HHS Public Access
Author manuscript
Am J Drug Alcohol Abuse. Author manuscript; available in PMC 2017 July 01.

Published in final edited form as:
Am J Drug Alcohol Abuse. 2016 July ; 42(4): 412–421. doi:10.3109/00952990.2016.1151522.

Local Spatial Clustering in Youths Use of Tobacco, Alcohol and Marijuana in Boston, Massachusetts, USA

Dustin T. Duncan, ScD[1,2,3,4,5], **Michael Rienti, Jr., MA**[6,7], **Martin Kulldorff, PhD**[8], **Jared Aldstadt, PhD**[6], **Marcia C. Castro, PhD**[9,10], **Rochelle Frounfelker, MPH, MSSW**[11], **James H. Williams, AB**[1], **Glorian Sorensen, PhD, MPH**[12,13,14], **Renee M. Johnson, PhD, MPH**[14], **David Hemenway, PhD**[15], and **David R. Williams, PhD, MPH**[11,13,16]

[1]Department of Population Health, New York University School of Medicine, New York, NY USA

[2]College of Global Public Health, New York University, New York, NY USA

[3]Center for Drug Use and HIV Research, New York University College of Nursing, New York, NY USA

[4]Population Center, New York University College of Arts and Science, New York, NY USA

[5]Center for Data Science, New York University, New York, NY USA

[6]Department of Geography, University at Buffalo, State University of New York, Buffalo, NY USA

[7]Center for Health and Social Research, SUNY Buffalo State, Buffalo, NY USA

[8]Department of Population Medicine, Harvard Medical School and Harvard Pilgrim Health Care Institute, Boston, MA USA

[9]Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, MA USA

[10]Harvard Center for Population and Development Studies, Harvard University, Cambridge, MA USA

[11]Department of Social and Behavioral Sciences, Harvard T.H. Chan School of Public Health, Boston, MA USA

[12]Center for Community-based Research, Dana-Farber Cancer Institute, Boston, MA USA

[13]Lung Cancer Disparities Center, Harvard T.H. Chan School of Public Health, Boston, MA USA

[14]Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD USA

[15]Department of Health Policy and Management, Harvard T.H. Chan School of Public Health, Boston, MA USA

[16]Departments of African and African American Studies, and Sociology, Harvard University, Cambridge, MA USA

Correspondence to: Dustin T. Duncan, ScD, New York University School of Medicine, Department of Population Health, Spatial Epidemiology Lab, 227 East 30th Street, Office 621, New York, NY 10016. Tel: 646-501-2674. Fax: 646-501-2706. Dustin.Duncan@nyumc.org.

Abstract

Background—Understanding geographic variation in youth drug use is important for both identifying etiologic factors and planning prevention interventions. However, little research has examined spatial clustering of drug use among youth using rigorous statistical methods.

Objectives—The purpose of this study is to examine spatial clustering of youth use of tobacco, alcohol, and marijuana.

Methods—Responses on tobacco, alcohol, and marijuana use from 1,292 high school students ages 13-19 who provided complete residential addresses were drawn from the 2008 Boston Youth Survey Geospatial Dataset. Response options on past month use included "none", "1-2", "3-9", and "10 or more". The response rate for each substance was approximately 94%. Spatial clustering of youth drug use was assessed using the spatial Bernoulli model in the SatScan™ software package.

Results—Approximately 12%, 36%, and 18% of youth reported any past-month use of tobacco, alcohol, and/or marijuana, respectively. Two clusters of elevated past tobacco use among Boston youths were generated, one of which was statistically significant. This cluster, located in the South Boston neighborhood, had a relative risk of 5.37 with a p-value of 0.00014. There was no significant localized spatial clustering in youth past alcohol or marijuana use in either the unadjusted or adjusted models.

Conclusion—Significant spatial clustering in youth tobacco use was found, and this type of research can be used for local targeting of drug abuse prevention interventions. Finding a significant cluster in the South Boston neighborhood provides reason for further investigation into neighborhood characteristics that may shape adolescents' substance use behaviors. Future research should evaluate the underlying reasons behind spatial clustering of youth substance use.

Introduction

During adolescence, drug use is associated with increased risk for substance use disorders, as well as other health and social problems (e.g., poor school performance) including sexually transmitted infections (e.g., HIV) (1-5). Research also suggests that behaviors developed in adolescence, such as drug use, are often continued into adulthood (6-9).

Historically, understanding and promoting healthy behaviors in youth and other populations has focused on individual-level variables, such as age and gender. It is increasingly recognized that contextual neighborhood factors play a role in healthy behaviors, including drug use. Since Tobler's First Law of Geography ("everything is related to everything else, but near things are more related than distant things") was first articulated over 40 years ago (10), spatial statistics has developed strong ties to public health research and practice, and it continues to be an area of very active investigation (11). One methodological application of spatial statistics in public health is the identification of disease clusters. Examining spatial patterns in health-related outcomes (such as cancer) can help researchers and policymakers understand and intervene in certain neighborhoods or other types of locations (12). For example, if there is a cluster of cancer, researchers can examine environmental factors that might be linked to increased cancer rates in a certain geographic area. A relatively large amount of research has examined spatial patterns in diseases such as cancer. In contrast, only recently have researchers begun to examine spatial patterns of other health-related factors such as obesity and tobacco use. Understanding geographic variation of youth drug use is important for prevention interventions and can also elucidate potential etiologic factors, especially at the neighborhood level. For example, if alcohol and marijuana use cluster spatially, then prevention specialists (e.g., intervention scientists) can target those neighborhoods.

