Title: Modeling geographic and demographic variability in residential concentrations of environmental tobacco smoke using national datasets

Running Title: Modeling residential environmental tobacco smoke

Teresa Chahine, ScD[1*]; Bradley Schultz, MS[2]; Valerie Zartarian, PhD[2]; SV Subramanian, PhD[1]; John Spengler, ScD[1]; James Hammitt, ScD[1]; Jonathan I Levy, ScD[1,3]

[1]Harvard School of Public Health; [2]EPA, Office of Research and Development, National Exposure Research Laboratory; [3] Boston University School of Public Health

*Corresponding author. Email: tchahine@post.harvard.edu, Tel. 617-669-2235, Fax: 617-384-8859, Express Mail to: 50 Commonwealth Avenue #904, Boston MA 02116

ABSTRACT

Background: In spite of substantial attention toward environmental tobacco smoke (ETS) exposures and health risks, studies to date have not provided adequate information to apply broadly within community-scale risk assessments.

Objectives: Our study aims to estimate residential concentrations of particulate matter (PM) from ETS in sociodemographic and geographic subpopulations in the United States for the purpose of screening-level cumulative or comparative risk assessment.

Methods: We developed regression models to estimate residential cigarette smoking using data from the 2006-7 Current Population Survey – Tobacco Use Supplement and linked this model with models of air exchange using housing data from the 2007 American Housing Survey. Using repeated logistic and log-linear models (n=1000), we investigated whether household demographic and geographic variables available from the 2000 U.S. Census can be used to predict the likelihood of residential exposure to ETS and the concentration of ETS-PM in exposed households.

Results: We estimated a mean concentration of residential ETS-PM of 16 $\mu g/m^3$ among the 17% of homes with non-zero exposure (3 $\mu g/m^3$ overall), with substantial variability in exposures among homes. The highest likelihood of residential ETS exposure was found in the South and Midwest regions, in rural populations, and in low-income households. ETS-PM concentrations in exposed households were highest in the South and demonstrated a non-monotonic association with income, partly related to air exchange rate patterns.

Conclusions: We provide estimates of ETS-PM concentration distributions for different demographic and geographic subpopulations in the United States, providing a useful

starting point for communities interested in characterizing aggregate and cumulative health risks from indoor air pollutants.

Key words: environmental tobacco smoke, secondhand smoke, residential exposure, air exchange rate, sociodemographic factors, fine particulate matter

INTRODUCTION

Environmental Tobacco Smoke (ETS) is an indoor air pollutant of great interest and concern for agencies including the U.S. Environmental Protection Agency (EPA) and the Centers for Disease Control and Prevention (CDC), among others (USEPA, 1992). The CDC in its 2006 Surgeon General's report estimated between 24,300 and 69,600 excess deaths per year from cardiovascular disease and between 3,423 and 8,866 excess deaths from lung cancer attributable to ETS nationwide (CDC, 2006). In its risk assessment and subsequent analyses, the California EPA estimated that health effects in children nationwide include more than 200,000 episodes of childhood asthma per year (new cases and exacerbations); and between 150,000 and 300,000 cases of lower respiratory illness, among other adverse health outcomes (CA EPA, 1997).

While these studies clearly indicate that the public health burden of ETS exposure is substantial, the health risk assessments have multiple limitations. These national risk estimates use a crude surrogate of ETS exposure, based on survey data indicating whether or not individuals live with smokers. This exposure metric was chosen to correspond with the epidemiological studies applied in these national risk assessments, as quantitative measurements of ETS concentrations and exposure have not been available for all participants in these studies (CDC, 2006). Such an approach does provide reasonable estimates of national-scale health risks, albeit with some uncertainties, but the reliability of this surrogate measure may vary across demographic subpopulations, and the approach lacks sufficient resolution to identify potentially meaningful community differences in

ETS exposure. The smoking literature has shown significant sociodemographic and geographic variation of smoking prevalence and intensity (Shavers et al., 2005; Shavers et al., 2006; Datta et al., 2006; Osypuk et al., 2006); furthermore, housing characteristics such as home volume and air exchange rate will significantly influence the exposure implications of indoor smoking (Klepeis and Nazaroff, 2006). Demographic factors correlated with smoking activity may also be correlated with housing characteristics of interest, influencing the patterns of ETS concentrations and potentially increasing within-community and between-community variability.

