighborhoods. For reasons previously articulated, in this study, we selected line-based egocentric network buffers as opposed to polygon-based egocentric network buffers. However, we note that the impact of different network buffers (e.g., line-based vs. polygon-based network buffers) on geospatial exposures will likely be minimal in urban environments, including Boston (34). A static buffer around a person's home is only 1 way to measure an egocentric neighborhood. Recent research has used the actual space-time paths of participants by using global positioning system technology (22, 47). The decision to focus on any buffer surrounding a location is based on the assumption that the youth under study spend sufficient time in that geographical area to influence their behavior and health, which may or may not be the case (7, 8). Furthermore, also related to the uncertain geographical context problem (7, 8), this study addresses only 1

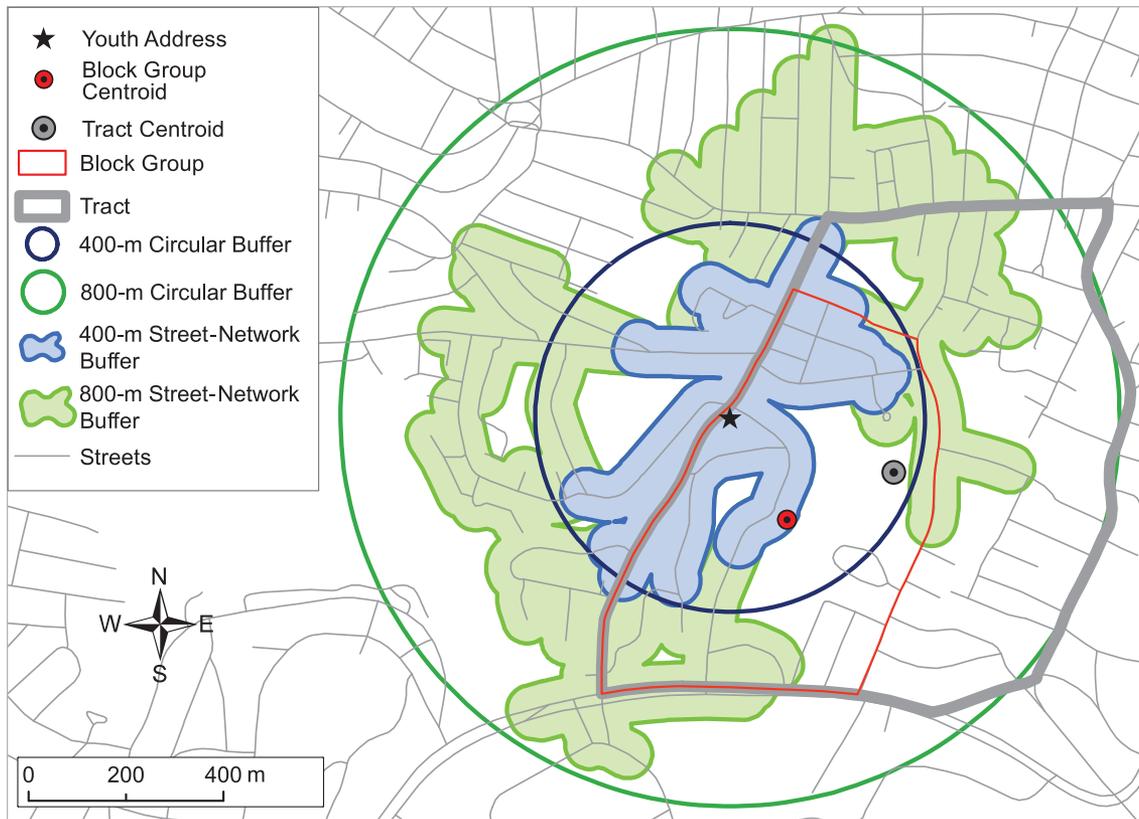


Figure 1. Neighborhood definitions for an individual youth in the 2008 Boston Youth Survey Geospatial Dataset, Boston, Massachusetts.

aspect of the problem; relationships between contextual variables and the individual-based outcome variable(s) may be different for different areal delineations of neighborhood but were not examined in this study. Moreover, we did not account for other geographically based places where youth could access tobacco, such as in their school neighborhoods. Finally, although youth in our sample came from neighborhoods across Boston (which increases the generalizability of our findings), the study is limited to an urban, geographically constrained area; therefore, the external validity of the results might be limited. Future research is needed to address the limitations of this study by examining how other neighborhood definitions (including other egocentric neighborhood definitions, such as global positioning system paths) influence spatial misclassification and by examining nonurban contexts and various neighborhood-based amenities.