Relatively few studies have examined spatial patterning of drug use as opposed to related factors (e.g., crime, built environment features), and even fewer have focused on spatial clustering of drug use by means of statistical methods. The existing research has notable but addressable limitations, which this study begins to address. First, the vast majority of research conducted on spatial patterns in drug use is largely limited to adult samples. Although studies on youth populations and spatial clustering of drug use behavior remain rare, some studies have been conducted (13-16). For example, McVie and Norris (2006) mapped youth use of illicit drugs and cannabis throughout Edinburgh, Scotland, and Généreux and colleagues (2012) assessed clusters of smoking behaviors in residents over age 15 in Montréal, Canada. Similarly, Chaney and Rojas-Guyler (2015) examined and found spatial patterns of adolescent use of alcohol, tobacco, and marijuana across the five-county Cincinnati, Ohio region. Second, most of this work on spatial patterns in drug use has not used spatial statistics. Instead, most work conducted on spatial patterns in drug use among adults employ simple visualization methods that are unable to clearly identify statistically significant spatial clusters because they rely on “eyeballing” rather than a rigorous statistical test to identify spatial clusters (17-22). In addition, visualization methods do not allow the researcher to adjust for potential confounding variables (such as respondents’ age), which may lead to inaccurate conclusions. Beyond visualization methods, an even simpler method is to examine the prevalence of a health outcome by a geographic region and compare differences (e.g., urban vs. rural difference). Sreeramareddy and colleagues (2011), for example, found regional patterns of tobacco consumption, with rural areas having a higher prevalence of smoking. Third, most work on spatial patterns in drug use has been conducted using administrative neighborhood definitions (such as census tracts or ZIP codes), including the few studies that have used spatial statistics (15, 24). As an example, ZIP code was the geographic unit of analysis in the study conducted by Chaney and Rojas-Guyler (2015) that examined spatial patterns of adolescent drug use in Cincinnati, Ohio. Using administrative neighborhoods minimizes variation between neighborhoods, and empirical research has shown that such aggregation can indeed reduce the power to detect spatial clusters (25, 26). In addition, using administrative neighborhood definitions can result in spatial misclassification and make it difficult to plan more localized interventions (27).

This paper addresses each of the above limitations and illustrates a methodology for assessing local elevated drug use among youth populations. The study sample was drawn from high school students in Boston, Massachusetts (USA), 82% of whom were 14-17 years old at the time they were surveyed. We used the spatial scan statistic and the SaTScan™ software to detect statistically significant clusters of youth drug use while accounting for several covariate factors. Instead of aggregating case data into bounded administrative units, we use individual-level georeferenced observations when conducting analyses, thereby improving the spatial accuracy and precision of identified clusters. This type of methodology could potentially lead to more focused and effective prevention measures dedicated to youth drug abuse. The objective of this study was to examine geographic variation and spatial clustering of youth use of tobacco, alcohol, and marijuana. In line with spatial theory and existing research, we hypothesized that there will be spatial patterns in youth drug use, i.e., spatial clusters will exist.

Methods

Data Collection and Spatial Sample

Data come from the 2008 Boston Youth Survey (BYS) Geospatial Dataset, which includes 9^{th}-12^{th} grade students in the Boston Public Schools system (Boston, Massachusetts) who took the BYS and provided their complete residential addresses (28-30). Similar to the percentage of those schools included in the BYS survey (31), approximately 74% of Boston Public School students in the 2007-2008 academic year were eligible for free or reduced-price meals; the same percentage of students were Black or Hispanic. Schools that served adults, students transitioning back to school after incarceration, suspended students, and students with severe disabilities were ineligible. A total of 22 public high schools in Boston participated in the 2008 BYS (32 schools were eligible). The primary reason for school non-participation was scheduling difficulties. Participating and non-participating eligible schools did not have statistically significant differences in key school characteristics (e.g., racial/ ethnic composition of students, drop-out rates, standardized test scores, student mobility rate). A list of unique classrooms within each participating school was obtained to generate the classroom-level sample, 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. The survey was administered to students by trained staff in the spring of 2008 during 50-minute class periods. Passive consent was sought from parents and students were read an assent statement prior to survey administration. Of the 2,725 students enrolled in the classrooms selected for participation, 1,878 (response rate = 68.9%) completed the survey. Approximately 86% of non-participants were absent from school on the day of survey administration. We obtained and geocoded complete address information to the nearest intersection from 68.8% of the students who took the survey (n = 1,292). There were no differences in past-month drug use behaviors (i.e., tobacco use, alcohol use and marijuana use) between students who provided geocodeable information and those who did not. The Human Subjects Committee (i.e., the institutional review board) at the Harvard School of Public Health reviewed and approved the original study. Further approval for the geospatial analysis underlying this paper was not necessary because the BYS data did not allow for the identification of individual respondents and because these were secondary analyses.

Tobacco, Alcohol and Marijuana Use

The BYS evaluated frequency of drug use. Past-month tobacco use was assessed with the question, "In the past 30 days, on how many days did you use tobacco, including smoking cigarettes or cigars, or chewing tobacco?" Alcohol was assessed with the question, "In the past 30 days, on how many days did you drink alcohol?" Marijuana was assessed with the question, "In the past 30 days, on how many days did you use marijuana?" Response options for all drug use items included a) "none", b) "1-2", c) "3-9", and d) "10 or more". Item were adapted from the 2005 national Youth Risk Behavioral Surveillance System (YRBS) survey. The drug use items have demonstrated good test-retest reliability, especially past-month tobacco use (kappa=76.2) (32).

Statistical Analyses

Prior to conducting spatial analyses, descriptive statistics were assessed for each of the individual variables. To evaluate spatial clustering in youth drug use, we used spatial scan statistics. We examined local spatial patterns as opposed to global patterns, as local spatial clustering methods facilitate the identification of specific "clusters", i.e., local spatial clustering methods facilitate knowing precisely *where* spatial clustering occurs. In other words, local clustering methods can directly identify regions that are significantly different from their "neighbors". Spatial scan statistics (a popular method of local spatial clustering) does so by comparing each scan window to its k geographical neighbors (as opposed to the variance of all scan windows). In contrast, global spatial clustering methods merely detect whether or not spatial clustering is present in a given area. Because our data are at the individual level, and to maximize spatial resolution, we did not aggregate the data into administrative levels (e.g., census tracts or counties in the U.S.). Empirical research has shown that aggregation of spatial units can reduce the ability to detect spatial clusters (25, 26), which is not surprising, as there can be much within-unit variation in larger spatial units.