There is a growing interest in quantifying ETS exposure at the community level because of its ubiquity in indoor environments and its association with multiple health outcomes of concern as described above. Community organizations, local governments, and federal agencies have cited ETS as an environmental issue of concern within the context of community-based cumulative assessment of indoor pollutants such as particulate matter (PM) associated with asthma (Zartarian and Schultz, 2009). However, national-scale analyses lack sufficient resolution for the above-mentioned reasons and rarely are local data available on ETS exposure (USEPA, 1992).

A substantial number of published studies have provided insight about ETS concentrations in indoor microenvironments or biomarkers of ETS exposures, but none have generated the data that would be needed to characterize exposure systematically at a community level across the U.S. Studies conducted prior to the widespread introduction of smoking restrictions in commercial and recreational facilities directly measured concentrations of ETS-PM using personal monitors, but the evidence from these studies cannot be used directly to construct exposure models in residential settings (Repace, 2004; Repace et al., 2006; Jenkins et al., 1996; Jenkins et al., 2001). Given increasing indoor smoking restrictions in public places including workplaces as well as retail, hospitality, and commercial venues, the majority of the exposure and risk burden may shift to the residential environment, but no study to date has developed broad-based and generalizable models of ETS exposure inside homes. Previous residential studies have either characterized ETS contributions to residential exposures within hypothetical simulation models or using measurements but without a nationally generalizable model framework (Dockery and Spengler, 1981; Leaderer, 1990; Ozkaynak et al., 1996; Klepeis and Nazaroff, 2006; Myatt et al., 2008). It is therefore unknown how variability in smoking patterns correlates with housing characteristics that influence indoor concentrations of ETS, necessary in determining variability in residential ETS concentrations and risk. Studies using cotinine as a biomarker of ETS exposure among children have demonstrated significant associations with parental education, race/ethnicity, income, and home size, but did not yield models directly applicable to all locations at fine geographic resolution and had challenges in separating predictors of exposure from factors that could influence metabolism (Mannino et al., 2001; Max et al., 2009; Marano et al., 2009).

The aim of this study is to develop a model that can be used to systematically estimate residential ETS concentrations in demographic and geographic subpopulations using publicly available data across the U.S. for the purpose of screening-level cumulative or comparative risk assessment. We develop and apply regression models based on national datasets to estimate residential cigarette smoking and air exchange rate patterns in order to characterize the distribution of PM concentrations from ETS in U.S. homes. We investigate whether household demographic and geographic variables that are widely available at fine geographic resolution can be used to predict (1) the likelihood of residential exposure to ETS-PM and (2) concentration levels of ETS-PM in exposed households.

METHODS

Study Population

The study population for the ETS model consisted of 98,329 individuals residing in 39,107 homes for which occupant interviews were conducted in the 2007 National American Housing Survey (AHS) (U.S. Department of Housing and Urban Development, 2009). The AHS is a nationally representative survey of the housing stock in the U.S. Information collected during the survey included occupants' demographic characteristics, which we used to model the likelihood of smoking; and housing characteristics, which we used to model resulting ETS concentrations. Table 1a summarizes the steps conducted to estimate and model variability in ETS concentrations. Table 1b provides a summary of parameters and assumptions used to calculate the estimated ETS-PM emissions and equilibrium concentrations. Both the model steps and parametric assumptions are described in more detail below. Unless otherwise stated, all steps in the analyses below were conducted using SAS 9.2 (SAS Institute Inc.).

CPS-TUS Smoking Models

Two smoking models were developed for this study, using data from the 2006/7 Current Population Survey – Tobacco Use Supplement (CPS-TUS) (NCI, 2007), n = 227, 428. These data have previously been used within multilevel logistic regressions to predict the probability of smoking based on individual characteristics (age, sex, poverty, race/ethnicity, nativity, education, occupation, employment, marital status, and number of people in the household) and area characteristics (state smoking laws, state taxes, percent poverty of core-based statistical area); more details about the underlying statistical approach are described elsewhere (Chahine, 2010).

For the purpose of our analyses, we needed to make multiple modifications to the previous modeling framework (Chahine, 2010), given the focus in this analysis on residential exposures as well as constraints in available data within other steps of the modeling (Table 1a). In our first model (Model 1a), we used as our outcome variable smoking at home rather than smoking overall. The TUS asked respondents whether smoking is prohibited inside the home, permitted in some places or at some time inside the home, or permitted everywhere inside the home. Subjects who replied that smoking is prohibited inside the home were assigned a value of 0 for the outcome; all other subjects were assigned a value of 1. Because no further detail was provided in the survey, for the purpose of this analysis we did not differentiate between subjects who were permitted to smoke in some places or at some time inside the home vs. everywhere inside the home.

