Predictors of tobacco outlet density nationwide: a geographic analysis

Daniel Rodriguez,[1] Heather A Carlos,[2] Anna M Adachi-Mejia,[2,3] Ethan M Berke,[2,4] James D Sargent[2,3]

[1]LaSalle University, Graduate Clinical Counseling Psychology, Philadelphia, Pennsylvania, USA
[2]Cancer Control Research Program, Norris Cotton Cancer Center, Lebanon, New Hampshire, USA
[3]Pediatrics, Dartmouth Medical School, Lebanon, New Hampshire, USA
[4]Community and Family Medicine, Dartmouth Medical School, Lebanon, New Hampshire, USA

Correspondence to
Dr Daniel Rodriguez, 1900 West Olney Avenue, Philadelphia, PA 19141, USA;
drodrig63@yahoo.com

ABSTRACT

Objective To elucidate how demographics of US Census tracts are related to tobacco outlet density (TOD).

Method The authors conducted a nationwide assessment of the association between socio-demographic US Census indicators and the density of tobacco outlets across all 64 909 census tracts in the continental USA. Retail tobacco outlet addresses were determined through North American Industry Classification System codes, and density per 1000 population was estimated for each census tract. Independent variables included urban/rural; proportion of the population that was black, Hispanic and women with low levels of education; proportion of families living in poverty and median household size.

Results In a multivariate analysis, there was a higher TOD per 1000 population in urban than in rural locations. Furthermore, higher TOD was associated with larger proportions of blacks, Hispanics, women with low levels of education and with smaller household size. Urban—rural differences in the relation between demographics and TOD were found in all socio-demographic categories, with the exception of poverty, but were particularly striking for Hispanics, for whom the relation with TOD was 10 times larger in urban compared with rural census tracts.

Conclusions The findings suggest that tobacco outlets are more concentrated in areas where people with higher risk for negative health outcomes reside. Future studies should examine the relation between TOD and smoking, smoking cessation, as well as disease rates.

Almost 50 years after the 1964 Surgeon General's report implicating smoking as a cause of lung cancer, heart disease and chronic lung disease, smoking remains the leading preventable cause of death and disease in the USA, resulting in some 443 000 deaths annually.[1,2] One contextual factor that may be related to an individual's propensity to smoke is access to tobacco, including the number of and proximity to stores that sell tobacco or 'tobacco outlets'. Tobacco outlet density (TOD) is a community contextual health factor, which, along with air and water quality, the availability of parks and playgrounds for recreation or stores selling nutritious foods, could impact a community's health.[3] Indeed, the results of several recent studies suggest a possible association between greater TOD and an increased likelihood of taking up smoking.[4–6] Although the results of these early studies are mixed,[7,8] and more research is clearly needed to clarify the extent of the relation, these findings have nevertheless prompted some researchers to suggest managing density of

tobacco outlets as an important component of comprehensive tobacco control.[9]

Regional studies have found that TOD is higher in areas with lower median household incomes and those of minority race or ethnicity, particularly Hispanic and black.[5,10–12] Coincidentally, one of the most striking findings in modern surveys of smokers is that smoking prevalence varies greatly across social class, making smoking predominantly a health risk factor today among the poor. For example, in 2007, the overall prevalence of smoking among US adults was 19.7%; however, prevalence varied markedly by education level[13]: 'Adults who had a General Education Development (GED) diploma (44.0%) and those with 9—11 years of education (33.3%) had the highest prevalence of current smoking. Those who had an undergraduate or graduate degree had the lowest smoking prevalence (11.4% and 6.2%, respectively)'. Having lost the affluent segment, the poor segment is a clientele the tobacco industry can ill afford to lose.

If increased TOD in poor areas affects tobacco use there, this could exacerbate or maintain observed disparities in smoking prevalence. Increased TOD could affect tobacco use by, for example, lowering the cost of a tobacco purchase and increasing the propensity to purchase and use tobacco products.[10,14] Higher TOD could also increase the initiation of new smokers by raising exposure to storefront advertising[15,16] and increasing the visibility of individuals purchasing and consuming tobacco, thereby reinforcing a favourable subjective norm for smoking.[17,18] This approach would be an effective marketing strategy as the perceived prevalence of smoking is associated with smoking, particularly in youth,[19–23] the age group at greatest risk for smoking.[24,25] Furthermore, higher retail tobacco density could derail quit attempts by increasing product visibility and encouraging impulse buying,[18,26] and increasing the visibility and impact of promotional discounts for cigarettes.[27] In this scenario, disproportionately high tobacco density in specific regions contributes to increased initiation, promotes higher daily use and limits cessation of tobacco use. This would logically lead to excess disease based on their greater neighbourhood access to tobacco.[12]