To adjust for confounding during testing, several other variables from the survey were included as covariates: age (years), gender (male, female), and race/ethnicity (White, Black, Hispanic, Asian, Other). In doing so, we avoided having to remove variables from the test or perform operations, such as imputation, that increase the likelihood of misestimation. Tests were performed both with and without covariate adjustment (33). In these analyses, we used a three-sample strategy in which the sample size was maximized for each outcome type in order to increase the statistical power of the tests and minimize any bias associated with conducting a complete case analysis. Seventy-five students were missing data on past-month tobacco use, seventy-one were missing data on past-month alcohol use, and seventy-six were missing data on past-month marijuana use. These students were excluded from the respective use samples, resulting in a final sample size of 1,217 students for tobacco use, 1,221 for alcohol use, and 1,216 for marijuana use.

We assessed for geographical areas with high rates of tobacco, alcohol, and marijuana use. The spatial scan statistic tests were performed using the SaTScan™ software (34). Because the data represents individual instances of use or abstinence, the Bernoulli model was used. For each substance two input files, cases and controls, were generated based on participants' survey responses. The format of these files was binary in nature, with respondents who answered b) "1-2", c) "3-9", and d) "10 or more" times to past 30 day use being assigned values of 1 in the case files and 0 in the control files, while respondents who answered a) "none", to past use were assigned values of 0 in the case files and 1 in the control files. Location data, in the form of geographic coordinates for each participant was also required. SaTScan™ assessed the patterns of past substance use for spatial clustering by imposing a circular window centered on each observation. The radius of this circle was continuously increased in size in order to analyze clusters of different sizes with different sets of neighboring data locations within them. Each of the circles at each observation is a potential cluster (34).

When using the Bernoulli model, SaTScan™ tests the null hypothesis of constant risk of usage throughout the study area with the alternative hypothesis that there is elevated risk within the circular window as compared to outside of it, e.g. a high proportion of cases. This assessment is made using a likelihood ratio test, and under the Bernoulli model this likelihood function is equal to the following equation:

$$\left(\frac{C}{n}\right)^{c}\left(\frac{n-c}{n}\right)^{n-c}\left(\frac{C-c}{N-n}\right)^{C-c}\left(\frac{(N-n)-(C-c)}{N-n}\right)^{(N-n)(C-c)} I\,()$$

Here, C is the total number of cases, c is the number of observed cases located within the test window, N is equal to the combined total of cases and controls in the data set, and n is equal to the number of cases on controls inside the test window. $I\,()$ is an indicator function equal to 1 for each window (34).

These analyses produced clustering results for tobacco, alcohol, and marijuana based solely on the case-control data. In order to adjust for covariates with the Bernoulli model we used the multiple data set function available in SaTScan™. This function requires separate case and control files for each discrete covariate category used in the analysis, but is limited to a total of 12 categories. To meet these requirements we maintained two groups for gender – male and female, and created 2 groups for age – 13-17 year olds and 18-19 year olds, and three groups for race/ethnicity – White, Black/Hispanic, and Asian/Other, for a total of 12 categories (e.g., 13-17 year old White males). When using multiple data sets SaTScan™ calculated the log likelihood ratio for each of the 12 data sets in each circular scan window. Any log likelihood ratio for data sets with less than the expected number of cases in the scan window were multiplied by negative one. To get the overall log likelihood for a scan window the individual log likelihood ratios for each category were then summed (34).

In both the unadjusted and adjusted analyses SaTScan™ identified the most likely cluster (i.e., that with the highest likelihood ratio test statistic) and additional secondary clusters. Each test performed scanned for areas of high values, or elevated risk. To reduce redundant results clusters whose center were within the boundary of a more likely cluster were not reported. The significance of these clusters was determined using a Monte Carlo simulation consisting of 999 random replications. For each cluster generated without covariate adjustment, the number of observed cases and expected cases, relative risk, and *p*-value were reported. For clusters generated using covariate adjustment only the *p*-value associated with the overall results of the combined categories was reported. These results were mapped using ArcGIS software (ESRI, Redlands, CA).

Results

Descriptive Statistics

Approximately 12% of youth reported any past-month tobacco use, while approximately 36% and 18% of youth reported any past-month alcohol and marijuana use, respectively.

Table 1 presents the proportion of respondents reporting positive substance use by demographic variables and the associated 95% confidence interval. Certain patterns exist across all of the substances. As age increases past use of both alcohol and marijuana increases by approximately 15%, while past use of tobacco remains relatively stable. Approximately 6% more males than females reported past use of tobacco and marijuana, but this difference falls to 2% for past alcohol use. White respondents reported the highest past use and Asian respondents the lowest past use for all three substances. The largest difference in reported past use was associated with tobacco. Twice as many white respondents reported past tobacco use than respondents in the next highest category (Table 1).

Spatial Scan Statistics

Using the Bernoulli method, SaTScan™ generated a number of spatial clusters illustrating alcohol, marijuana, and tobacco use among Boston youths while using both the case-control data by itself and when controlling for the age, gender, and race/ethnicity. These clusters are presented by substance and in order of likelihood. In each analysis Cluster 1 is the most likely cluster.

Tobacco—SaTScan™ generated two clusters (see Table 2 and Figure 1) of elevated past tobacco use among Boston youths. Of these, Cluster 1 was statistically significant. Located in the South Boston neighborhood, the relative risk (RR) for this cluster was 5.37 with a *p*-value of 0.00014.

One statistically significant cluster, Cluster 1 (*p*-value of 0.035) and six non-significant potential clusters (see Table 3 and Figure 2) were identified by scanning for tobacco use when controlling for covariates.