In our second model, we used TUS data on the number of cigarettes smoked daily for the subset of people who smoke at home, which was predicted using linear regression (Model 1b). We were not able to include occupation and employment in our models, which were statistically significant predictors previously (Chahine, 2010), because insufficient information was provided on these covariates in the AHS dataset. We were also not able to include random effects at the levels of the state and the core-based statistical area, due to insufficient geographic identifiers in the AHS dataset. A household random parameter was included in our models to account for the nested structure of the AHS data. Dummy variables were added to represent census region and metropolitan status (central city vs. suburb). Models 1a and 1b were created using MLwiN 2.16 (Rasbash et al., 2002).

24% of the TUS study population did not complete the portion of the survey dedicated to rules of smoking at home and work. These subjects were either unemployed, self-employed, or retired. Summary statistics were performed on this subpopulation to determine whether they differed from the general population in demographics and smoking prevalence. A further 1% of the TUS study population did not know or refused to answer this portion of the questionnaire. Therefore the total sample size upon which Model 1a was built was 169,061. Only people who smoked at home were included in the analysis for Model 1b (n=18,529), and an additional 164 people were excluded due to missing values for number of cigarettes smoked daily, resulting in a total sample size of 18,265 for Model 1b.

AHS Smoking Predictions

As described above, we used Model 1a to calculate the predicted probability of smoking at home for each individual in the AHS, according to their sociodemographic characteristics (age, sex, poverty, race/ethnicity, nativity, education, marital status, and number of people in the household) and geographic covariates (census region and urban status). The predicted probability was then used to assign a binary home smoking status for each individual, to appropriately capture the fact that each household either does or does not have a smoker. This was done by generating a random number between 0 and 1 and determining whether this number was less than the individual's predicted probability of smoking at home (resulting in an assignment of a home-smoking status of 1) or greater than the predicted probability of smoking at home (resulting in an assignment of a home-smoking status of 0). The resulting simulated dataset of homesmokers vs. non-homesmokers was then carried forward in the subsequent analyses described below. This process was repeated 1000 times to construct 1000 simulated datasets. All individuals under the age of 18 were assigned a home-smoking status of zero because the underlying TUS data on which the smoking models were built included adults only.

For those with a home-smoking status of 1, we then calculated the predicted number of cigarettes smoked daily according to their sociodemographic characteristics and geographic location using Model 1b. Based on previous literature, on average people spend approximately one third of their time outside of the home, one third of their time inside the home during waking hours, and one third of their time inside the home during non-waking hours (Klepeis, 1999; Nazaroff and Singer, 2004). Therefore, we assumed that our study population spends one half of its waking time inside of the home, and multiplied the predicted daily number of cigarettes smoked by a factor of 0.5 to obtain the predicted daily number of cigarettes smoked at home (Table 1b). We assumed a constant rate of smoking; although it is possible that more cigarettes are smoked outside of working hours given the increasing smoking restrictions in public places, no information was available on this and we chose to use a linear rate based on the previous literature (Klepeis and Nazaroff, 2006).

Air Exchange Model

To estimate air exchange rates for the houses in the AHS dataset, we used a previously published regression model developed using a residential leakage database for single-family detached homes (Chan et al., 2005). Homes were separated into conventional and low income, and model parameters were applied to calculate normalized leakage based on year built and floor area. Floor area was reported as an ordinal variable in the AHS, while year built was reported in interval categories. For our calculations, each home was assigned the midpoint of its year-built category. Year built was top-coded at 2000 and floor area at 600 m^2 (which fell at the 96th percentile of the AHS data), to maintain consistency with the residential leakage database used to develop the air exchange model. Floor area was missing for 7914 houses and was estimated by multiplying the number of rooms in these houses by the median room size in the dataset (25 m^2). Normalized leakage was then translated into air changes per hour using an equation as described by Chan et al., 2005. An additional factor of two was applied for multi-unit housing structures, based on a previous study which analyzed all currently available data on multi-unit housing structures and reported that multi-unit housing structures are on average twice as leaky per unit area as single-unit homes, with little systematic variation in building leakage by construction type, building activity type, height, size, or location within the U.S. (Price et al., 2003). While our approach is quite uncertain, direct application of the Chan et al. model to multi-family housing was not appropriate and no additional data sources were available; we consider the implications of this assumption within our analysis.