One problem with the evidence base supporting the above scenario is that most of the research in this area fails to account for population density. It is plausible that higher TOD in poor neighbourhoods is due to higher population density in poorer or minority neighbourhoods, which would logically support a greater area-based density of tobacco outlets to efficiently distribute product to the

neighbourhood population. Another weakness in the extant research is the regional nature of the studies published to date, which limits generalisability.[5] [10–12] The results of these studies indicate that lower median household income and the proportion of blacks are associated with a higher density of tobacco outlets.[5] [10–12] However, the results are mixed for Hispanics, which could be a result of undercoverage of Hispanics in the regions studied. Finally, few studies have examined whether the relation between socioeconomic status and retail outlet density holds equally well for urban and rural areas. Because tobacco outlets in rural areas have greater geographic reach, they may serve both poor and more affluent tobacco users living there. Thus, we would expect to find less striking disparities in rural compared with urban communities; it is difficult to test this hypothesis in regional studies.

This study adds to the literature by assessing the association between socio-demographic indicators and the density of tobacco outlets across the 64 909 census tracts in the entire continental USA, using a retail outlet density measure that accounts for underlying population density. To assure that findings for poverty, blacks and Hispanics within a census tract were not spurious, we added measures of education and average household size to the model. The national sample allowed us to determine definitively whether the association with socio-demographic variables and TOD was modified by urban versus non-urban status. Finally, this study adds to the existing literature by validating commercially derived densities against densities derived from licensing data in one state, something that has not been done to our knowledge.

METHODS
Instrumentation
Tobacco outlet density
A national data set of tobacco outlets was created using North American Industry Classification System (NAICS) codes. The Office of Management and Budget developed NAICS for use by Federal Statistical Agencies, classifying all business establishments based on their primary activity. We obtained geocoded data from the NAICS Association (http://www.naics.com) for all likely points of sale for tobacco products, including establishments coded as tobacco stores, grocery stores, gas stations and convenience stores, identifying 306 695 addresses nationwide.

We produced a nationwide density surface for the tobacco outlets using adaptive bandwidth kernel density estimation[28] [29] using the LandScan Global Population Data Set, which estimates the underlying population based on satellite imagery.[30] Adaptive bandwidth kernel density estimation allows the influence (the bandwidth) of each tobacco outlet to be limited to a surrounding population, or in our case, 1000 people. In choosing 1000 persons, we estimated that 1000 would be about the number of persons needed to support a small retailer. Our estimation is supported by the data, in that the median TOD in urban areas is close to 1.0 (it is 0.74). This mathematically constrains the influence of a single outlet to a small spatial area in regions where the population density is high (eg, urban areas), while in rural areas, the influence of the tobacco outlet is allowed to be larger geographically. For regions that are sparsely populated, we limited the influence of each outlet to a 25 km radius to prevent the density calculation from expanding to a spatially unreliable distance. We then used this density surface to calculate the mean TOD for each census tract by averaging the densities contained within each polygonal census tract. The resultant density unit, our criterion variable for this study, is

measured in tobacco outlets per 1000 population. Data and maps for states and communities are available from the senior author on request.

Validation of the dependent measure
Several studies have raised concerns about the accuracy of NAICS and other commercial data sets. For example, in a field study[31] of food outlets in rural towns, researchers found that only 39% of outlets were accurately characterised by Google Earth and Yahoo Yellow Pages. Since no one has validated commercial data for tobacco outlets, we assessed the concordance between densities calculated above for Washington State, obtaining addresses for all tobacco retailers (n=6386) from the Washington State Department of Health as of 1 February 2012. We successfully geocoded 97% of the licensees based on the street address (n=6196). Our national data set of tobacco retailers was from 2007 and contained 5623 outlets in Washington State. The difference in the number of outlets can be accounted for even with the 4+ year difference in compilation date, by inclusion of vending machines in the licenses data set and because the commercial data set did not include pharmacies and warehouse stores. Additionally, the commercial data set included all stores of a particular category (eg, grocery or convenience store), some of which may not have sold cigarettes.