Alcohol—Using only the case-control data, SaTScan™ generated six non-significant potential clusters (see Table 4) for alcohol use.

Six non-significant potential clusters (see Table 5) for past alcohol use were generated when controlling for covariates.

Marijuana—When strictly analyzing case-control data of youth past marijuana use in Boston, SaTScan™ generated two non-significant potential clusters (see Table 6).

When controlling for age, gender, and race/ethnicity three non-significant potential clusters of youth marijuana use were identified (see Table 7).

Overall, we found no significant localized spatial clustering in youth past alcohol or marijuana use in either the unadjusted or adjusted models. However, we found significant localized spatial clustering in youth past tobacco use, highlighting locations of elevated youth substance use with both case-control data alone and covariate adjustment.

Discussion

In this study, we examined spatial clustering of youth use of tobacco, alcohol, and marijuana among high school students in Boston, Massachusetts. The results show that there is elevated use of tobacco use among the youth population in certain areas of Boston, and they indicate that the spatial clustering of these elevated risk behaviors is largely concentrated in south/southeast Boston. When using only case data, and when controlling for age, gender, race/ethnicity, and missing indicator variables, the most likely significant clusters for tobacco use were spatially stable. In these analyses, the most likely cluster appeared in approximately the same position and covered approximately the same area. Rarely have studies used the Bernoulli method with covariate adjustments, further highlighting the novelty of the current study.

It is difficult to compare our results with others, as we aware of only one study evaluating spatial clustering of tobacco use in a population containing youth (15) and none examining alcohol and marijuana use; however, some related work with has been conducted in this area. Généreux and colleagues (2012) assessed spatial differences in smoking between 2003 and 2009 based on social inequality in Montréal, Canada among residents 15 years or older using the Local Moran's *I* test. As the data for each age group in the study were aggregated to the boundaries of local community service centers, it is not possible to compare our results to the youth included in this study; however, education level and income were identified as important factors impacting smoking behaviors. McVie and Norris (2006) mapped youth use of illicit drugs and cannabis among 16 year-olds throughout Edinburgh, Scotland. They used a visualization method and performed regression models to assess the impact of neighborhood-level factors on substance use. Their results indicate that youth illicit drug use was increased in areas of high recorded street crime, and that youth cannabis use was higher in locations with economic prosperity and/or younger, more mobile populations.

What, then, accounts for the spatial patterns observed in this study? Spatial clustering in an attribute (e.g., drug use) can occur due to a spatial interaction (true contagion) or a spatial reaction to a common feature (apparent contagion). Research suggests that one's social network can influence one's patterns of smoking and use of other drugs (35-37), which is an example of true contagion. This finding has been found among adolescents, whose likelihood of smoking, drinking alcohol, and/or using illicit drugs including marijuana rises with the prevalence of substance use in their friend group and social network (38-40). This may be a factor contributing to the spatial patterns in tobacco use observed in our study. However, in addition to social networks, characteristics of the neighborhood environment (apparent contagion) in which substance use takes place may play an important role in shaping patterns of drug use (as briefly discussed). Neighborhood environments with high rates of neighborhood poverty have been found to be associated with higher prevalence of drug use (41, 42). Residents of such disadvantaged neighborhood environments have elevated risk of exposure to a broad range of psychosocial stressors. Animal and human studies have found that chronic stress early in life and during early adulthood predicts alcohol and drug dependence in later adulthood (43-45). High rates of neighborhood violence, physical disorder (e.g., broken windows, vandalism, liter, empty alcohol containers), and social disorder (e.g., alcohol use, prostitution, drug addiction), especially within urban areas, are also associated with elevated rates of drug use (46-50). Several studies have shown that such neighborhood characteristics shape adolescents' substance use behaviors, including the initiation of alcohol and marijuana use during adolescence (40, 41, 46, 48). Reboussin and colleagues (2014) examined the relationship between urban neighborhood environments and marijuana use during high school among African American youth. Through a longitudinal survey study conducted in Baltimore, MD, they found that urban neighborhoods with increased drug activity, violence, and neighborhood disorder heightened the likelihood of youth either starting to use, or increasing their use, of marijuana. In one recent study using data from Monitoring the Future, U.S. high school seniors living in areas with high rates of perceived drug sales reported higher rates of illicit drug use than their counterparts living in areas without "open" (visible) neighborhood drug selling (4). While there may be some overlap between one's social network and one's neighborhood environment (e.g., individuals may socialize in their neighborhood), past evidence shows that social networks and characteristics of the neighborhood environment independently play an important role in shaping patterns of drug use; therefore, both causes should be probed as possible explanations for our findings as well as simultaneously exploring both true contagion and apparent contagion theories by simultaneously examining one's neighborhood environment and their social network, which synergistically might have greater effects on heath and behavior than one singly (52).

Study Implications and Future Research

This research can be used for local targeting of drug abuse prevention interventions and in particular tobacco use prevention interventions. Understanding the reasons for spatial clustering in youth drug use indeed may not just guide the geographic locations for interventions, but also intervention development. As previously mentioned, spatial clustering in an attribute can occur due to a spatial interaction (true contagion) or due to a spatial reaction to a common feature (apparent contagion). For example, spatial interaction processes include neighborhood peer effects, which could occur when youth interact with other youth in their neighborhoods and this induces similar levels of health and wellbeing (such as tobacco use). If environmental factors, on the other hand, influence the attribute (e.g. tobacco use), the process would be a spatial reaction process. Disentangling these effects remains methodologically difficult. Future research should be conducted to evaluate potential sources of spatial clustering in youth drug use including in South Boston. South Boston is 76.3% White, but this number has been dropping in recent years. In order to investigate the relevance of South Boston to the presence or absence of such clustering would require additional data on social, behavioral, and environmental risk factors not collected in the survey we took our data from. To evaluate the true contagion hypothesis, researchers would need data on social networks (which may be difficult to obtain). An evaluation of apparent contagion would include examining factors such as alcohol retailers and crime as correlates of youth alcohol use.