Conclusion

Neighborhood definitions influence measurements of spatial accessibility to tobacco retailers. These analyses suggest that, when estimating an individual's exposure, researchers should use egocentric neighborhood definitions whenever possible. The use of larger administrative neighborhood definitions can bias exposure estimates for proximity measures. These findings have significant implications for future

epidemiologic research. Researchers need to think carefully about which contextual units to use and to undertake sensitivity analyses.

ACKNOWLEDGMENTS

Author affiliations: Department of Social and Behavioral Sciences, Harvard School of Public Health, Boston, Massachusetts (Dustin T. Duncan, Ichiro Kawachi, S.V. Subramanian, David R. Williams); Lung Cancer Disparities Center, Harvard School of Public Health, Boston, Massachusetts (Dustin T. Duncan, Ichiro Kawachi, David R. Williams); Department of Geography, University at Buffalo, State University of New York, Buffalo, New York (Jared Altstadt); Department of Environmental Health, Harvard School of Public Health, Boston, Massachusetts (Steven J. Melly); Department of African and African American Studies, Harvard University, Cambridge, Massachusetts (David R. Williams); and Department of Sociology, Harvard University, Cambridge, Massachusetts (David R. Williams).

At the time of the study, D.T.D. was supported by the Alonzo Smythe Yerby Postdoctoral Fellowship at Harvard School of Public Health. D.R.W. was supported in part by a grant from the National Cancer Institute (grant 1P50CA 148596) to the Lung Cancer Disparities Center at Harvard

School of Public Health. The 2008 Boston Youth Survey was funded by a grant from the Centers for Disease Control and Prevention (grant U49CE00740) to the Harvard Youth Violence Prevention Center at Harvard School of Public Health. The Robert Wood Johnson Foundation's Active Living Research Program (grant 67129 to D.T.D.) supported the development of the Boston Youth Survey geospatial data set. This study was supported by the Robert Wood Johnson Foundation Health and Society Scholars Seed Grant Program (grant to D.T.D.), Harvard Center for Population and Development Studies, Harvard School of Public Health.

The Boston Youth Survey was conducted in collaboration with the Boston Public Health Commission, Boston's Office of Human Services, Boston Public Schools, and the Office of the Mayor, the Honorable Thomas M. Menino. The survey would not have been possible without the participation of the faculty, staff, administrators, and students of Boston Public Schools, as well as faculty, staff, and students of Harvard School of Public Health and City of Boston employees who participated in survey administration. We thank Jeff Blossom for providing technical assistance with building this geospatial data set and for creating the figure used in this study. We thank Susan Kum for assisting in developing and reviewing the R code for the statistical analysis. We gratefully acknowledge the efforts of Rochelle Frounfelker for her assistance with the preparation of this manuscript.

Conflict of interest: none declared.