ETS Concentration Estimation

Screening-level estimates of residential ETS concentrations were calculated using a single-zone mass-balance equation, assuming perfect mixing: $C = (Q/V)/(a+k)$, where:

C = equilibrium concentration ($\mu g/m^3$)

Q = emission rate (μg/hour)

V = volume of housing unit (m^3)

a = air exchange rate (1/hour)

k = deposition rate (1/hour)

Emission and deposition parameters were selected based on previously reported central estimates: we used 10 mg per cigarette for PM emissions and 0.1/hr for particle deposition loss-rate coefficient (Klepeis and Nazaroff, 2006). These previously reported central estimates were based on an assimilation of multiple previous studies (Klepeis et al., 2003; Martin et al., 1997; Xu et al., 1994).

Statistical Analysis of Estimated ETS-PM Concentrations

Multiple analyses were conducted on the resulting ETS-PM concentrations to develop broadly-applicable models describing the association of ETS-PM with sociodemographic and geographic variables available across the U.S. The analyses were performed at the household level rather than the level of the individual, because the individuals in our study population are clustered within households. Given available information, we assumed that all individuals within a household are exposed to the same equilibrium concentrations of residential ETS, though time-activity pattern differences could clearly lead to differential personal exposures.

We restricted the household predictors to variables which are available both in the AHS dataset and cross-tabulated in the U.S. Census by census tract. A census tract is a small statistical subdivision of a county, usually containing between 2500-8000 persons (U.S. Census Bureau, 2000). We chose census tracts as our geographic resolution because they were designed to be homogeneous with respect to population characteristics, economic status, and living conditions. The U.S. Census provides summary tables for singular household variables, and cross-tabulations of selected variables. The largest cross-tabulation of multiple householder demographic variables available is: Householder Race by Householder Age by Household Income. We also included dummy variables for census region and urban/rural status.

A logistic regression model was fit to estimate the predicted probability of non-zero ETS-PM in the indoor home environment with respect to the above variables (Model 2a). For the subset of people with non-zero residential ETS concentrations, a log-linear regression model was fit to describe the distribution of ETS-PM concentrations in the indoor home environment (Model 2b). These regression models were fit for each of the 1000 simulated datasets, and we report results representing the mean across regression model fits.

RESULTS

Models 1a and 1b: Home smoking prevalence and daily cigarettes smoked in CPS-TUS

The average prevalence of home-smokers in the TUS dataset was 10.9%, compared to an overall smoking prevalence of 17.7%. The difference is attributable to the fact that 39.7% of smokers reported having smoke-free homes. The mean number of total daily cigarettes smoked by home-smokers was 16.2. Examining the characteristics of the subset of the study population with missing data on home smoking rules showed similar overall smoking prevalence to the overall population and no significant differences in sociodemographic covariates, except that it contains more people in the 18-24 age group and fewer people in the never-married group than the general study population.

Models 1a and 1b shared many common patterns with respect to sociodemographic and geographic covariates (Tables 2 and 3). The reference sociodemographic population for Models 1a and 1b was White non-Hispanic women aged 45-54 with a high school degree who are native U.S. citizens, currently married, have an annual household income between $30-60,000 and live with two or more people; the reference geographic categories were Census region 3 (i.e. South) and Metropolitan balance (i.e. not a central city). Compared to the reference population, higher odds of smoking at home and higher number of daily cigarettes among home-smokers were found in White Non-Hispanic males, native U.S. citizens, and those with lower education levels. A U-shaped pattern was observed for the association between age and smoking at home, peaking at age 45-54 in both models. Living in a metropolitan suburb was associated with lower odds of smoking at home, as was living in the Western U.S. (Census Region 4). The effect of gender was found to differ by race in both models, as indicated by statistically significant interaction terms between the two variables.

Significant effect modification was found between Census region and race in Model 1a (Table 2). This is consistent with previous findings which showed the effect of race to differ by state (Osypuk et al., 2006; Chahine, 2010). However, the interaction between these two variables was not statistically significant in Model 1b, and was not retained in the final model (Table 3). Other key differences between Model 1a and Model 1b include the fact that those who live alone were more likely to smoke at home but smoked fewer cigarettes per day relative to home-smokers living with two or more people, as well as the fact that those who live in metropolitan central cities were more likely to smoke but smoked significantly fewer cigarettes per day at home. As indicated by the 95% confidence intervals in Table 3, while annual household income was one of the most significant predictors of smoking at home, it did not predict the daily number of cigarettes smoked by people who smoke at home. In general, as can be seen from the confidence intervals in Tables 2 and 3, most predictors had weaker associations with the outcome in Model 1b compared to Model 1a. This could be due both to smaller sample size and to a lesser effect of demographic and geographic factors on the smoking intensity of home-smokers.