Given increased geo-computational abilities in recent years, conducting statistical tests to detect spatial clusters using desktop computers is now possible, and can be done relatively quickly. However, multiple spatial clustering methods exist, and there is little guidance as to which methods to use. The Local Moran's *I* test is the most widely used method for evaluating local autocorrelation across space (11). Other popular methods include *G*-statistics and spatial scan statistics such as those used here. Few studies have compared cluster detection methods, and those that do, e.g., (53), illustrate the need to compare and assess the validity of results from multiple spatial cluster detection methods, including newly developed methods. Future research on methodology should focus on identifying optimal cluster detection methods given the type and level of aggregation of available data and the spatial scale being analyzed. As previously mentioned, several studies have shown that aggregating data can negatively impact cluster detection (25, 26); and it is an axiom in geography that studying a phenomenon at different spatial scales often requires different analytical approaches.

Study Limitations

The limitations of this study should be noted. Reliance on self-report of drug use is one of them. We recognize that some selection bias might exist in that youth with high levels of drug use may not have taken the survey. It is also important to highlight that positional accuracy is important in spatial analysis, and errors can exist in spatial datasets, in part due to geocoding methods (54). The intersection addresses we obtained may also contribute to location misclassification. However, the effect of using intersections on location misclassification is likely to be minimal, since all study subjects live in an urban environment, which generally has a dense street network with small block sizes. Importantly, we found no evidence of geographic bias (55) because there were no differences by the drug use outcomes with regards to who provided geocodeable information and who did not. Results from this study might only be generalizable to low-income youths in similar urban locations at similar spatial scales. A limitation of using circular scanning windows is that the cluster may be irregularly shaped, and as such, using a circular window may miss a cluster. Recently, other methods have been used to improve the detection of clusters with arbitrary shapes. Tango and Takahashi (2005) and others (57, 58) presented a flexibly shaped spatial scan statistic which is used to detect arbitrarily shaped clusters by constructing scanning windows of irregular shapes. Finally, the survey did not have information on certain variables such as parental education or parental income, which could be a source of residual confounding.

Conclusion

Significant spatial clustering in youth tobacco use were found. These findings can be used for geographically targeting of drug abuse prevention interventions, which is seldom done. Future research should evaluate the reasons for youth drug use spatial clustering, e.g., whether spatial clusters of youth drug use are caused by neighborhood peer-effects and/or neighborhood environmental factors (e.g., alcohol and tobacco outlets, drug trafficking), and determine the best methods for analyzing spatial clusters.

Acknowledgements

The 2008 Boston Youth Survey (BYS) was conducted in collaboration with the Boston Public Health Commission (Barbara Ferrer, Director), Boston's Office of Human Services (Larry Mayes, Chief), Boston Public Schools (Carol Johnson, Superintendent) 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 dataset. We thank Daniel Hagen for reviewing an early draft of this manuscript, and we thank Thayane Moreira and William Goedel for assistance with formatting the references.

Sources of Support: Dr. Dustin Duncan was supported by his New York University School of Medicine Start-Up Research Fund to work on this project. 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 (David Hemenway, PhD, Principal Investigator). A grant to Dustin Duncan from Robert Wood Johnson Foundation's Active Living Research Program (Grant # 67129)supported the development of the Boston Youth Survey Geospatial Dataset. This study was supported by the Robert Wood Johnson Foundation Health and Society Scholars Seed Grant Program (grant to Dustin Duncan), Harvard Center for Population and Development Studies, Harvard School of Public Health.

References

1. Gruber AJ, Pope HG. Marijuana use among adolescents. Pediatric Clinics of North America. 2002; 49(2):389–413. [PubMed: 11993290]
2. Chen C-Y, O'Brien MS, Anthony JC. Who becomes cannabis dependent soon after onset of use? Epidemiological evidence from the United States: 2000–2001. Drug and Alcohol Dependence. 2005; 79(1):11–22. [PubMed: 15943940]
3. Behrendt S, Wittchen H-U, Höfler M, Lieb R, Beesdo K. Transitions from first substance use to substance use disorders in adolescence: Is early onset associated with a rapid escalation? Drug and Alcohol Dependence. 2009; 99(1):68–78. [PubMed: 18768267]
4. Duncan DT, Palamar JJ, Williams JH. Perceived neighborhood illicit drug selling, peer illicit drug disapproval and illicit drug use among US high school seniors. Substance Abuse Treatment, Prevention, and Policy. 2014; 9(1):35.
5. Duncan DT, Hatzenbuehler ML, Johnson RM. Neighborhood-level LGBT hate crimes and current illicit drug use among sexual minority youth. Drug and Alcohol Dependence. 2014; 135:65–70. [PubMed: 24326203]
6. Kelder SH, Perry CL, Klepp K-I, Lytle LL. Longitudinal tracking of adolescent smoking, physical activity, and food choice behaviors. American Journal of Public Health. 1994; 84(7):1121–1126. [PubMed: 8017536]
7. Chassin L, Presson CC, Rose JS, Sherman SJ. The natural history of cigarette smoking from adolescence to adulthood: demographic predictors of continuity and change. Health Psychology. 1996; 15(6):478. [PubMed: 8973929]
8. Chassin L, Presson CC, Pitts SC, Sherman SJ. The natural history of cigarette smoking from adolescence to adulthood in a midwestern community sample: multiple trajectories and their psychosocial correlates. Health Psychology. 2000; 19(3):223. [PubMed: 10868766]
9. Chassin L, Flora DB, King KM. Trajectories of alcohol and drug use and dependence from adolescence to adulthood: the effects of familial alcoholism and personality. Journal of Abnormal Psychology. 2004; 113(4):483. [PubMed: 15535782]
10. Tobler WR. A computer movie simulating urban growth in the Detroit region. Economic Geography. 1970:234–240.
11. Waller, LA.; Gotway, CA. Applied Spatial Statistics for Public Health Data. Wiley-Interscience; Hoboken: 2004.
12. Cromley, EK.; McLafferty, SL. GIS and Public Health. 2nd Edition. Guilford Press; New York, NY: 2011.
13. Petronis KR, Anthony JC. A different kind of contextual effect: geographical clustering of cocaine incidence in the USA. Journal of epidemiology and community health. 2003; 57(11):893–900. [PubMed: 14600117]

14. McVie, S.; Norris, P. Neighbourhood Effects on Youth Delinquency and Drug Use. Centre for Law and Society, University of Edinburgh; Edinburgh: 2006.