REFERENCES

- McKenzie R. *The Neighborhood: A Study of Local Life in the City of Columbus, Ohio*. Chicago, IL: The University of Chicago Press; 1923.
- Osypuk TL, Galea S. What level macro? Choosing appropriate levels to assess the relation between space and population health. In: Galea S, ed. *Macrosocial Determinants of Population Health*. New York, NY: Springer Media; 2007:399–436.
- Matthews SA. Spatial polygamy and the heterogeneity of place: studying people and place via egocentric methods. In: Burton L, Kemp SP, Leung M, et al, eds. *Communities, Neighborhoods and Health: Expanding the Boundaries of Place*. New York, NY: Springer; 2011:35–55.
- Openshaw S, Taylor P. A million or so correlation coefficients: three experiments on the modifiable area unit problem. In: Wrigley N, ed. *Statistical Applications in the Spatial Sciences*. London, United Kingdom: Pion Ltd; 1979:127–144.
- Arbia G. *Spatial Data Configuration in the Statistical Analysis of Regional Economics and Related Problems*. Dordrecht, the Netherlands: Kluwer; 1989.
- Wong D. The modifiable areal unit problem (MAUP). In: Fotheringham A, Rogerson PA, eds. *The SAGE Handbook of Spatial Analysis*. London, United Kingdom: SAGE Publications; 2009:104–124.
- Kwan M-P. The uncertain geographic context problem. *Ann Assoc Am Geogr*. 2012;102(5):958–968.
- Kwan M-P. How GIS can help address the uncertain geographic context problem in social science research. *Annals of GIS*. 2012;18(4):245–255.
- Rothman K, Greenland S. *Modern Epidemiology*. Philadelphia, PA: Lippincott, Williams and Wilkins; 1998.
- Spielman S, Yoo EH. The spatial dimensions of neighborhood effects. *Soc Sci Med*. 2009;68(6):1098–1105.
- Spielman SE, Yoo E-H, Linkletter C. Neighborhood contexts, health, and behavior: understanding the role of scale and residential sorting. *Environ Plann B Plann Des*. 2013;40(3):489–506.
- Bow C, Waters NM, Faris PD, et al. Accuracy of city postal code coordinates as a proxy for location of residence. *Int J Health Geogr*. 2004;3(1):5.
- Healy M, Gilliland JA. Quantifying the magnitude of environmental exposure misclassification when using imprecise address proxies in public health research. *Spat Spatiotemporal Epidemiol*. 2012;3(1):55–67.
- Apparicio P, Abdelmajid M, Riva M, et al. Comparing alternative approaches to measuring the geographical accessibility of urban health services: distance types and aggregation-error issues. *Int J Health Geogr*. 2008;18(7):7.
- Neckerman K, Bader MD, Richards CA, et al. Disparities in the food environments of New York City public schools. *Am J Prev Med*. 2010;39(3):195–202.
- Duncan DT, Aldstadt J, Whalen J, et al. Validation of walk score for estimating neighborhood walkability: an analysis of four US metropolitan areas. *Int J Environ Res Public Health*. 2011;8(11):4160–4179.
- Schwartz B, Stewart WF, Godby S, et al. Body mass index and the built and social environments in children and adolescents using electronic health records. *Am J Prev Med*. 2011;41(4):e17–e28.
- Sparks A, Bania N, Leete L. Comparative approaches to measuring food access in urban areas: the case of Portland, Oregon. *Urban Stud*. 2011;48:1715–1737.
- Boone-Heinonen J, Gordon-Larsen P, Kiefe CI, et al. Fast food restaurants and food stores: longitudinal associations with diet in young to middle-aged adults: the CARDIA Study. *Arch Intern Med*. 2011;171(13):1162–1170.
- Forsyth A, Van Riper D, Larson N, et al. Creating a replicable, valid cross-platform buffering technique: the sausage network buffer for measuring food and physical activity built environments. *Int J Health Geogr*. 2012;11:14.
- Thornton L, Pearce JR, Macdonald L, et al. Does the choice of neighbourhood supermarket access measure influence associations with individual-level fruit and vegetable consumption? A case study from Glasgow. *Int J Health Geogr*. 2012;11:29.
- Boruff BJ, Nathan A, Nijenstein S. Using GPS technology to (re)-examine operational definitions of 'neighbourhood' in place-based health research. *Int J Health Geogr*. 2012;11:22.
- Pokorny S, Jason LA, Schoeny ME. The relation of retail tobacco availability to initiation and continued smoking. *J Clin Child Adolesc Psychol*. 2003;32(2):193–204.
- Novak S, Reardon SF, Raudenbush SW, et al. Retail tobacco outlet density and youth cigarette smoking: a propensity-modeling approach. *Am J Public Health*. 2006;96(4):670–676.
- Henriksen L, Feighery EC, Schleicher NC, et al. Is adolescent smoking related to the density and proximity of tobacco outlets and retail cigarette advertising near schools? *Prev Med*. 2008;47(2):210–214.
- McCarthy W, Mistry R, Lu Y, et al. Density of tobacco retailers near schools: effects on tobacco use among students. *Am J Public Health*. 2009;99(11):2006–2013.
- Chan W, Leatherdale ST. Tobacco retailer density surrounding schools and youth smoking behaviour: a multi-level analysis. *Tob Induc Dis*. 2011;9(1):9.

