available at www.sciencedirect.com

Public Health

journal homepage: www.elsevier.com/puhe

Original Research

Tobacco outlet density and demographics: Analysing the relationships with a spatial regression approach

D. Yu [a,*], *N.A. Peterson* [b], *M.A. Sheffer* [c], *R.J. Reid* [a], *J.E. Schnieder* [d]

[a] *Montclair State University, Montclair, NJ, USA*
[b] *Rutgers University, NJ, USA*
[c] *University of Wisconsin – Madison, USA*
[d] *Health Economics Consulting Group, LLC, Morrisville, NJ, USA*

ARTICLE INFO

Article history:
Received 2 October 2009
Received in revised form
11 February 2010
Accepted 24 March 2010
Available online 11 June 2010

Keywords:
Tobacco outlet density
Spatial regression
Demographic factors
New Jersey

SUMMARY

Objective: Studies of relationships between tobacco sales and socio-economic/sociodemographic characteristics are well documented. However, when analysing the data that are collected on geographic areas, the spatial effects are seldom considered, which could lead to potential misleading analytical results. This study addresses this concern by applying the spatial analysis method in studying how socio-economic factors and tobacco outlet density are related in New Jersey, USA.

Study design: A spatial regression method applied to tobacco outlet and socio-economic data obtained in 2004 in New Jersey, USA.

Method: This study assessed the association between tobacco outlet density and three demographic correlates – income, race and ethnicity – at the tract level of analysis for one state in the north-eastern USA. Data for 1938 residential census tracts in the state of New Jersey were derived from 2004 licences for 13,984 tobacco-selling retail outlets. Demographic variables were based on 2000 census data. When applying a regression model, the residuals of an ordinary least squared (OLS) estimation were found to exhibit strong spatial autocorrelation, which indicates that the estimates from the OLS model are biased and inferences based on the estimates might be misleading. A spatial lag model was employed to incorporate the potential spatial effects explicitly.

Results: Agreeing with the OLS residual autocorrelation test, the spatial lag model yields a significant coefficient of the added spatial effect, and fits the data better than the OLS model. In addition, the residuals of the spatial regression model are no longer autocorrelated, which indicates that the analysis produces more reliable results. More importantly, the spatial regression results indicate that tobacco companies attempt to promote physical availability of tobacco products to geographic areas with disadvantageous socio-economic status. In New Jersey, the percentage of Hispanics seems to be the dominant demographic factor associated with tobacco outlet distribution, followed by median household income and percentage of African Americans.

Conclusion: This research applied a spatial analytical approach to assess the association between tobacco outlet density and sociodemographic characteristics in New Jersey at the census tract level. The findings support the common wisdom in the public health research

* *Corresponding author.*
E-mail address: yud@mail.montclair.edu (D. Yu).

doi:10.1016/j.puhe.2010.03.024

domain that tobacco outlets are more densely distributed in socio-economically disadvantaged areas. However, incorporating the spatial effects explicitly in the analysis provides less biased and more reliable results than traditional methods.

Introduction

Economic theory posits that the net price faced by consumers of products such as tobacco is, in part, a function of search costs.[1,2] Search costs generally include costs associated with obtaining price and quality information,[3,4] but also include other costs associated with transacting, such as time and distance travelled by individuals to the point of sale.[5,6] Thus, search costs for tobacco consumers are expected to be lower as the density of tobacco-selling retail outlets increases.

The relationship between retail density and consumption has been established for tobacco use.[7–9] In addition, it is generally found that tobacco outlets tend to be disproportionately located in neighbourhoods that are characterized with socio-economic disadvantage.[8,10–12] However, the extension of state and local policies to tobacco zoning has thus far been limited. The majority of those policies focus on licence revocation in the event of tobacco sales to minors.[13] In a recent review article discussing the role of land use planning in the control of alcohol, tobacco, firearms and fast food, Ashe *et al.*[14] maintain that most states do not have laws that would pre-empt local regulation of tobacco sales, and that there is a potentially effective role for local tobacco zoning policies in access restrictions targeted at areas frequented by children and other price-sensitive populations.

