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

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.
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.

© 2010 The Royal Society for Public Health. Published by Elsevier Ltd. All rights reserved.

Introduction

Economic theory posits that the net price faced by consumers of products such as tobacco is, in part, a function of search costs.\textsuperscript{1,2} Search costs generally include costs associated with obtaining price and quality information,\textsuperscript{3,4} but also include other costs associated with transacting, such as time and distance travelled by individuals to the point of sale.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.\textsuperscript{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.,\textsuperscript{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\textsuperscript{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.\textsuperscript{10,11} These studies support previous research showing that areas with a higher percentage of minority residents had greater tobacco outlet density.\textsuperscript{16,17} The majority of the studies, however, have rather limited discussion of the potential spatial effects\textsuperscript{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.\textsuperscript{15–28}

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,\textsuperscript{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,\textsuperscript{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\textsuperscript{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 disadvantaged socio-economic status. 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 have established a geographic adjacency approach to setting 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 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.

Two spatial regression analyses are often mentioned in the spatial analysis literature: 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 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

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Analytical results for the spatial lag model and comparison with ordinary least squared (OLS) regression.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
</tr>
<tr>
<td>(Intercept)</td>
<td>0.35</td>
</tr>
<tr>
<td>Percentage of African Americans</td>
<td>0.13</td>
</tr>
<tr>
<td>Percentage of Hispanics</td>
<td>0.80</td>
</tr>
<tr>
<td>Median household income</td>
<td>-4.56E-06</td>
</tr>
</tbody>
</table>

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 < 2.2e-16.
Lagrange Multiplier test for residual autocorrelation test value: 1.5017, P = 0.22041 (test for OLS residual autocorrelation, P < 2.2e-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 is often employed to rectify such problems.
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 ($P < 2.2e-16$, 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. 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**