28. Johns M, Sacks R, Rane M, et al. Exposure to tobacco retail outlets and smoking initiation among New York City adolescents [published online ahead of print May 23, 2013]. *J Urban Health*. (doi:10.1007/s11524-013-9810-2).
29. Danaei G, Ding EL, Mozaffarian D, et al. The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors. *PLoS Med*. 2009;6(4):e1000058.
30. Azrael D, Johnson RM, Molnar BE, et al. Creating a youth violence data system for Boston, Massachusetts. *Aust N Z J Criminol*. 2009;42(3):406–421.
31. Duncan DT, Castro MC, Gortmaker SL, et al. Racial differences in the built environment—body mass index relationship? A geospatial analysis of adolescents in urban neighborhoods. *Int J Health Geogr*. 2012;11(1):11.
32. Duncan DT, Aldstadt J, Whalen J, et al. Validation of walk scores and transit scores for estimating neighborhood walkability and transit availability: a small-area analysis. *GeoJournal*. 2013;78(1):407–416.
33. Block JP, Christakis NA, O'Malley AJ, et al. Proximity to food establishments and body mass index in the Framingham Heart Study offspring cohort over 30 years. *Am J Epidemiol*. 2011;174(10):1108–1114.
34. Oliver LN, Schuurman N, Hall AW. Comparing circular and network buffers to examine the influence of land use on walking for leisure and errands. *Int J Health Geogr*. 2007;6:41.
35. Hipp JR, Boessen A. *Ego*hoods as waves washing across the city: a new measure of “neighborhoods”. *Criminology*. 2013; 51(2):287–327.
36. Timperio A, Crawford D, Telford A, et al. Perceptions about the local neighborhood and walking and cycling among children. *Prev Med*. 2004;38(1):39–47.
37. Colabianchi N, Dowda M, Pfeiffer KA, et al. Towards an understanding of salient neighborhood boundaries: adolescent reports of an easy walking distance and convenient driving distance. *Int J Behav Nutr Phys Act*. 2007;4:66.
38. Duncan DT, Piras G, Dunn EC, et al. The built environment and depressive symptoms among urban youth: a spatial regression study. *Spat Spatiotemporal Epidemiol*. 2013;5:11–25.
39. Friedman M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *J Am Stat Assoc*. 1937;32(200):675–701.
40. Ribisl KM, Kim AE, Williams RS. Are the sales practices of Internet cigarette vendors good enough to prevent sales to minors? *Am J Public Health*. 2002;92(6):940–941.
41. Ribisl KM, Williams RS, Kim AE. Internet sales of cigarettes to minors. *JAMA*. 2003;290(10):1356–1359.
42. Ribisl KM, Jo C. Tobacco control is losing ground in the Web 2.0 era: invited commentary. *Tob Control*. 2012;21(2): 145–146.
43. Messer L, Kaufman JS. Using census data to approximate neighborhood effects. In: Oakes JM, Kaufman JS, eds. *Methods in Social Epidemiology*. San Francisco, CA: Jossey-Bass; 2006:209–236.
44. Chen J, Rehkopf DH, Waterman PD, et al. Mapping and measuring social disparities in premature mortality: the impact of census tract poverty within and across Boston neighborhoods, 1999–2000. *J Urban Health*. 2006;83 (6):1063–1084.
45. Li W, Land T, Zhang Z, et al. Small-area estimation and prioritizing communities for tobacco control efforts in Massachusetts. *Am J Public Health*. 2009;99(3):470–479.
46. Li W, Kelsey JL, Zhang Z, et al. Small-area estimation and prioritizing communities for obesity control in Massachusetts. *Am J Public Health*. 2009;99(3):511–519.
47. Wiehe SE, Carroll AE, Liu GC, et al. Using GPS-enabled cell phones to track the travel patterns of adolescents. *Int J Health Geogr*. 2008;7:22.