Laws *et al.*[15] gathered data on tobacco outlets from neighbourhoods in Boston, Massachusetts. They showed that neighbourhoods with higher percentages of businesses that sold tobacco were more likely to have lower average per capita income, and that residents in neighbourhoods with lower income were predominately Latino or African American. Hyland *et al.*[16] used licensing data from residential census tracts in Erie County, New York to examine whether tobacco outlet density was associated with income and race. Similar to the Boston study, they found that tracts with lower median household income and a higher percentage of African Americans had greater densities of tobacco-selling retail outlets. These findings are supported by Hackbarth *et al.*,[17] who found that African American and Hispanic neighbourhoods in the Chicago area were disproportionately targeted for outdoor advertising of alcohol and tobacco. A recent review by Schaap and Kunst[18] also indicates that smoking prevalence is highly associated with specific socio-economic groups, which are usually grouped based on ethnicity and household income in the USA.

Recent studies have examined the issue of tobacco outlet density at the county and the census block group level.[10,11] These studies support previous research showing that areas with a higher percentage of minority residents had greater tobacco outlet density.[16,17] The majority of the studies, however, have rather limited discussion of the potential spatial effects[19] when using data collected based upon geographic units such as census tracts. However, it is extensively discussed in the fields of geography, spatial statistics and spatial econometrics that ignoring possible spatial effects in data analysis might lead to potential unreliable and even misleading results.[19–28]

The present study extends previous research by assessing the association between tobacco outlet density and three demographic variables – income, race and ethnicity – at the census tract level. Differing from previous studies, this study will examine such associations from a spatial analysis standpoint. The existence of spatial effects in tobacco outlet distribution in New Jersey will be examined, explicitly taking into account the spatial effects by a spatial regression model. Given the emphasis on controlling tobacco use and reducing health disparities in public health, as well as the lack of research on the topic of tobacco outlet density,[14] the objective is to test the validity of the demographic–tobacco outlet link by investigating associations from a spatial analytical standpoint. By incorporating the spatial effects explicitly in the model, the authors believe that the approach models the relationship more reliably.

Methods

Addresses of all 15,037 licensed tobacco-selling retail outlets in the state of New Jersey in 2004 were obtained from the New Jersey Department of the Treasury. The 2000 TIGER/Line files were used to extract 2000 census data for the 1938 residential census tracts in the state and geocode the licensed tobacco-selling retail outlets; 13,984 addresses have been successfully geocoded. The total population of the state of New Jersey in 2000 was 8,414,350, with 1,211,750 individuals who were African America (14.4%) and 1,117,191 individuals who were Hispanic or Latino (13.3%). Following prior research conducted at the tract level,[16,29] the number of tobacco outlets per 10 km of road was used as the primary density measure. All the public streets in New Jersey available in the 2000 census TIGER road file (see http://www.state.nj.us/dep/gis/tgr2000shp.html) were included in the calculation. Census tracts in New Jersey have an average size of 10.4 km^2 and an average population (in 2000) of 4348 people. The choice of census tract in this study is a balance between representation of neighbourhood and data manageability. Smaller units such as the census block or block groups would render too much variation, increasing analytical instability, while larger units such as counties would aggregate data too much and prevent the analysis from being useful. Median household income, the percentage of African American residents and the percentage of Hispanic residents were based on 2000 census data.

Regression analysis was employed in this study to examine the tobacco outlet–demographic link. In this research, it is argued that tobacco companies attempt to promote physical availability of tobacco products by increasing the amount or density of tobacco-selling retail outlets in geographic areas with disadvantageous socio-economic status.[30–33] The regression models intend to establish a statistic relationship between the density of tobacco outlets (dependent variable) and the percentage of African Americans, the percentage of Hispanics, and the median income at the census tract level. Previous studies in both the tobacco and alcohol literature tend to point out a casual relationship between the availability of tobacco or alcohol products and the neighbourhood socio-economic status.[8,10,11,34,35] Longitudinal analysis with time series data would be required to determine the existence of a casual relationship.[34] This certainly merits further discussion but is beyond the scope of the current study.

During the preliminary data analyses, it was found that the dependent variable (tobacco outlet density) was highly skewed. Such data will likely violate the specification assumptions of ordinary least squared (OLS) regression. Moreover, a simple spatial autocorrelation test of the regression residuals based on raw data indicated that there was significant spatial autocorrelation, which points to the possible misspecification problem of using raw data. Based on this observation, a base 10 logarithm transformation on the dependent variable (tobacco outlet density) was performed to reduce the skewness of the dependent variable (more details are provided in Table 1).

