Smoking Prevalence in Urban and Rural Populations: Findings from California between 2001 and 2012

Running title: Smoking in Urban and Rural Californian Populations

Number of words: 4221
Number of Figures: 2
Number of Tables: 3
Abstract

Background: Tobacco smoking and related health problems are still major public health concerns in the United States despite the declining smoking prevalence.

Objectives: This study explored differences in smoking prevalence between urban and rural areas potentially relevant to tobacco control efforts in California.

Methods: Public use adult smoking data from the California Health Interview Survey (CHIS) between 2001 and 2011-2012 were analyzed. A total of 282931 adults were surveyed across the six CHIS cycles. A ZIP code-based geographic classification (Urban, Second-City, Suburban, and Town/Rural) was used to examine the association between smoking prevalence and area of residency.

Results: The overall smoking prevalence in California decreased from 17.0% in 2001 to 13.8% in 2011-2012. Within each CHIS cycle, the Town/Rural areas had the highest smoking prevalence, followed by Urban and Second-City areas, and Suburban areas had the lowest. Pooled data from all CHIS cycles showed a similar pattern, with rates in Urban, Second-City, Suburban and Town/Rural areas being 15.2%, 15.2%, 13.1% and 17.3%, respectively. Weighted multivariate logistic regression analysis indicated significantly higher odds of smoking in Urban, Second-City and Town/Rural areas compared to Suburban areas (all adjusted odds ratios >1.10), although this trend varied by race/ethnicity, being present in non-Hispanic Whites and not present in Hispanics.

Conclusions: Town/Rural and Urban populations of California are consistently at higher risk of smoking than Suburban populations. These results indicate a need for population-specific tobacco control approaches that address the lifestyle, behavior, and education of disparate populations within the same state or region.

Key Words: smoking prevalence, rural, urban, California
INTRODUCTION

Although smoking prevalence rates in the United States (US) dropped from about 43% in 1965 to 18% in 2014, smoking-related diseases are still the leading preventable cause of morbidity and mortality. More than 42 million individuals still smoke today in the US, and nearly one-half million adults will die prematurely each year because of smoking. The total estimated annual economic costs due to tobacco in the US are over 289 billion dollars (1). Therefore, efforts to further reduce tobacco smoking are still crucial for the health of the general population.

A few studies among adults in the US have shown that smoking prevalence may vary between individuals residing in urban areas and those in rural areas. Data from the US Behavior Risk Factor Surveillance System (BRFSS) indicate that smoking prevalence tends to be higher among adults aged 18 years and older from rural areas than similarly aged adults in either urban areas (2) or suburban areas (3), but this finding is not consistent across different states (4, 5). Different urban/rural classifications have also been employed in these studies, such as metropolitan statistical area (MSA) based classification of urban and rural (2), classification into urban, suburban and rural areas (3), urban, suburban and nonmetropolitan areas (4), remote, rural and metro areas (6), and urban, hub (micropolitan), and rural areas (5). Different classifications of geographic areas and regional influences have made comparisons difficult and motivate efforts to better understand trends in tobacco use in rural, urban, and suburban populations across time.

Data from other countries also indicate inconsistent findings regarding rural and urban smoking. A survey of German residents aged 10 years or older showed that small city and metropolitan residents exhibited higher odds of being current or former smokers than those living in rural communities (7). Data from Russia among adults (≥18 years old) showed that both men and women living in urban areas smoked more often than those living in rural areas (8). A study of Greek adults (≥20 years old), however, showed no significant difference between urban and rural adult smoking (9). Although heterogeneous survey populations and disparate social
and environmental situations within different countries contributed to this variation of findings, differences in smoking prevalence between urban and rural areas are evident worldwide.