Air Exchange and ETS-PM Simulations in AHS

Estimated air changes per hour (ACH) varied from a low of 0.07 to a high of 2.7. The mean estimated ACH across households was 0.5/hr, with a median of 0.4/hr and standard deviation of 0.4/hr. Slight regional variations were observed, with the highest air exchange rates in the Northeast (mean 0.6/hr, median=0.5/hr) and lowest in the South (mean 0.4/hr, median=0.3/hr).

Across all 1000 simulated datasets, the mean prevalence of smoking at home in our AHS study population was 10.9%, similar to the TUS study population, with a mean number of cigarettes predicted for homesmokers of 14.9, slightly lower than the mean of 16.2 observed in the TUS. Approximately 17% of homes had non-zero ETS concentrations. The mean household concentration of ETS-PM in the total population across individual homes and simulations was 2.8 $\mu g/m^3$, with a median of zero and a standard deviation (SD) of 8.9. Looking at the exposed subset only, the mean household concentration was 16.3 $\mu g/m^3$ (median 13.0 $\mu g/m^3$, SD 15.9); the 99th percentile was 67 $\mu g/m^3$ and the maximum concentration reached over 499 $\mu g/m^3$, indicating that a small percentage of the population is predicted to be exposed to extremely high concentrations of ETS-PM. Figure 1 presents box-plots of ETS-PM concentrations for key sociodemographic and geographic covariates.

Models 2a and 2b: Statistical Analysis of Estimated ETS-PM Concentrations

Odds of ETS exposure (Model 2a) showed generally similar patterns of association with many householder covariates as was observed in the underlying smoking prevalence model with individual covariates (Model 1a), with the exception of race (Table 4). Concentrations of ETS-PM in the exposed population (Model 2b) showed similar patterns of association with householder covariates as was observed in the underlying number of cigarettes model with individual covariates (Model 1b), with the exception of Census region (Table 5). While households in the Midwest region had the highest odds of ETS exposure, exposed households in the South had the highest mean concentrations after controlling for household income, householder age, and urban/rural status. This reflects the role of housing characteristics in determining ETS-PM concentrations, as the lowest air exchange rates were estimated in the South.

A non-monotonic relationship was observed between household income and ETS-PM concentration in the exposed homes, where households in the second lowest income quartile had the highest concentration (Table 5). This is a direct result of the low-income air exchange model provided by Chan et al. (2005) and the factor of two that we further applied to attached units (which are most prevalent in the lowest income category). R-square values for the log-linear regression models were on the order of 0.2, indicating that a substantial proportion of variability in ETS-PM concentrations remains unexplained by the covariates controlled for in our model. Standard errors of the individual models (n=1000) were on the same order as the standard deviation of the mean parameter estimates (intercept and coefficients, denoted by β) for both models shown in Tables 4 and 5, indicating that the model is appropriately characterizing uncertainty (results not shown).

DISCUSSION

Our results demonstrate that the substantial variability in residential ETS concentrations, driven by variability in both smoking patterns and housing characteristics, can be reasonably explained by household sociodemographic and geographic variables which are publicly available at the census tract level. Although many influential factors could not be captured within the constrained set of covariates available through census cross-tabulations, the models are interpretable and help identify sociodemographic and geographic subpopulations in the United States which are at higher risk of elevated ETS exposures.

The mean estimated ETS-PM concentration of 16 $\mu g/m^3$ for exposed households in our study falls on the lower side of the range of concentrations measured in previous large-scale residential ETS studies (Dockery and Spengler, 1981; Leaderer, 1990; Ozkaynak et al., 1996). Our results are also in general agreement with recent modeling studies: Klepeis et al reported means ranging from 6.6-49 $\mu g/m^3$ from multiple simulations using a simple box model approach; and Myatt et al reported a mean of 15 $\mu g/m^3$ and a median of 17.8 $\mu g/m^3$ using the CONTAM model (Klepeis and Nazaroff, 2006; Myatt et al., 2008). These studies incorporated time-activity patterns and ventilation patterns but relied on a hypothetical smoking population for their simulations, while our study is a snapshot characterization of equilibrium concentrations based on parameters estimated from housing characteristics and incorporating smoking variability from a national survey. Thus, while these studies showed that exposure concentrations may vary *within* a household depending on ventilation behavior and time-activity patterns of occupants, our study suggests that time-averaged concentrations can vary *between* households depending on smoking patterns and house type, volume, and age (all of which show variation among sociodemographic and geographic subpopulations).