However, even after transformation, it was found that the residual of the regression was still highly spatially autocorrelated. Testing spatial autocorrelation using Moran's I yields a significant value of 0.29 at a confidence level of 99%. This is to be expected when using cross-sectional data on a geographic area.[20–22] The existence of spatial autocorrelation intuitively indicates that some information is repeated, and the degrees of freedom are less than that assumed by the OLS regression for an independently distributed variable.[22,26]

Table 1 – Analytical results for the spatial lag model and comparison with ordinary least squared (OLS) regression.

	Estimate	SE	z-value	Pr(>|z|)
(Intercept)	0.35	0.032	10.91	0.00
Percentage of African Americans	0.13	0.038	3.36	0.00
Percentage of Hispanics	0.80	0.059	13.52	0.00
Median household income	−4.56E-06	4.11E-07	−11.14	0.00

SE, standard error.
Number of observations: 1938.
Dependent variable: base 10 logarithm transformed tobacco outlet density (transformation offset: 0.3).
Spatial effect coefficient, Rho: 0.48459, Likelihood Ratio (LR) test value: 361.51, $P < 2.2\text{e-}16$.
Akaike Information Criterion (AIC): 1485.2 (AIC for OLS regression: 1844.7).
Lagrange Multiplier test for residual autocorrelation test value: 1.5017, $P = 0.22041$ (test for OLS residual autocorrelation, $P < 2.2\text{e-}16$).

This implies that application of the OLS regression technique is not appropriate for the study data. The actually decreased degrees of freedom of the regression model will likely lead to biased and/or misleading coefficient estimations and inferences. In practice, maximum likelihood based spatial regression[19] is often employed to rectify such problems.

Discussion of spatial regression techniques has been booming recently in the field of spatial econometrics.[19,20,22] They were developed in the recognition that observations over geographical space are likely to be correlated with one another. To incorporate such information practically into the modelling scheme, certain spatial structure needs to be imposed.[19,22] Such structure is usually established through defining a set of neighbours for each observation, while members in the neighbourhood are those that have potential interaction with the one in question. It has been noted that definition of the neighbourhood is 'a matter of considerable arbitrariness'.[22] Since this study focused on how incorporating spatial effects will impact on the analytical results, without losing generality, a geographic adjacency approach was taken. Noting that the basic analytical spatial units are census tracts, the geographic adjacency approach means that the neighbourhood for a particular observation is the collection of other census tracts with which it shares borders. A spatial weights matrix was derived from such definition to represent such spatial structure numerically. Elements of the matrix were row-standardized to facilitate interpretation and computation.[19]

Two spatial regression analyses are often mentioned in the spatial analysis literature[19–21]: the spatial lag and spatial error models. Spatial autocorrelation in the residuals could be the result of autocorrelation in the dependent variable (spatial lag) or autocorrelation in the error term due to spatially autocorrelated predictors are not included in the model (spatial error). The Lagrange Multiplier test and Robust Lagrange Multiplier index are able to point to the more appropriate model; the higher the Robust Lagrange Multiplier index, the greater the appropriateness of the model.[21] Since R^2 is no longer applicable in spatial regression models due to the residual autocorrelation,[27] comparison of spatial regressions and OLS regression is based on the Akaike Information Criterion (AIC); the lower the AIC, the better the fit of the model.[21] A drop of 3 in the AIC value is usually deemed to be an acceptable improvement.[36] All the calculations were conducted using the SPDEP package[37] in the R environment.[38] The Robust Lagrange Multiplier index indicates that the lag model tends to be more significant, hence the spatial lag model is the more appropriate specification of the data, indicating that the spatial autocorrelation in the OLS model's residuals is likely to be the result of a spatially autocorrelated dependent variable (tobacco outlet density at census tract level).

Results

The results of the spatial lag regression are presented in Table 1. Four points stand out immediately. First, by comparing the AICs, it can be seen that the spatial lag model fits the data much better than the OLS model (a drop from OLS 1844.7 to spatial lag model 1485.2). Second, the significant coefficient of the spatial effect, Rho, agrees with the authors' previous test that spatial effects need to be taken into consideration explicitly. Third, it is also important to note that the residual of the spatial lag model is no longer spatially autocorrelated ($P = 0.22041$), whereas in the OLS case, the residual is significantly autocorrelated ($P < 2.2e\text{-}16$, Table 1). This further confirms that the spatial lag specification of the model successfully incorporates the essential spatial effects in the modelling scheme, and hence provides more appropriate analytical results. Fourth, the magnitude of the z-values clearly indicates that in New Jersey, of the three sociodemographic variables, the percentage of Hispanics is the most important variable associated with tobacco outlet density, followed by median household income and the percentage of African Americans. This finding indicates that tobacco companies attempt to promote the physical availability of tobacco products to geographic areas with disadvantageous socio-economic status. The z-values further reveal that tobacco companies' promotion efforts are rather sensitive to the Hispanic population due to the relatively low smoking prevalence among Hispanics.[39] The coefficients indicate that a 1% increase in Hispanics in the census tract could result in a 6.3 unit increase in tobacco outlet density. In addition, a 1% change in African Americans could result in a 1.3 unit increase in tobacco outlet density, and a $10,000 increase in median household income in the neighbourhood could result in a 0.9 unit decrease in tobacco outlet density.