Finer resolution of the geographic distribution of smoking prevalence will help to better target future tobacco control efforts. The California tobacco control program, which decreased smoking prevalence across all sectors of society (10), has been one of the most successful nationally and globally. However, it is not known if this success is also evident according to rural or urban residency, or whether there is a difference in rural versus urban smoking prevalence independent of common smoking risk factors, such as age, sex, ethnicity, socioeconomic status, etc. The California Health Interview Survey (CHIS) is a population-based telephone survey of California conducted every other year since 2001. It is the largest state health survey ever conducted and one of the largest health surveys in the nation. CHIS collects extensive information on health-related conditions and behaviors, including smoking behavior (11, 12). Utilizing the CHIS public use data, we explored smoking prevalence in California from 2001 to 2012 using a finer 4-level geographic classification derived from ZIP codes of respondents' place of residence (13). Our aim was to examine smoking prevalence by geographic classification, but independent of the common smoking risk factors.

METHODS

Data

Data for this study were from the CHIS 2001, 2003, 2005, 2007, 2009 and 2011-2012 cycles, six cycles for which a total of 282931 adults were surveyed (14). The CHIS is administered biennially by the University of California, Los Angeles (UCLA) Center for Health Policy Research (UCLA-CHPR). To capture the rich diversity of the California population, CHIS is conducted in five languages: English, Spanish, Chinese (Mandarin and Cantonese dialects), Vietnamese, and Korean. The CHIS began in 2001 as a biennial survey with a 12 month survey
window and was modified in 2011 to a continuous sampling protocol. Information is collected for all age strata; data reported in this study were from adults (≥18 years old) only.

The CHIS employs a dual-frame, multi-stage sample design. The random-digit-dial (RDD) sample includes both landline and cellular telephone numbers (cell phone data were officially collected since 2007). To produce population estimates from the CHIS data, weights are applied to the sample data to compensate for the probability of selection and a variety of other factors. For each survey, one base weight is created and used for point estimation; 80 post-stratification replicate weights are created and used for jackknife confidence interval estimation. For details about the CHIS sampling procedures and weighting methodology, please refer to the online CHIS reports (12).

The overall response rate for the CHIS decreased from 37.7% in 2001 to 15.0% for landline and 17.8% for cellular phones in 2011-2012. Survey response rates tend to be lower in California than nationally, and over the past decade response rates have been declining both nationally and in California. However, the CHIS response rates are historically comparable to response rates of other scientific telephone surveys in California, such as the California BRFSS (12).

Measures

In all cycles of the CHIS, two questions about cigarette smoking were asked: “Altogether, have you smoked at least 100 or more cigarettes in your entire lifetime (Yes, No)?” “Do you now smoke cigarettes every day, some days, or not at all (Every day, Some Days, Not at all)?” Those answering “Yes” to the first question and “Every day” or “Some days” to the second question were considered as current smokers and included in the smoking prevalence rate calculation.

Multiple classifications of urban/rural areas are available in the CHIS data and most are binary, except a 4-level geographic classification from Claritas Inc., which was used in this study.
as it provided finer geographic classification for the purpose of the study. The Claritas urbanization classification has been developed and refined by Claritas Inc. since the 1980s because the US Census Bureau does not provide adequate standard measures (13, 15-17). The goal of this approach is to establish an objective classification that is less boundary-dependent. It is based on the population density of a specific geography in the context of its surrounding area. This urban-rural dimension has been used in multiple national surveys and studies (13, 15-17). The Claritas 4-level geographic classification assigns ZIP codes in California to 4 urbanization categories (Urban, Second-City, Suburban, and Town/Rural) based on analysis of population density grids of regularly updated geo-boundaries, redistricting updates, and population estimates. The ZIP code for each case was then assigned to its corresponding urbanization category. In cases that reported ZIP code of a PO Box location rather than a ZIP code for a residence, Claritas Inc. provided the “parent ZIP codes” for these PO Box locations. The urbanization categories assigned to the “parent” ZIP codes were used to classify these cases. Those cases with no ZIP code information were imputed to assign a value for the urbanization variable. Definitions of the 4-level urbanization categories are (provided by UCLA-CHPR):

- **Urban**: ZIP codes associated with dense neighborhoods that represent the central cities of most major metropolitan areas (more than 4,150 persons per square mile (PPSM)).
- **Second City**: ZIP codes associated with moderate-density neighborhoods in population centers (more than 1,000 and fewer than 4,150 PPSM).
- **Suburban**: ZIP codes associated with moderate-density neighborhoods that are not surrounded by urban or second-city population centers (estimated to be more than 1,000 PPSM and not in an urban or second city population center).
- **Town or Rural**: ZIP codes associated with isolated small towns or less-developed areas on the exurban frontier (estimated to be more than 210 but fewer than 1000 PPSM); or
small villages and rural hamlets surrounded by productive farmland or wide-open spaces (estimated to be 210 or fewer PPSM).