The question pertinent to the present study was not the level of one-to-one correspondence for tobacco outlets but the correspondence between census tract TOD. We used the same kernel density procedure to estimate TOD for each of the 1318 census tracts in Washington State with the license data (TOD_L) and compared these with densities from the commercial data set (TOD_C). The question we sought to answer was whether use of commercial sources introduced bias into our study. We first examined the distributions for each variable and found them to be broadly similar (median (IQR) for TOD_L 0.32 (0.13−1.92) and for TOD_C 0.30 (0.13−0.78)), with mean TOD being significantly higher with the license data set. A lowess smoothed plot of TOD_C against TOD_L was linear, and the least squares regression confirmed that TOD_L was about 6% higher than TOD_C (β=1.06, t=61, R^2=0.74). We created a new variable (TOD_L−TOD_C) to represent possible bias and fit a least squares regression with independent variables for urbanicity, per cent families in poverty, per cent Hispanic, per cent black and population of the census tract as independent variables, and it failed to reach statistical significance F (7, 1310)=1.94, p=0.06, despite the large sample size. Finally, we fit the model included in the present manuscript for Washington State using first TOD_L then TOD_C and found that the conclusions for both models were the same. While not national in scope, this validation provides very strong evidence that, although commercial data sets may underestimate the number of tobacco outlets, these underestimates are not biased with respect to the variables of interest in our study.

Socio-demographic variables
Predictor variables were based on US Census level data regarding the following socioeconomic and ethnic group indicators: proportion of black race, proportion of Hispanic ethnicity, proportion of families with income below the poverty level, proportion of women older than 25 years without a high school diploma or equivalent and average household size. These census data were extracted from the 2000 US Census for all census tracts within the continental USA.[32] [33] We selected these socio-demographic variables as they are known to be related to smoking.[34] Furthermore, we assessed household size as it is an indicator of

residential crowding, a health disparity indicator, that is, related to poverty, and has become an increasingly important issue given the recent influx of immigrant families into the USA.[35][36]

Urbanicity

We used the Rural-Urban Commuting Area (RUCA) classification system to derive a binary indicator of urban versus rural census tracts. The RUCA classification system uses commuter patterns to derive a four-tiered classification system of census tracts into Metropolitan (Urban), Micropolitan (large rural town), Small Rural Town and Isolated Rural census tracts. We assessed the socio-demographic variables at each level of the four-tiered RUCA classification system,[37] noting that differences were most pronounced when comparing urban to the non-urban census tracts (ie, Urban versus Micropolitan, Small Rural Town and Isolated Rural). Thus, we determined that a binary urban versus non-urban variable would best represent differences in the relations between our socio-demographic variables and TOD.

Data analysis

We assessed the effects of the socio-demographic variables and the two-tier urban/non-urban binary variable on our outcome variable, log tobacco output density per 1000 population within a census tract, with a multiple regression analysis. Predictor variables were entered using a hierarchical forced-entry method in two blocks. In the first block, the binary urban/non-urban variable was entered with the quantitative socio-demographic variables: average household size and proportion of Hispanic, black, families living in poverty and women older than 25 years without a high school diploma. All model quantitative predictor variables were log transformed and centred about their means to make the relation with the dependent variable more linear and to aid in the interpretation of interaction effects once added. We had tested the relation of each predictor variable with our criterion variable for an R^2 improvement with a quadratic trend, and in no case did the quadratic trend improve substantively upon a simple linear trend. In the second block, the quantitative predictor variables along with the interaction of each quantitative variable with the urban/non-urban binary variable were entered into the model.

The significant interactions were illustrated by plotting TOD on the y-axis and representing predictor variables on the x-axis at 1SD below and above the predictor variable's mean. Slope calculations were based on the multiple regression formulae ($y=b_0+b_1X+b_2Y+b_3XY$), including the specific interaction terms Y (urban=1, non-urban=0). Thus, for instance, with respect to non-urban census tracts (Y=0), the slope for log proportion of Hispanics (X) included the intercept term (b_0) and the gradient (b_1) for Hispanics, but neither the interaction (XY) gradient (b_2) nor the gradient for Y (b_3), as each was multiplied by zero (non-urban, Y=0). To calculate the slope for urban

census tracts, we reverse coded the moderator (0=urban, 1=non-urban) so again our formula only included the intercept term and the gradient for Hispanics. All data analysis was conducted with PASW Statistics V.18.0.

RESULTS

Descriptive statistics

Descriptive statistics are presented in table 1. Medians and IQRs were used as summary statistics as the majority of variables were positively skewed and positively kurtotic. Of the nearly 65 000 census tracts in the continental USA, approximately 45 000 are categorised as urban within the two-tiered RUCA classification system. The median density of tobacco outlets per 1000 population in a census tract was higher for urban (median=0.74, IQR=1.46) than for non-urban (median=0.22, IQR=0.29) census tracts. All comparisons between urban and non-urban census tracts were statistically significant, p<0.0001, even when differences were small (eg, average household size). However, these significant relations are a likely result of the large sample size.