In our multiple logistic and linear regression models we attempted to capture variability in log odds of ETS exposure (Model 2a) and in the concentration of ETS-PM in exposed populations (Model 2b) using covariates available across the U.S. from the Census. While the majority of the estimated parameters were consistent with those of Models 1a and 1b (e.g., a U-shaped curve for age, higher exposure in rural locations and in the Midwest), some parameters had differing relationships for ETS than for smoking. In particular, race was not a significant predictor of residential ETS exposure in Model 2a. One possible explanation for these results is that the grouping of the variables was modified in order to be consistent with how these variables are defined and presented in the 2000 Census. Race/ethnicity were grouped together in Model 1a, to maintain integrity with the original smoking model from which it was developed, which was determined based on the previous literature. However, the 2000 Census presents race without the ethnicity component. In Model 1a, the log odds for the Black-Non Hispanic indicator variable was borderline significant at $p=0.046$, while the log odds for the Hispanic indicator variable was highly significant at $p < 0.001$. The association captured in Model 1a may have been largely due to the ethnicity component rather than the race component, which was not captured by the race variable in Model 2a.

Our findings can be additionally evaluated by contrasting them with studies examining cotinine levels among children, taken from the National Health and Nutrition Examination Survey (NHANES). Among children with reported smoke exposure in the home (paralleling our Model 2b), cotinine levels were significantly higher among households with low parental education, white ethnicity (relative to Mexican-American ethnicity), and households with fewer rooms, with a borderline significant effect of poverty status and no significant difference by region or white/black race (Mannino et al., 2001). While our findings are not directly comparable given different covariates, and the NHANES study includes factors influencing metabolism and dosimetry, both studies reinforce the importance of housing factors and more complex associations with household income than seen in studies of smoking patterns.

One of the major uncertainties in our analysis relates to our modeling of air exchange rates. We applied a leakage model developed for single-family detached units which takes into consideration floor area, year built, and low-income housing status to estimate air exchange, and adapted the model to accommodate attached units. The literature on air exchange rates in attached housing units is scarce, and theoretical understanding of factors influencing air exchange could lead us to argue for either higher or lower air exchange relative to detached units. On the one hand, the lack of a ceiling/roof providing a direct pathway to the outdoors and the smaller ratio of exterior wall area per unit of interior volume in attached homes may point to smaller air exchange rates. On the other hand, larger buildings may contain more opportunity for leakage such as cracks and leaks; the natural physical forces that move air (e.g., wind and stack effect) are more pronounced in taller buildings; and there may be weaker financial incentives for energy-efficient construction or retrofits (Diamond et al., 1996). Our choice of multiplying by a factor of two for attached units is based on a study which analyzed all available data on multi-family homes and reported greater leakage and a larger air exchange rate than single-family homes (Price et al., 2003). The available data are not statistically representative and consist of measurements of indoor-outdoor air changes per hour (ACH) in individual apartments within sixteen different apartment buildings. The observed rates of 0.5 to 2 ACH were approximately twice those of single-family homes, and leakiness values in the same studies (3 to 8 $L/s \cdot m^2$) were also approximately twice those of single-family homes. The study employed Bayesian hierarchical modeling to address problems caused by small samples sizes, and no systematic variation was found with construction type, building activity type, height, size, or location of the buildings. To assess the sensitivity of our results on this assumption, we removed the factor of two and repeated our analysis. This did not change the general patterns observed in Model 2a and 2b (results not shown). In the absence of a larger body of knowledge on air exchange in multi-family housing, we relied on this analysis, but additional data collection would clearly be warranted.

Our study has a number of additional limitations. First, lack of availability of state and metropolitan area identifiers for over half the study population in the AHS inhibited our ability to make use of more detailed geographic inputs that could have been included in the smoking model from the CPS-TUS. While this reduced our predictive power, these inputs were previously shown to have a smaller contribution to predicting smoking variability than individual socioeconomic and demographic variables (Chahine, 2010). Given our inclusion of geographic information in the form of Census region and urban status indicator variables, we believe that our model still captures much of the variability in smoking that can be predicted with available covariates. Second, the use of a simple box model to estimate ETS concentrations may not adequately capture indoor conditions, given the multi-compartment nature of most houses. For a more sophisticated exposure assessment, the use of a more detailed exposure model taking into account the differences in smoking rules within a house, the number of rooms in the house, the movement of people within the house, and more detailed ventilation information is required. However, because the aim of our study is to provide screening-level estimates, we chose a simple box model. Third, although the previously published air exchange model used in our study established that higher leakage rates are found in low-income homes, the data used for such homes came from a state database and the model likely does not capture variability in leakage in different regions of the U.S. (Chan et al., 2005).