**Statistical Analysis**

In order to obtain appropriate population estimates and compensate for the probability of selection and a variety of other factors, all descriptive summaries and statistical models reported were population weighted. Standard errors (SE) for confidence interval calculations and hypothesis testing were calculated by the paired unit jackknife method (JK2) using one base weight and 80 replicate weights as mentioned above (12).

Descriptive statistics for demographic characteristics and geographic locations were reported, with weighted percentages and 95% confidence intervals (95%CI) calculated as 1.96xSE. Smoking prevalence rates were evaluated overall and across demographic and geographic population density strata in each CHIS cycle as well as in the pooled data from all cycles. Rao-Scott chi-square tests (18) were performed to examine the differences in smoking prevalence rates across demographic and geographic strata. Smoking prevalence within each ethnic stratum across the geographic locations and related chi-square tests were also evaluated in each cycle as well as in the pooled data.

In order to examine possible predictors of cigarette smoking, weighted univariate logistic regression analyses were conducted in the pooled data, with current cigarette smoking (yes/no) as the dependent variable and each of the key demographic characteristics, the four geographic areas or the CHIS cycles as the independent variable. Weighted multivariate logistic regression analyses were also conducted using the same dependent variable, and all independent variables from the univariate analyses. In order to identify effect modification of the association between smoking and the geographic areas by demographic characteristics, a set of stratified weighted multivariate logistic analyses were performed within non-Hispanic Whites and Hispanics, respectively, to compare the odds of being a smoker across geographic areas within
each of the ethnic stratum, while controlling for other demographic characteristics. Adjusted odds ratios (AOR) and 95%CI were reported for these weighted logistic regression analyses. All analyses were performed in SAS version 9.3 (SAS Institute Inc., 2011).

RESULTS

Smoking prevalence across demographic strata

Population size and weighted smoking prevalence estimates across strata of sex, age, ethnicity, education, working status, and geographic areas are shown in Table 1. The overall smoking rates decreased from 17.0% in 2001 to 13.6% in 2009, and remained relatively stable at 13.8% in 2011-2012.

Smoking rates were significantly different across ethnic strata in most CHIS cycles (all P’s<0.001) except in 2009 (P=0.14). Non-Hispanic Whites had the highest smoking rates and Hispanics had the lowest in all cycles.

There were also significant differences across strata of sex, age, education, and employment status in all CHIS cycles (all P’s<0.0001). In each cycle, males had higher smoking rates than females; those between 25-44 years old had the highest smoking rates and those ≥65 years old had the lowest; those with high school education had the highest smoking rates and those with college or above education had the lowest; those unemployed but looking for work had higher smoking rates than those with full time jobs, those with part time jobs and those unemployed but not looking for work.

Smoking Prevalence across geographic areas

As shown in Figure 1, there were significant differences in adult population smoking prevalence across geographic areas in each CHIS cycle (all P’s<0.02); those living in Town/Rural areas had the highest smoking rates, those living in Suburban areas had the lowest, and those living in Urban or Second-City areas had smoking rates in between these two types.
of areas. The smoking rates therefore showed a “U-shaped” pattern where Town/Rural and Urban populations had higher smoking prevalence than Suburban populations.