Multivariate model

The results of the multiple regression analysis are presented in table 2. In the first block (Model 1), all main effects were significant (p<0.0001). Of the six predictor variables in Model 1, five had positive effects on the outcome variable, and only average household size had a negative association. Compared with non-urban census tracts, urban census tracts were associated with a 32% increase in TOD per 1000 population. Among the continuous variables, proportion of Hispanic and proportion of families living in poverty had the largest association with TOD: a 1% increase in each measure was associated with a 0.91% and 0.83% increase in density, respectively. These findings are consistent with an argument that socio-demographic variables are risk factors for greater TOD per 1000 population. Unexpectedly, there was an inverse relation between average household size and TOD per 1000 population. Specifically, each 1% change in the average household size was associated with a 0.57% decline in TOD per 1000 population.

To explore whether the main effects were similar across urban and rural census tracts, we entered the five interaction terms in the second block (Model 2). Only the urban by poverty interaction was not significant (p=0.62), suggesting that there is no difference in the association between the proportion of families living in poverty within a census tract and TOD between the urban and non-urban census tracts. TOD per 1000 population was higher for urban than for non-urban census tracts, regardless of the level of the predictor variable (figure 1). The most striking differences were seen when comparing proportion of Hispanics in urban versus non-urban census tracts, with the slope for the effect of the proportion of Hispanics being much

Table 1 Descriptive statistics

Variable	Total (n=64 909)		Urban (n=44 635)		Rural (n=20 228)	
	Median	IQR	Median	IQR	Median	IQR
Tobacco outlet density per 1000 population	0.43	1.08	0.74	1.46	0.22	0.29
Proportion of black	0.03	0.13	0.04	0.16	0.01	0.05
Proportion of Hispanic	0.03	0.10	0.05	0.14	0.01	0.03
Proportion of families living in poverty	0.07	0.11	0.06	0.12	0.08	0.08
Proportion of women >25 without a HS diploma	0.17	0.18	0.16	0.19	0.19	0.14
Average household size (people)	2.58	0.48	2.59	0.59	2.56	0.31

These are raw values, not log transformed.
HS, high school.

Table 2 Results of the hierarchical multiple regression analysis

Predictor variables	Estimate	SE	β*	t Value	p Value
Model 1					
Constant	0.31	0.003	—	114.28	<0.0001
Log proportion of Hispanic	0.91	0.01	0.29	68.52	<0.0001
Log proportion of black	0.43	0.01	0.17	41.74	<0.0001
Log proportion of families living in poverty	0.83	0.02	0.16	34.29	<0.0001
Log proportion of women older than 25 years without a high school diploma	0.34	0.02	0.09	17.33	<0.0001
Log average household size	−0.57	0.01	−0.18	−54.02	<0.0001
Urban (1=urban, 0=rural)	0.32	0.003	0.33	93.90	<0.0001
Model 2					
Constant	0.26	0.003	—	81.38	<0.0001
Log proportion of Hispanic	0.11	0.03	0.03	4.12	<0.0001
Log proportion of black	0.10	0.02	0.04	4.13	<0.0001
Log proportion of families living in poverty	0.75	0.05	0.15	15.11	<0.0001
Log proportion of women older than 25 years without a high school diploma	0.05	0.04	0.01	1.33	0.183
Log average household size	−0.26	0.02	−0.08	−12.53	<0.0001
Urban (1=urban, 0=rural)	0.36	0.004	0.37	97.97	<0.0001
Urban by log proportion of Hispanic	1.01	0.03	0.29	33.16	<0.0001
Urban by log proportion of black	0.43	0.03	0.15	16.65	<0.0001
Urban by log proportion of families living in poverty	0.03	0.06	0.005	0.50	0.617
Urban by log proportion of women older than 25 years without a high school diploma	0.33	0.04	0.08	7.73	<0.0001
Urban by log average household size	−0.44	0.02	−0.13	−18.35	<0.0001

$R^2=0.39$ for Step 1, $\Delta R^2=0.02$ ($p<0.0001$).
The column header Estimate is the unstandardised parameter estimate. For all predictor variables except the binary urban/non-urban variable, the estimate represents the per cent change in the tobacco outlet density per 1000 population for a 1% change in the predictor variable since both the predictor and the outcome variables are log transformed. In the case of the binary urban/non-urban predictor variable, which is not log transformed, the estimate is the ($100 \times$ estimate) per cent change in the outcome variable for a unit change in the predictor variable. For the interaction terms, the interpretation is the additional percentage increase for urban census tracts versus rural census tracts. Thus, for instance, a 1% increase in the proportion of Hispanics in non-urban census tracts is associated with a 0.11% increase in the tobacco outlet density per 1000 population. For urban communities, there is an additional 1.01% change totalling to a 1.12% increase for a 1% increase in the proportion of Hispanics compared with a 0.11% increase in non-urban census tracts.
*Standardized parameter estimate.

steeper in the urban (slope=1.12) than in the non-urban (slope=0.11) settings. For proportion of blacks, there was a different pattern, with a significantly steeper slope in the non-urban (slope=0.96) than in the urban (slope=0.53) census tracts. The association between proportion of women older than 25 years without a high school diploma and TOD per 1000 population was no longer significant once its interaction term was included in the model, suggesting that its effect on TOD was driven by urban as opposed to non-urban census tracts. Finally, for household size, as the average household size increased, the density of tobacco outlets decreased, but the decrease was steeper in the urban (slope=−0.70) than in non-urban (slope=−0.26) census tracts.