15. Généreux M, Roy M, Montpetit C, Azzou SAK, Gratton J. Regional surveillance of social and geographic inequalities in smoking: The case of Montréal, Canada. Health & Place. 2012; 18(2): 240–249. [PubMed: 22019850]

16. Chaney RA, Rojas-Guyler L. Spatial patterns of adolescent drug use. Applied Geography. 2015; 56:71–82.

17. Subramanian S, Nandy S, Kelly M, Gordon D, Smith GD. Patterns and distribution of tobacco consumption in India: cross sectional multilevel evidence from the 1998-9 national family health survey. BMJ. 2004; 328(7443):801–806. [PubMed: 15070637]

18. Tomintz MN, Clarke GP, Rigby JE. The geography of smoking in Leeds: estimating individual smoking rates and the implications for the location of stop smoking services. Area. 2008; 40(3): 341–353.

19. Chahine T, Subramanian S, Levy J. Sociodemographic and geographic variability in smoking in the US: A multilevel analysis of the 2006–2007 Current Population Survey, Tobacco Use Supplement. Social Science & Medicine. 2011; 73(5):752–758. [PubMed: 21813218]

20. Hermes K, Poulsen M. Small area estimates of smoking prevalence in London. Testing the effect of input data. Health & Place. 2012; 18(3):630–638. [PubMed: 22281441]

21. Corsi DJ, Lear SA, Chow CK, Subramanian S, Boyle MH, Teo KK. Socioeconomic and geographic patterning of smoking behaviour in Canada: a cross-sectional multilevel analysis. PloS ONE. 2013; 8(2):e57646. [PubMed: 23469038]

22. Linder JA, Rigotti NA, Brawarsky P, Knotos EZ, Park ER, Klinger EV, Marinacci L, Li W, Hass JS. Use of Practice-Based Research Network Data to Measure Neighborhood Smoking Prevalence. Preventing Chronic Disease: Public Health Research, Practice, and Policy. 2013; 10:1–10.

23. Sreeramareddy CT, Ramakrishnareddy N, Kumar H, Sathian B, Arokiasamy J. Prevalence, distribution and correlates of tobacco smoking and chewing in Nepal: a secondary data analysis of Nepal Demographic and Health Survey-2006. Substance Abuse Treatment, Prevention, and Policy. 2011; 6(33.10):1186.

24. Li W, Land T, Zhang Z, Keithly L, Kelsey JL. Small-area estimation and prioritizing communities for tobacco control efforts in Massachusetts. American Journal of Public Health. 2009; 99(3):470. [PubMed: 19150913]

25. Ozonoff A, Jeffery C, Manjourides J, White LF, Pagano M. Effect of spatial resolution on cluster detection: a simulation study. International Journal of Health Geographics. 2007; 6(1):52. [PubMed: 18042281]

26. Luo L. Impact of spatial aggregation error on the spatial scan analysis: a case study of colorectal cancer. Geospatial Health. 2013; 8(1):23–35.

27. Duncan DT, Kawachi I, Subramanian S, Aldstadt J, Melly SJ, Williams DR. Examination of how neighborhood definition influences measurements of youths' access to tobacco retailers: a methodological note on spatial misclassification. American Journal of Epidemiology. 2014; 179(3):373–381. [PubMed: 24148710]

28. Azrael D, Johnson RM, Molnar BE, Vriniotis M, Dunn EC, Duncan DT, Hemenway D. Creating a youth violence data system for Boston, Massachusetts. Australian & New Zealand Journal of Criminology. 2009; 42(3):406–421.

29. Duncan DT, Castro MC, Gortmaker SL, Aldstadt J, Melly SJ, Bennett GG. Racial differences in the built environment–body mass index relationship? A geospatial analysis of adolescents in urban neighborhoods. International Journal of Health Geographics. 2012; 11(11):10.1186. [PubMed: 22471893]

30. Duncan DT, Aldstadt J, Whalen J, Melly SJ. Validation of Walk Scores and Transit Scores for estimating neighborhood walkability and transit availability: a small-area analysis. GeoJournal. 2013; 78(2):407–416.

31. Green JG, Dunn EC, Johnson RM, Molnar BE. A multilevel investigation of the association between school context and adolescent nonphysical bullying. Journal of School Violence. 2011; 10(2):133–149. [PubMed: 21532943]

32. Brener ND, Collins JL, Kann L, Warren CW, Williams BI. Reliability of the youth risk behavior survey questionnaire. American Journal of Epidemiology. 1995; 141(6):575–580. [PubMed: 7900725]

33. Aldstadt, J. Spatial Clustering. In: Fischer, M.; Getis, A., editors. Handbook of Applied Spatial Analysis: Software Tools, Methods and Applications. Springer; New York, NY: 2010. p. 279-300.