Pooled smoking prevalence from all CHIS cycles across strata of demographic characteristics within each geographic area was also evaluated (Table 2). As expected, the differences across the four areas were significant (P<0.0001). Those living in Suburban areas had the lowest rates (13.1%, 95%CI: 12.6%-13.6%), those living in Town/Rural areas had the highest (17.3%, 95%CI: 16.7%-17.8%), and those living in Urban and Second-City areas had rates in between (15.2%, 95%CI: 14.7%-15.6%, and 15.2%, 95%CI: 14.7%-15.7%, respectively). The smoking rates across strata of each demographic characteristic were significantly different within each geographic area (all P’s <0.0001, except that across ethnic strata in Suburban areas where P=0.049), and showing the same pattern as described above in each CHIS cycle.

Additional analyses of the pooled data comparing education levels across geographic areas showed that among those living in Suburban areas, 40.1% (95%CI: 39.6%-40.7%) had at least college education, this proportion was significantly higher than that of all other education levels in the same areas (<high school 11.0%, 95%CI: 10.5-11.4; high school 23.0, 95%CI: 22.5-23.5; some college 25.9%, 95%CI: 25.4-26.4), and also significantly higher than the same proportion in Urban (32.4%, 95%CI: 32.0%-32.8%), Second-City (29.1%, 95%CI:28.6%-29.5%), and Rural (24.9%, 95%CI:24.2%-25.5%) areas. Analyses within those living in Town/Rural areas showed that 68.3% (95%CI: 67.7%-68.9%) of the rural population were non-Hispanic Whites, and the smoking rates in non-Hispanic Whites (12.6%, 95%CI: 12.1%-13.1) were significantly higher than that in all other ethnicities (Hispanic: 2.4%, 96%CI: 2.1%-2.6%; Others: 2.3%, 95%CI: 2.1%-2.6%; (P<0.0001).

**Smoking prevalence within ethnic strata**
Since non-Hispanic Whites and Hispanics are the two largest ethnic strata in California, comparisons of smoking rates across the geographic areas within each of the two ethnic strata are shown in Figure 2.

For non-Hispanic Whites, smoking rates across the four geographic areas were significantly different in each CHIS cycle (all P’s<0.011), with the same pattern as the entire population such that Urban and Town/Rural or Second-City areas had higher smoking rates than Suburban areas. For Hispanics, smoking rates across the four areas differed significantly only for the 2003 CHIS (P=0.0066), whereby people living in Urban and Rural areas had higher smoking rates.

**Predictors of smoking**

Using the pooled data from all CHIS cycles, weighted univariate logistic regression analysis showed that cigarette smoking was significantly associated with living in Town/Rural, Urban and Second-City areas, being male, being <65 years old, being non-Hispanic, lacking a college education, and being unemployed but looking for work or working full time (data not shown).

As shown in Table 3, after controlling for demographic characteristics and survey cycles, weighted multivariate logistic regression analysis indicated that living in Urban (AOR=1.21, 95%CI: 1.13-1.28), Second-City (AOR=1.11, 95%CI: 1.04-1.18) and Town/Rural (AOR=1.21, 95%CI: 1.13-1.29) areas was associated with significant higher odds of being a smoker compared to living in Suburban areas. Stratified analyses showed that the U-shaped trend in smoking prevalence by population density was present in non-Hispanic Whites (Urban AOR=1.24 (95%CI: 1.16-1.34), Second-City AOR=1.11 (95%CI: 1.03-1.20) and Town/Rural AOR=1.23 (95%CI: 1.14-1.33) versus Suburban), but not in Hispanics.

**DISCUSSION**
During the 12 years of CHIS data collection there was a general decline in overall smoking prevalence in California. Yet, after controlling for other demographic confounders, non-Hispanic Whites living in Urban, Second-City and Town/Rural areas had significantly higher odds of being a smoker than those living in Suburban areas. In contrast, smoking prevalence for Hispanics did not vary by urban versus rural areas.

As shown in Figure 1, a “U-shaped” pattern for smoking prevalence by geographical distribution was observed across all cycles of the CHIS, with smoking higher in Urban, Town/Rural and Second-City areas than in Suburban areas. Consistent with our findings, a similar “U-shaped” association between geographic areas (population density) and smoking prevalence had been described for the California Tobacco Surveys (CTS) representative of California (10). In that study, those living in less than 1000 Person Per Square Mile (PPSM) areas (representing rural areas) reported the highest smoking rates, those living in 2000-5000 PPSM areas (representing suburban areas) reported the lowest, and those living in >5000 PPSM areas (representing urban areas) reported intermediate smoking rates (10).