DISCUSSION

This study clearly indicates, in a national sample, that race, ethnic and socioeconomic factors are positively associated with TOD. Furthermore, the study demonstrates that these associations are not simply a function of higher population density for the tracts where poor and minority populations tend to live. These data do not permit us to understand why such disparities exist, in part, because we cannot ascertain whether higher TOD is a specific industry targeting strategy,[12] contributing to more tobacco use among those living in poverty or whether individuals living in poverty require higher TOD to satisfy their higher demand for tobacco products. However, the study points out the need for longitudinal research studies to understand what factors dictate tobacco points of sale in low-income communities,[38] and whether such points of sale affect smoking among the individuals that live there.

One novel finding in this study is the striking difference in how TOD is related to socio-demographic variables in urban versus rural settings and race/ethnicity. Interestingly, poverty

seems to confer risk for higher retail TOD regardless of whether the community is urban or rural. In contrast, the relation of tobacco outlets to the proportion of Hispanics living in a census tract is much more pronounced in urban than in rural areas, whereas the opposite is seen when comparing urban and rural areas for the proportion of black population. For Hispanics, this difference may be the result of their increased population growth rate and a greater proportion of tobacco outlets being distributed to Hispanic urban centres to meet the demand of new smokers; from 2000 to 2006, Hispanics accounted for 50% of the growth in the US population.[39] Although the rates of smoking are lower among Hispanics than black and white adults, Hispanic adolescents smoke at twice the rate of black youth (18% vs 9%).[34 40] Thus, it is possible that tobacco companies target larger Hispanic urban communities as a source of potential new smokers. We see the opposite pattern for the proportion of blacks. Although there is a positive relation between the proportion of blacks in a census tract and the density of tobacco outlets in both urban and rural census tracts, the positive relation is more pronounced in rural than in urban census tracts. This finding may be the result of certain rural areas with disproportionately high rates of smoking (eg, the Black Belt counties of rural Alabama) that naturally result in a greater density of tobacco outlets to meet the increased need for cigarettes.[41] For instance, the national average of regular smoking for black men was 24% across the USA, whereas the proportion of weekly smokers was 40% in the Black Belt sample. Although this is but one rural region of the USA, and the proportion of black regular smokers in the Black Belt region of Alabama is still lower than white smokers,[42] it is plausible that larger percentages of smokers in various rural communities may affect the relation of race to density across rural regions. Therefore, such diverse effects suggest a need for more detailed