Discussion

Due to the recognition that ignoring spatial effects may lead to unreliable estimates, the purpose of this study was to apply a spatial analytic approach to assess the association between tobacco outlet density and three demographic correlates – income, race and ethnicity – at the tract level for one state in the north-eastern USA. Results of spatial lag regression revealed that tobacco outlet density is higher in census tracts where there are more minority groups and more people with lower socio-economic status. These results agree with previous analyses of tobacco outlet distribution and demographic characteristics. These findings, however, differ from previous research and have critical implications for future research and practice. By explicitly taking into account the inherent spatial association in the observations, this study has been able to produce more reliable estimates in determining the relationships between tobacco outlet and demographic variables. Public health policy makers should give careful consideration to this spatial analytic approach when formulating tobacco control measures that attempt to close the tobacco use and disparities gap, since inaccurate analysis could potentially render policies ineffective.

Ethical approval

None sought.

Funding

None declared.

Competing interests

None declared.

REFERENCES

1. Johnson RN. Search costs, lags and prices at the pump. *Rev Indust Organ* 2004;**20**:33–50.
2. Stigler GJ. The economics of information. In: Stigler GJ, editor. *The organization of industry*. Chicago: University of Chicago Press; 1968. p. 171–90.
3. Lee JM. Effect of a large increase in cigarette tax on cigarette consumption: an empirical analysis of cross-sectional survey data. *Public Health* 2008;**122**:1061–7.
4. Ahmad S, Franz GA. Raising taxes to reduce smoking prevalence in the US: a simulation of the anticipated health and economic impacts. *Public Health* 2008;**122**:3–10.
5. Clarke PM. Cost-benefit analysis and mammographic screening: a travel cost approach. *J Health Econ* 1998;**17**: 767–87.
6. Konishi H, Sandfort MT. Expanding demand through price advertisement. *Int J Indust Organ* 2002;**20**:965–94.
7. Pokorny SB, Jason LA, Schoeny ME. The relation of retail tobacco availability to initiation and continued smoking. *J Clin Child Adolesc Psychol* 2003;**32**:193–204.
8. Novak SP, Reardon SF, Raudenbush SW, Buka SL. Retail tobacco outlet density and youth cigarette smoking: a propensity-modeling approach. *Am J Public Health* 2006;**96**: 670–6.
9. Henriksen L, Feighery EC, Schleicher NC, Cowling DW, Kline RS, Fortmann SP. Is adolescent smoking related to the density and proximity of tobacco outlets and retail cigarette advertising near schools? *Prev Med* 2008;**47**:210–4.
10. Peterson NA, Lowe JB, Reid RJ. Tobacco outlet density, prevalence, and demographics at the county-level of analysis. *Subst Use Misuse* 2005;**40**:1627–35.
11. Reid RJ, Peterson NA, Lowe JB, Hughey J. Tobacco outlet density and smoking prevalence: does racial concentration matter? *Drugs Educ Prev Policy* 2005;**12**:233–8.
12. Yu DL, Peterson NA, Robert RJ. Exploring the impact of non-normality on spatial non-stationarity in geographically weighted regression analyses: tobacco outlet density in New Jersey. *GISci Remote Sens* 2009;**46**:329–46.
13. Forster JL, Wolfson M. Youth access to tobacco: policies and politics. *Ann Rev Public Health* 1998;**19**:203–35.
14. Ashe M, Jernigan D, Kline R, Galaz R. Land use planning and the control of alcohol, tobacco, firearms, and fast food restaurants. *Am J Public Health* 2003;**93**:1404–8.
15. Laws MB, Whitman J, Bowser DM, Krech L. Tobacco availability and point of sale marketing in demographically contrasting districts of Massachusetts. *Tob Control* 2002;**11**:ii71.
16. Hyland A, Travers MJ, Cummings KM, Bauer J, Alford T, Wieczorek WF. Tobacco outlet density and demographics in Erie County, New York. *Am J Public Health* 2003;**93**: 1075–6.
17. Hackbarth DP, Schnopp-Wyatt D, Katz D, Williams J, Silvestri B, Pfleger M. Collaborative research and action to control the geographic placement of outdoor advertising of alcohol and tobacco products in Chicago. *Public Health Rep* 2001;**116**:558–68.
18. Schaap MM, Kunst AE. Monitoring of socio-economic inequalities in smoking: learning from the experiences of recent scientific studies. *Public Health* 2009;**123**:103–9.
19. Anselin L. *Spatial econometrics: methods and models*. Dordrecht: Kluwer Academic Publishers; 1988.