The consistency of the above results from two independent California surveys lends support to the robustness of our finding of high smoking prevalence in the least, as well as most densely populated areas in California. A cohort analysis of smoking prevalence employing CTS data from 1990 to 2008 revealed that low density populations had a higher smoking prevalence across cohorts, regardless of age (10). These findings demonstrate consistency of the nature and direction of the association of smoking and urban/rural place of residence, thereby minimizing the potential for chance findings or systematic error in our results.

Previous studies from the US indicated that rural areas had higher smoking prevalence than urban areas. Data from the US Behavior Risk Factor Surveillance System (BRFSS) showed that between 1994 and 1996 national smoking rates were 22.4% in urban areas and 24.6% in rural areas; this urban/rural difference did not change between 2000 and 2001 (2). Data from the BRFSS 2008 (3) showed a distribution of smoking prevalence between urban,
suburban and rural areas similar to our findings, but our analyses were more detailed, accounted for confounders, and data were collected over a much longer time period. The higher smoking rates in Second-City areas than in Urban areas in the 2007 and 2009 CHIS compared to other cycles represent an anomaly that cannot be accounted for by the variables examined in this study. One possible explanation is suggested by findings indicating that the 2008 financial crisis had a noticeable effect on the increase of smoking rates in the US (19). It may be that the recession led to migration and/or increased levels of stress, both of which have been associated with variations in smoking prevalence (4, 5, 7, 20). Although the population proportions of each geographic area remain relatively stable over time as shown in Table 1, it is possible that similar proportions of residents with different smoking status migrated between Urban and Second-City areas. However, this interpretation is speculative and further studies are needed to support this proposed explanation.

Analyses of pooled CHIS data showed that higher odds of smoking were associated with sociodemographic characteristics such as male sex, younger age, non-Hispanic ethnicity, lower education, full time worker or unemployed but looking for a job. However, controlling for all these sociodemographic factors did not change the observed differences of smoking prevalence by area. Thus the observed variations across geographic areas demand other explanations.

The higher smoking prevalence in Town/Rural areas may be explained by lower socioeconomic status, cultural norms, reach of tobacco control efforts, and access to health services in these areas (3, 10). The lower smoking prevalence in Suburban areas likely reflects the high portion of residents (40.1%) with college or above education. Higher education level generally indicates higher socioeconomic status, different social norms, and more knowledge about health risk behaviors, which are consistently associated with lower rates of smoking (3, 10). Factors associated with higher smoking prevalence in rural populations include disparities in tobacco control programs and resources, adoption of smoke-free policies, sociodemographic characteristics such as lower income and educational attainment, lower health insurance
coverage and reduced access to treatment services, and targeted marketing by the tobacco industry (3, 10, 21-24). Higher smoking rates in urban areas have been attributed to higher levels of stress from living in urban areas (4, 5, 7, 20). Other sources of stress may also lead to increased smoking rates. For example, Gallus and colleagues (19) found that more than 30% of unemployed people in the US were current smokers, and suggested that this higher smoking prevalence was likely related to stress from job loss. Our finding that individuals who are unemployed but looking for work have a higher smoking prevalence than those in the other employment statuses may reflect higher levels of stress associated with job seeking, however the present data cannot address this directly.

Further analysis of the pooled CHIS data within different ethnic strata showed that non-Hispanic Whites living in Town/Rural areas had higher smoking rates than those in other geographic areas. Yet, smoking rates were similar across these areas for Hispanics. Since non-Hispanic Whites are the majority ethnic population in California, it can be deduced that the “U-shaped” pattern of the population was driven by this group. In other words, one may assume that those who lived in low population density areas with higher smoking rates were mostly non-Hispanic Whites. Further analysis of the pooled data confirmed this assumption: more than two thirds of the population in rural areas were non-Hispanic Whites with smoking rates five times higher than those of other ethnicities. Thus, it is reasonable to infer that the tobacco-related health burden for non-Hispanics is likewise proportionally higher. Decreasing smoking prevalence within this geographically dispersed rural population in California represents a distinct challenge given the limited institutions and public health resources available to address smoking.