Figure 1 Urban versus rural differences in the effects of the socio-demographic variables on tobacco outlet.
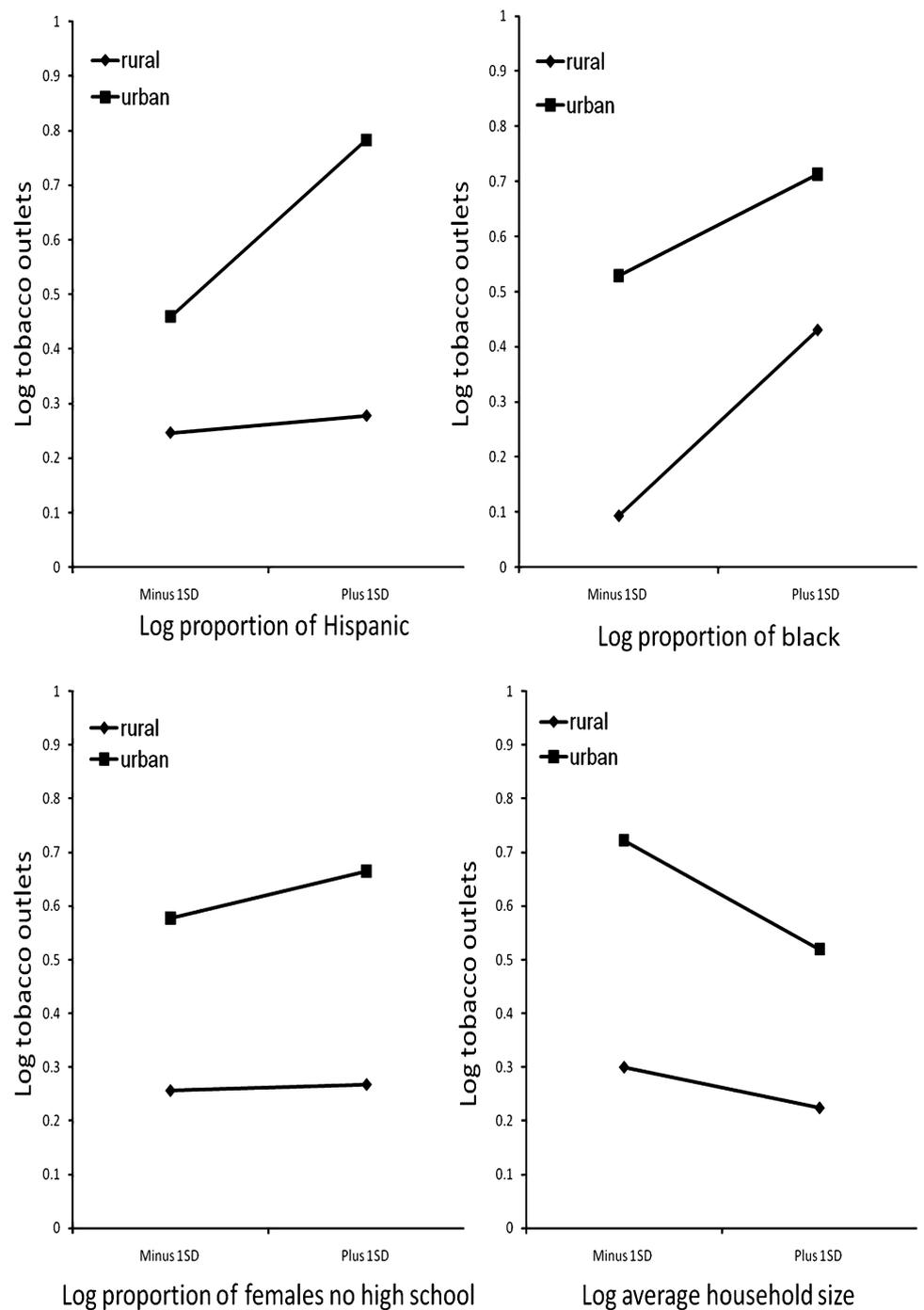
study of the role of tobacco-related health disparities within urban and rural settings.

Another interesting finding is the negative relation between average household size and tobacco outlets in urban and non-urban census tracts. Although historically, household size has been considered an indicator of overcrowding and lower socio-economic status,[43] our findings suggest that in the case of tobacco outlets, these assumptions may not hold. Perhaps the increase in mother-only families along with the sharp decline in dual-parent families with children younger than 18 years over the past 4 decades can explain this finding, suggesting that smaller households would be related to less income.[44] The disproportionately greater proportion of black and to some extent Hispanic children living in mother-only families[44] bolsters this possibility as greater proportions of both race/ethnic groups

are associated with an increased TOD. Finally, the increased immigration of Hispanics has resulted in a shift in the cultural acceptance of larger households, as there is a greater tendency for recent immigrant Hispanics to reside in a single household to reduce expenditures so as to support families abroad.[43] This would mean that even as income increases, there is a reduced likelihood of changing living arrangements.

There are some limitations to these findings. We did not include all possible tobacco outlets due to practical considerations regarding use of NAICS codes to define likely points of tobacco sale. For instance, some department stores and pharmacies (eg, Wal-Mart, Costco, CVS) may be tobacco outlets.[45] However, many stores in these classifications do not sell tobacco (eg, Sears, Dollar Stores, hospital pharmacies), and this is why we excluded them. Additionally, available data sets on retailers

Research paper

What this paper adds

▶ While research supports that TOD may be related to health disparity variables (eg, race, ethnicity, poverty), studies thus far have used regional data potentially limiting the generality of their findings.
▶ This study extends the prior research by assessing the effects of health disparity variables on TOD nationally using all 64 909 census tracts in the continental USA and generates many important questions for further study.
▶ This study also adds to the existing literature by validating commercially derived densities against densities derived from licensing data in one state, something that has not been done to our knowledge.

have their own limitations, in that their sensitivity to the actual retail environment is only on the order of 60%–70%.[46] However, the validation procedure conducted in Washington State reassures that use of commercial data does not bias the results for the relationships we studied. Another limitation to this study is that we did not alter the bandwidth by store category, so the model assumes that a small convenience store selling tobacco serves the same population size (1000 persons) as a large supermarket. It is also possible that the relations found in this study are specific only to census tracts and that conclusions would differ with alternative areas (eg, cities or counties). Finally, this study cannot distinguish whether tobacco companies target poor communities or simply respond to higher demand there. This is the biggest limitation to the research on this topic and underlines the need for longitudinal research to assess how changes in community structure, demographics and smoking relate to changes in the retail environment (or vice versa). Additionally, there is a need for more research on whether retail tobacco density and proximity affects smoking onset, maintenance and cessation.