Beyond the concerns mentioned above with estimating air exchange rates in multi-family housing, the scaling factor of two addresses indoor-outdoor air exchange only; our study does not take into account exchange between units in the same building, which may be a substantial contributor of ETS, nor the variability in air exchange rates for different units in the same building (Bohac et al., 2007). Another limitation of our model is that the simulated home-smoking includes adults aged 18 and above only, which may result in an underestimation of residential ETS exposure. The CPS-TUS and other comparable national surveys such as the CDC's Behavioral Risk Factor Surveillance System (BRFSS) do not collect data on smoking in children and teenagers. While data are available in the literature on smoking in children and teenagers, these data were collected in separate youth-focused surveys (CDC, 2010). We recommend that children and teenagers be included in the study population for future CPS-TUS data collection. Finally, our simulation method does not formally take into account which individuals in our dataset live in the same homes. Although Model 1a incorporates within-house correlations in smoking into the beta parameters used to estimate predicted probability of home-smoking in this study, the random number generation method does not further incorporate information on individuals living in the same home in the AHS dataset (i.e., the likelihood that a smoker will live with another smoker, beyond the demographic factors that predict individual behaviors).

More generally, the results of our ETS-PM regression models shed light on the average differences between subpopulation groups as represented by household income, householder race, and householder's age, using variables that are universally available at the census tract level. However, many other factors remain which influence residential ETS exposure and which have not been controlled for in our analysis. These factors could include further individual-level variables in addition to variables operating at the neighborhood, town, metropolitan, and state levels. Therefore, although our estimates are an improvement on national averages, use of our model should be restricted to screening purposes, to identify likely exposure hot spots where more detailed local data could be gathered and to inform screening-level community cumulative exposure assessments.

Despite these limitations, this study leverages data from three national datasets to provide screening estimates of residential ETS exposure in demographic and geographic subpopulations across the U.S. While previous studies have used average smoking and air exchange statistics in ETS exposure simulations, our study captures variability in residential smoking patterns and housing characteristics, including factors that correlate with both smoking behaviors and air exchange. Our modeling approach, while constrained in the predictors that were considered, would theoretically allow for ETS exposure to be systematically estimated for a census tract given its location and demographic cross-tabulations. Future studies should investigate correlations between ETS and other indoor air pollutants in the home in order to determine whether subpopulations with highest ETS exposure are also more highly exposed to high concentrations from other indoor sources. Given the high level of concern among communities for prevalent health risks such as asthma, more awareness on the hazards of ETS exposure is needed at the community level.

Although no relative risks are established for health outcomes from ETS-PM, evidence based on ambient PM with appropriate adjustments for personal exposures vs. outdoor concentrations could allow for a screening-level characterization of risk from residential

ETS-PM exposure. To contextualize this risk, the annual average residential ETS-PM concentration of 16 μg/m^3 that we have estimated for exposed households in the U.S. is higher than the annual average outdoor $PM_{2.5}$ concentration of 12 μg/m^3 reported for 2007 (USEPA, 2008). Further, others have stated that exposure to indoor particulate matter is as hazardous, if not more hazardous, than exposure to outdoor particulate matter (Nazaroff and Singer, 2004; Repace, 2007). While only a subset of the population is exposed to residential ETS-PM, and relative risks may differ between the mixture of particles in ambient PM and those in ETS-PM, the health risks may be of a similar order of magnitude as those from ambient PM for this exposed subpopulation.

More generally, by providing estimates of ETS-PM concentration distributions for different subpopulations, we allow for future risk assessments that move beyond "living with a smoker (yes/no)" exposure classification. These estimates are a useful starting point for communities interested in characterizing cumulative health risks from indoor air pollutants, which are likely substantial contributors to population disease burdens.

Acknowledgements: This research was supported by EPA Student Services Contract #EP09D000560. We gratefully acknowledge Dr. Andrew Geller and Dr. Daniel Vallero of EPA's Office of Research and Development, National Exposure Research Laboratory, for providing technical review of the paper.