Similar findings of smoking prevalence between urban and rural areas in different ethnic strata were also found in the CTS data, which showed that the general pattern of prevalence for non-Hispanic Whites was similar to that seen for the entire population, and the prevalence within Hispanics was independent of population density of residence (10). The differences of
smoking prevalence between ethnic strata have also been observed at the national level. Data from the 2012-2013 cycle of the National Adult Tobacco Survey (NATS) showed that non-Hispanic Whites had higher smoking rates than Hispanics (17.2% vs. 14.6%) (25). When focusing on rural areas, this higher smoking prevalence in non-Hispanic Whites was evident even among teenagers (26), and this phenomenon has not only been observed in the US (27). These data together suggest that cultural influences on smoking behavior are intact regardless of urban or rural residency for these ethnic strata, but further investigation are required to establish which factors mediate these divergent findings according to ethnic background.

There are some possible limitations to this study. First, the smoking prevalence relied on self-report with no biomarkers to validate smoking status. However, this is the standard approach to measuring smoking in surveys and has construct validity. Second, the declining response rates may also affect the results, but the patterns observed were consistent across time despite declining response rate, which supports a lack of significant bias. Third, like other telephone surveys, households without a telephone (landline or cell phone) were not sampled. Multiple post-stratification weights were employed to address this and other sampling issues, which should minimize the potential influence of this kind of limitation. Fourth, due to the decline of landline users and increase of cell phone users, cell phone numbers have been included in the CHIS sampling since 2007. The addition of cell phones might have an impact on survey results; but as mentioned above, different post-stratification weights were used to minimize potential sampling issues. Further, the steady trend of decreasing smoking prevalence also indicates that the survey results are reliable with little or no influence attributable to the inclusion of cell phone numbers.

Strengths include that this was the first study to report smoking prevalence over a consecutive 12-year period in California, and the first to use the Claritas 4-level geographic classification to report finer resolution of urban/rural smoking prevalence. Other strengths contributing to the uniqueness of this study include large sample sizes at each time point, a
serial cross-sectional design covering a relatively large interval of time, consistent and rigorous sampling strategies, and the ability to survey respondents in multiple languages. This study was also the first to demonstrate different risks of being a current smoker across geographic locations within the two largest ethnic strata in California, non-Hispanic Whites and Hispanics.

Tobacco control efforts have been focused on urban and metropolitan areas, and there are clear disparities in tobacco control resources as well as in the success of such tobacco control programs in rural and low population density areas (22-24). Rural areas also have different social norms and behaviors which need to be taken into consideration by tobacco control programs (3, 10). Our study results clearly indicate a need for a population-specific tobacco control approach that addresses the lifestyle, behavior, and education of disparate populations within California in order to lower the higher than average smoking prevalence in such settings. Enforcement of the loosely implemented smoke free laws, education at an early age in schools, the use of multimedia and social network applications, and providing resources for quitting that are culturally in tune with rural communities are some examples of such a multipronged approach (21-23).

In summary, Town/Rural populations of California were consistently at higher risk of smoking than those living in Suburban areas; the more population-dense Urban areas of California also had high smoking prevalence, suggesting a “U-shaped” relationship of population density with smoking prevalence. This association is significant in non-Hispanic Whites, the largest ethnic population in California; while Hispanics, the largest minority population in California, reported consistently lower smoking rates that were independent of geographic location. In addition to living in Urban, Second-City, and Town/Rural areas, being non-Hispanic White was an additional risk factor for being a current smoker. The association between geographic locations and smoking prevalence persisted despite an ongoing decline in smoking prevalence across the cycle of surveys. Data from this study provide further information that
urban or rural settings are predictors of higher smoking prevalence, which may be relevant to future tobacco control efforts.
REFERENCES