The results of this study clearly demonstrate that access to tobacco is most pronounced in areas highly populated with individuals already vulnerable to negative health outcomes based on socio-demographic indicators.[47] These national findings raise concern that tobacco points of sale could promote or exacerbate disparities in smoking prevalence and its associated morbidity and mortality. However, further study is needed to determine if limits on retail outlet density should be incorporated into tobacco control policy.

REFERENCES

1. **Adhikari B,** Kahende J, Malarcher A, et al. Smoking-attributable mortality, years of potential life lost, and productivity losses—United States, 200–2004. MMWR Morb Mortal Wkly Rep 2008;**57**:1226–8.
2. The health consequences of smoking: a report of the Surgeon General (Atlanta, GA). Washington, DC: 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, 2004.
3. **Macintyre S,** Ellayway A, Cummins S. Place effects on health: how can we concetualise operationalise and measure them? Soc Sci Med 2002;**55**:125–39.
4. **Novak SP,** Reardon SF, Raudenbush SW, et al. Retail tobacco outlet density and youth cigarette smoking: a propensity-modeling approach. Am J Public Health 2006;**96**:670–6.
5. **Peterson NA,** Lowe JB, Reid RJ. Tobacco outlet density, cigarette smoking prevalence, and demographics at the county level of analysis. Subst Use Misuse 2005;**40**:1627–35.
6. **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**:210–14.
7. **Chuang Y,** Cubbin C, Ahn D, et al. Effects of neighbourhood socioeconomic status and convenience store concentration on individual level smoking. J Epidemiol Community Health 2005;**59**:568–73.
8. **Pearce J,** Hiscock R, Moon G, et al. The neighbourhood effects of geographical access to tobacco retailers on individual smoking behaviour. J Epidemiol Community Health 2009;**63**:69–77.
9. **Cohen JE,** Anglin L. Outlet density: a new frontier for tobacco control. Addiction 2009;**104**:2–3.
10. **Schneider JE,** Reid RJ, Peterson NA, et al. Tobacco outlet density and demographics at the tract level of analysis in Iowa: Implications ofr environmentally based prevention initiatives. Prev science 2005;**6**:319–25.
11. **Hyland A,** Travers MJ, Cummings KM, et al. Tobacco outlet density and demographics in Erie County, New York. Am J Public Health 2003;**93**:1075–6.
12. **Yu D,** Peterson NA, Sheffer MA, et al. Tobacco outlet density and demographics: analysing the relationships with a spatial regression approach. Public Health 2010;**124**:412–16.
13. **Thorne SL,** Malarcher A, Maurice E, et al. Cigarette smoking among adults - - United States, 2007. MMWR Morb Mortal Wkly Rep 2008;**57**:1221–6.
14. **Siahpush M,** Jones PR, Singh GK, et al. Association of availability of tobacco products with socio-economic and ratial/ethnic characteristics of neighbourhoods. Public Health 2010;**124**:525–9.
15. **Henriksen L,** Flora JA, Feighery E, et al. Effects on youth of exposure to retail tobacco advertising. J Appl Soc Psychol 2002;**32**:1771–89.
16. **Henriksen L,** Feighery EC, Schleicher NC, et al. Reaching youth at the point of sale: cigarette marketing is more prevalent in stores where adolescents shop frequently. Tob Control 2004;**13**:315–18.
17. **West JH,** Blumberg EJ, Kelley NJ, et al. Does proximity to retailers influence alcohol and tobacco use among Latino adolescents? J Immigrant Minor Health 2010;**12**:626–33.
18. **Paul CL,** Mee KJ, Judd TM, et al. Anywhere, anytime: retail access to tobacco in New South Wales and its potential impact on consumption and quitting. Soc Sci Med 2010;**71**:799–806.
19. **Wang MP,** Ho SY, Lo WS, et al. Overestimation of peer smoking prevalence Predicts smoking initiation among primary school students in Hong Kong. J Adolesc Health 2011;**48**:418–20.
20. **Franca LR,** Dautenberg B, Falissard B, et al. Peer substance use overestimation among French university students: a cross-sectional survey. Bmc Public Health 2010;**10**:169.
21. **Lai MK,** Ho SY, Lam TH. Perceived peer smoking prevalence and its association with smoking behaviours and intentions in Hong Kong Chinese adolescents. Addiction 2004;**99**:1195–205.
22. **Henry DB,** Kobus K, Schoeny ME. Accuracy and bias in adolescents' Perceptions of Friends' substance use. Psychol Addict Behav 2011;**25**:80–9.
23. **Reid JL,** Manske SR, Leatherdale ST. Factors related to adolescents' estimation of peer smoking prevalence. Health Educ Res 2008;**23**:81–93.
24. **CDC.** Smoking attributable mortality, years of potential life lost, and productivity losses. United States, 2000-2004. MMWR Morb Mortal Wkly Rep 2008;**57**:1226–8.
25. **Johnston LD,** O'Malley PM, Bachman JG, et al. Monitoring the Future National Survey Results on Drug Use, 1975-2007. **Vol. I**. Secondary school students. Bethesda, MD: NIH Publication No. 08-6418A, 2008.
26. **Reitzel LR,** Cromley EK, Li YS, et al. The effect of tobacco outlet density and proximity on smoking cessation. Am J Public health 2011;**101**:315–20.
27. **Ruel E,** Mani N, Sandoval A, et al. After the master settlement agreement: Trensds in the American tobacco retail environment from 1999 to 2002. Health Promot Pract 2004;**5**:99s–110s.
28. **Carlos HA,** Shi X, Sargent J, et al. Density estimation and adaptive bandwidths: a primer for public health practitioners. Int J Health Geogr 2010;**9**:8.
29. **Shi X.** Selection of bandwidth type and adjustment side in kernal density estimation over inhomogeneious backgrounds. Int J Geographical INformation Sci 2010;**24**:643–60.
30. **Laboratory ORN.** LandScan Global Population Database. Oak Ridge, TN: Oak Ridge National Laboratory, 2008.
31. **Longacre MR,** Primack BA, Owens PM, et al. Public Directory data sources do not accurately Characterize the food environment in two predominantly rural states. J Am Diet Assoc Apr 2011;**111**:577–82.
32. **Bureau UC.** Census 2000 Summary File 1. Washington, DC: US Census Bureau, 2002.