34. Kulldorff, M. SaTScan User Guide for Version 9.4. 2015.

35. Valente TW. Social network influences on adolescent substance use: An introduction. Connections. 2003; 25(2):11–16.

36. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. New England Journal of Medicine. 2008; 358(21):2249–2258. [PubMed: 18499567]

37. Fujimoto K, Valente TW. Social network influences on adolescent substance use: Disentangling structural equivalence from cohesion. Social Science & Medicine. 2012; 74(12):1952–1960. [PubMed: 22475405]

38. Alexander C, Piazza M, Mekos D, Valente T. Peers, schools, and adolescent cigarette smoking. Journal of Adolescent Health. 2001; 29(1):22–30. [PubMed: 11429302]

39. Huang GC, Soto D, Fujimoto K, Valente TW. The interplay of friendship networks and social networking sites: longitudinal analysis of selection and influence effects on adolescent smoking and alcohol use. American Journal of Public Health. 2014; 104(8):e51–e59. [PubMed: 24922126]

40. Tucker JS, De La Haye K, Kennedy DP, Green HD, Pollard MS. Peer influence on marijuana use in different types of friendships. Journal of Adolescent Health. 2014; 54(1):67–73. [PubMed: 24054813]

41. Duncan SC, Duncan TE, Strycker LA. A multilevel analysis of neighborhood context and youth alcohol and drug problems. Prevention Science. 2002; 3(2):125–133. [PubMed: 12088137]

42. Molina KM, Alegría M, Chen C-N. Neighborhood context and substance use disorders: a comparative analysis of racial and ethnic groups in the United States. Drug and Alcohol Dependence. 2012; 125:S35–S43. [PubMed: 22699095]

43. Becker HC, Lopez MF, Doremus-Fitzwater TL. Effects of stress on alcohol drinking: a review of animal studies. Psychopharmacology. 2011; 218(1):131–156. [PubMed: 21850445]

44. Enoch M-A. The role of early life stress as a predictor for alcohol and drug dependence. Psychopharmacology. 2011; 214(1):17–31. [PubMed: 20596857]

45. Mulia N, Schmidt L, Bond J, Jacobs L, Korcha R. Stress, social support and problem drinking among women in poverty. Addiction. 2008; 103(8):1283–1293. [PubMed: 18855817]

46. Lambert SF, Brown TL, Phillips CM, Ialongo NS. The relationship between perceptions of neighborhood characteristics and substance use among urban African American adolescents. American Journal of Community Psychology. 2004; 34(3-4):205–218. [PubMed: 15663207]

47. Latkin CA, Williams CT, Wang J, Curry AD. Neighborhood social disorder as a determinant of drug injection behaviors: a structural equation modeling approach. Health Psychology. 2005; 24(1):96. [PubMed: 15631567]

48. Copeland-Linder N, Lambert SF, Chen Y-F, Ialongo NS. Contextual stress and health risk behaviors among African American adolescents. Journal of Youth and Adolescence. 2011; 40(2): 158–173. [PubMed: 20213481]

49. Kuipers MA, van Poppel MN, van den Brink W, Wingen M, Kunst AE. The association between neighborhood disorder, social cohesion and hazardous alcohol use: a national multilevel study. Drug and Alcohol Dependence. 2012; 126(1):27–34. [PubMed: 22572208]

50. Patterson F, Seravalli L, Hanlon A, Nelson DB. Neighborhood safety as a correlate of tobacco use in a sample of urban, pregnant women. Addictive Behaviors. 2012; 37(10):1132–1137. [PubMed: 22688344]

51. Reboussin BA, Green KM, Milam AJ, Furr-Holden CDM, Ialongo NS. Neighborhood environment and urban African American marijuana use during high school. Journal of Urban Health. 2014; 91(6):1189–1201. [PubMed: 25323775]

52. Duncan DT. Putting Health and Behavior in Context: Current Knowledge and Future Directions on Research Embracing Ecological Perspectives. Behavioral Medicine. 2015; 41(3):172–176. [PubMed: 26332934]

53. Hanson CE, Wieczorek WF. Alcohol mortality: a comparison of spatial clustering methods. Social Science & Medicine. 2002; 55(5):791–802. [PubMed: 12190271]

54. Duncan DT, Castro MC, Blossom JC, Bennett GG, Gortmaker SL. Evaluation of the positional difference between two common geocoding methods. Geospatial health. 2011; 5(2):265–273. [PubMed: 21590677]

55. Oliver MN, Matthews KA, Siadaty M, Hauck FR, Pickle LW. Geographic bias related to geocoding in epidemiologic studies. International Journal of Health Geographics. 2005; 4:29. [PubMed: 16281976]

56. Tango T, Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. International Journal of Health Geographics. 2005; 4(1):11. [PubMed: 15904524]

57. Duczmal L, Cancado AL, Takahashi RH, Bessegato LF. A genetic algorithm for irregularly shaped spatial scan statistics. Computational Statistics & Data Analysis. 2007; 52(1):43–52.

58. Costa MA, Assunção RM, Kulldorff M. Constrained spanning tree algorithms for irregularly-shaped spatial clustering. Computational Statistics & Data Analysis. 2012; 56(6):1771–1783.

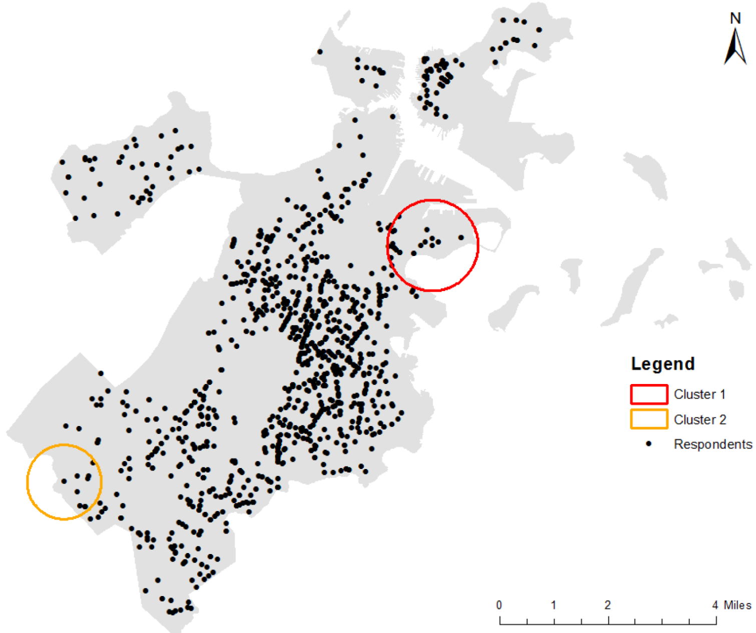

Figure 1.
Spatial clustering of tobacco use cases (unadjusted).