20. Anselin L, Rey S. Properties of tests for spatial dependence in linear regression models. *Geogr Anal* 1991;**23**:112–31.
21. Anselin L. Spatial econometrics. In: Baltagi B, editor. *A companion to theoretical econometrics*. Oxford: Basil Blackwell; 2001. p. 310–30.
22. Anselin L, Bera A. Spatial dependence in linear regression models with an introduction to spatial econometrics. In: Ullah A, Giles D, editors. *Handbook of applied economic statistics*. New York: Marcel Dekker; 1998. p. 237–89.
23. Bailey TC, Gatrell AC. *Interactive spatial data analysis*. Harlow: Longman; 1995.
24. Cliff A, Ord J. *Spatial processes, models and applications*. London: Pion; 1981.
25. Dubin R. Estimation of regression coefficients in the presence of spatially autocorrelated error terms. *Rev Econ Stat* 1988;**70**: 466–74.
26. Griffith D. *Advanced spatial statistics*. Dordrecht: Kluwer Academic Publishers; 1988.
27. Griffith D. Effective geographic sample size in the presence of spatial autocorrelation. *Ann Assoc Am Geograph* 2006;**95**: 740–60.
28. Yu DL, Wei YHD. Spatial data analysis of regional development in Greater Beijing, China, in a GIS environment. *Papers Region Sci* 2008;**87**:97–117.
29. Schneider JE, Reid RJ, Peterson NA, Lowe JB, Hughey J. Tobacco outlet density and demographics at the tract level of analysis in Iowa: implications for environmentally based prevention initiatives. *Prev Sci* 2005;**6**:319–25.
30. Barbeau EM, Wolina KY, Naumova EN, Balbach E. Tobacco advertising in communities: associations with race and class. *Prev Med* 2005;**40**:16–22.
31. Pierce JP, Gilmer TP, Lee L, Gilpin EA, de Beyer J, Messer K. Tobacco industry price-subsidizing promotions may overcome the downward pressure of higher prices on initiation of regular smoking. *Health Econ* 2005;**14**:1061–71.
32. Centers for Disease Control and Prevention. *Best practices for comprehensive tobacco control programs–2007*. Atlanta GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2007.
33. John R, Cheney MK, Azad MR. Point-of-sale marketing of tobacco products: taking advantage of the socially disadvantaged? *J Health Care Poor Underserv* 2009;**20**:489–506.
34. Cohen DA, Ghosh-Dastidar B, Scribner R, Miu A, Scott M, Robinson P, et al. Alcohol outlets, gonorrhea, and the Los Angeles civil unrest: a longitudinal analysis. *Soc Sci Med* 2006; **62**:3062–71.
35. Waller LA, Zhu L, Gotway CA, Gorman DM, Gruenewald PJ. Quantifying geographic variations in associations between alcohol distribution and violence: a comparison of geographically weighted regression and spatially varying coefficient models. *Stoch Environ Res Risk Assess* 2007;**21**:573–88.
36. Fotheringham AS, Brunsdon C, Charlton ME. *Geographically weighted regression: the analysis of spatially varying relationships*. West Sussex: John Wiley & Sons Ltd; 2002.
37. Bivand R. *Spatial dependence: weighting schemes, statistics and models*. Vienna: R Foundation for Statistical Computing. Available at: http://cran.r-project.org/src/contrib/Descriptions/spdep.html; 2010 [accessed 10.02.10].
38. R Development Core Team. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available at: http://www.R-project.org; 2010.
39. Centers for Disease Control and Prevention. Cigarette smoking among adults and trends in smoking cessation. *MMWR Morbid Mortal Wkly Rep* 2008;**58**:1227–32.