33. **Bureau UC.** *Census 2000 Summary File 3*. Washington, DC: US Census Bureau, 2002.
34. **CDC.** Vital signs: Current cigarette smoking among adults aged >= 18 years—United States, 2009. *MMWR Morb Mortal Wkly Rep* 2010;**59**:1135—40.
35. **Litt JS,** Goss C, Diao LH, *et al*. Housing environments and Child health Conditions among recent Mexican immigrant families: a population-based study. *J Immigr Minor Health* 2010;**12**:617—25.
36. **Evans GW,** Kantrowitz E. Socioeconomic status and health: the potential role of environmental risk exposure. *Annu Rev Public Health* 2002;**23**:303—31.
37. **Health WSDo.** *Guidelines for Using Rural-Urban Classification Systems For Public Health Assessment*. 2008. http://www.doh.wa.gov/data/guidelines/RuralUrban2.htm (accessed 31 May 2011).
38. **DeNavas-Walt C,** Proctor BD, Smith JC. Income, poverty, and health insurance coverage in the United States: 2008. *Current population reports: Consumer income* US Census Bureau, 2009.
39. **Ethnicity and Ancestry Branch, Population Division, US Census Bureau.** *Hispanics in the United States*. Suitland-Silver Hill, MD: US Census Bureau, 2006.
40. **CDC.** Cigarette us among high school students - United States, 1991-2009. *MMWR Morb Mortal Wkly Rep* 2010;**59**:797—801.
41. **Carroll WR,** Foushee HR, Hardy CM, *et al*. Tobacco use among rural African American Young adult males. *Otolaryngol Head Neck Surg* 2011;**145**:259—63.
42. **Shuaib F,** Foushee HR, Ehiri J, *et al*. Smoking, Sociodemographic Determinants, and Stress in the Alabama black Belt. *J Rural Health* 2011;**27**:50—9.
43. **Myers D,** Woo Lee S. Immigration cohorts and residential overcrowding in Southern California. *Demography* 1996;**33**:51—65.
44. **Kreider RM,** Elliott DB. *America's Families and Living Arrangements: 2007*. Suitland-Silver Hill, MD: US Census Bureau, 2009.
45. **Morton CM,** Peterson NA, Schneider JE, *et al*. Tobacco sales in community pharmacies: Remote decisions and demographic targets. *J Community Psychol* 2010;**38**:39—48.
46. **Liese AD,** Colabianchi N, Lamichhane AP, *et al*. Validation of 3 food outlet Databases: Completeness and Geospatial accuracy in rural and urban food Environments. *Am J Epidemiol* 2010;**172**:1324—33.
47. **Liao Y,** Bang D, Cosgrove S, *et al*. Surveillance of health status in minority communities—racial and ethnic approaches to community heath across the U.S. (REACH U.S.) Risk Factor Survey, United States, 2009. *MMWR Surveill Summ* 2011;**60**:1—41.