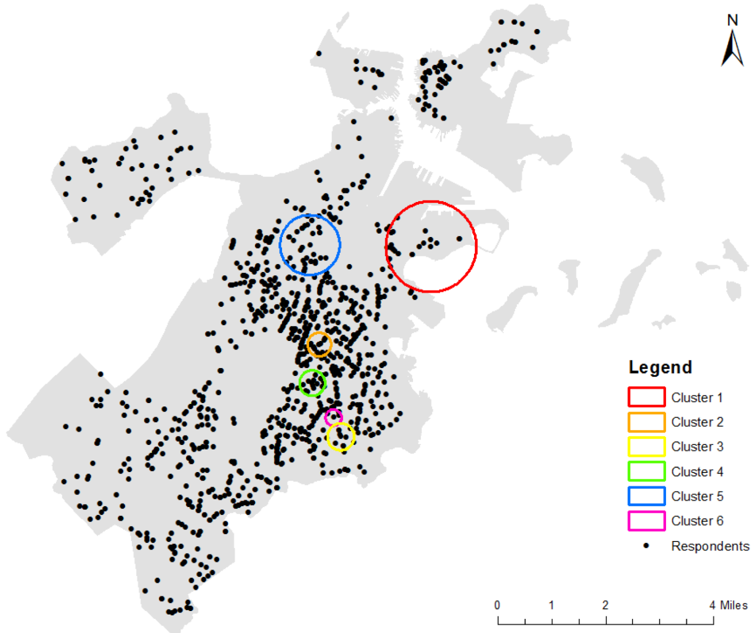

Figure 2.
Spatial clustering of tobacco use cases when adjusting for age, gender, and race/ethnicity categories.

Table 1

Percentage and 95% confidence interval of substance users by age, gender, and race/ethnicity for past-month tobacco, alcohol, and marijuana.

		Tobacco % (C.I.)	**Alcohol % (C.I.)**	**Marijuana % (C.I.)**
Age	≤ 14 yrs.	15.2 (±2.0)	34.8 (±2.7)	10.9 (±1.8)
	15 years	10.7 (±1.7)	32.2 (±2.6)	16.5 (±2.1)
	16 years	11.1 (±1.8)	38.7 (±2.7)	21.0 (±2.3)
	17 years	5.4 (±1.3)	41.6 (±2.8)	21.6 (±2.3)
	18 years	14.2 (±2.0)	41.4 (±2.8)	21.0 (±2.3)
	19 years	17.0 (±2.1)	49.1 (±2.8)	26.4 (±2.5)
Gender	Male	15.3 (±2.0)	39.9 (±2.8)	23.2 (±2.4)
	Female	10.7 (±1.7)	37.3 (±2.7)	16.7 (±2.1)
Race/ Ethnicity	White	33.9 (±2.7)	44.6 (±2.8)	27.3 (±2.5)
	Black	9.9 (±1.7)	36.0 (±2.7)	19.5 (±2.3)
	Hispanic	11.1 (±1.8)	45.2 (±2.8)	19.4 (±2.3)
	Asian	8.2 (±1.6)	24.7 (±2.4)	8.2 (±1.6)
	Other*	16.5 (±2.1)	34.9 (±2.7)	24.7 (±2.5)

*Includes non-Hispanic youth who were bi- or multi-racial, American Indian or Alaska Native, Native Hawaiian or Other Pacific Islander, or youth who did not fit into any of the specified race/ethnicity categories.

Note: Based on the total number of respondents who answered the past use question for each drug. Also, we collapsed 13 and 14 years due to the limited the number of 13 year olds.

Table 2

Number of cases, relative risk, and significance associated with spatial clusters of tobacco use based on unadjusted analysis.

Cluster ID	Observed Cases	Expected Cases	Relative Risk	*p*-value
1	14	2.81	5.37	<0.05
2	6	1.66	3.71	0.97

Table 3

Significance associated with spatial clusters of tobacco use when adjusting for age, gender, and race/ethnicity categories.

Cluster ID	*p*-value
1	<0.05
2	0.69
3	0.98
4	0.99
5	0.99
6	0.99
7	0.99

Table 4

Number of cases, relative risk, and significance associated with spatial clusters of alcohol based on unadjusted analysis.

Cluster ID	Observed Cases	Expected Cases	Relative Risk	*p*-value
1	9	3.47	2.63	0.11
2	89	66.25	1.42	0.36
3	31	19.26	1.65	0.75
4	10	4.62	2.19	0.92
5	9	4.24	2.15	0.99
6	23	14.64	1.60	0.99

Table 5

Significance associated with spatial clusters of alcohol when adjusting for age, gender, and race/ethnicity categories.

Cluster ID	*p*-value
1	0.34
2	0.65
3	0.91
4	0.93
5	0.99
6	0.99

Table 6

Number of cases, relative risk, and significance associated with spatial clusters of marijuana based on unadjusted analysis.

Cluster ID	Observed Cases	Expected Cases	Relative Risk	*p*-value
1	10	3.34	3.09	0.65
2	91	70.24	1.48	0.86

Table 7

Number of cases, relative risk, and significance associated with spatial clusters of marijuana when adjusting for age, gender, and race/ethnicity categories

Cluster ID	*p*-value
1	0.370
2	0.948
3	0.985