Table 1a. Steps conducted to model variability in residential ETS concentrations:

1) Develop multi-level regression models predicting home-smoking as a function of sociodemographic and geographic covariates available in the AHS, using data from the 2006-7 CPS-TUS.
   - Model 1a: predicted probability of smoking at home
   - Model 1b: predicted number of cigarettes smoked daily by home-smokers

2) Apply parameters from Models 1a and 1b to calculate predicted probability of smoking and predicted daily cigarettes among smokers for all individuals in the AHS based on their demographic and geographic covariates.

3) Assign binary home-smoking status for all individuals in the AHS dataset through multiple simulations using their predicted probability of smoking at home.

4) Estimate ETS-PM emissions per person by combining binary home-smoking status with predicted daily cigarettes, emission rates per cigarette, and time spent at home.

5) Apply previously developed air exchange model to estimate air changes per hour for each AHS home using floor area, year built, and unit type.

6) Sum ETS-PM emissions across each household and apply one-compartment mass balance model to obtain screening-level estimates of equilibrium ETS-PM concentrations in each AHS home, assuming perfect mixing.

7) Apply statistical analysis to describe association of ETS-PM concentrations in the home with demographic and geographic household covariates, constraining covariates to publicly available census-tract cross-tabulations (U.S. Census 2000) to allow for broad-based extrapolation.
   - Model 2a: predicted probability of ETS-PM exposure at home
   - Model 2b: predicted concentrations of ETS-PM in exposed households

Table 1b.
Parameters and model assumptions used to calculate ETS-PM emissions and equilibrium concentrations.
(References for each parameter can be found in the text.)

Parameter	**Assumption and central estimates used in model**
Total daily number of cigarettes smoked by homesmokers	Estimated by applying Model 1b for each individual in AHS 2007 dataset
Proportion of total waking hours spent at home	Half of total waking hours are assumed to be spent inside the home.
Rate of cigarette consumption	Cigarettes are assumed to be consumed at a constant rate during waking hours.
Particulate Matter (PM) emitted per cigarette	1×10^4 µg/cigarette
Volume of housing unit	Calculated from floor area provided in AHS 2007 dataset, assuming average height of three meters.
Air changes per hour	Estimated using unit type, year built, floor area
Particle deposition loss-rate coefficient	0.1/hr

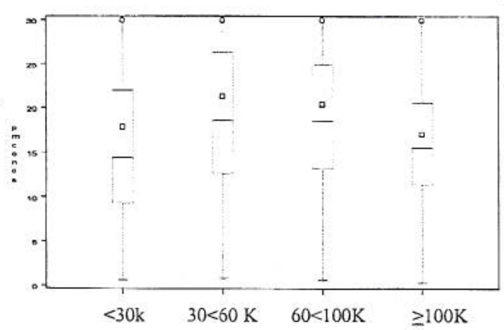
Table 2. Multilevel logistic regression model of home smoking as a function of geographic and demographic covariates (Model 1a). Reference categories not shown.	
PARAMETER	OR (95% CI)
Intercept	0.21 (0.20, 0.23)
Men	1.09 (1.05, 1.13)
<$30,000/year	1.42 (1.36, 1.49)
$60,000-99,000/year	0.67 (0.63, 0.70)
≥$100,000/year	0.44 (0.40, 0.47)
Family income not reported	0.82 (0.77, 0.87)
Age 18-24yrs	0.46 (0.43, 0.50)
25-34yrs	0.75 (0.71, 0.79)
35-44yrs	0.94 (0.90, 0.99)
55-64yrs	0.69 (0.66, 0.73)
65-74yrs	0.31 (0.29, 0.33)
≥75yrs	0.09 (0.08, 0.10)
Black Non-Hispanic	0.59 (0.50, 0.71)
Hispanic	0.49 (0.40, 0.59)
Other Non-Hispanic	1.01 (0.80, 1.27)
Not a native U.S. Citizen	0.40 (0.37, 0.44)
Formerly married	1.60 (1.52, 1.69)
Never married	1.98 (1.89. 2.08)
Less than High School Edu	1.51 (1.43, 1.58)
Some College	0.70 (0.67, 0.73)
College Degree	0.32 (0.30, 0.34)
Graduate Degree	0.21 (0.19, 0.23)
Living alone	1.26 (1.19, 1.33)
Living with 1 other person	1.23 (1.18, 1.29)
Northeast Census region (NE)	0.88 (0.83, 0.92)
Midwest Census region (MW)	0.94 (0.90, 0.99)
West Census region (W)	0.64 (0.60, 0.67)
Metropolitan Central City	1.10 (1.05, 1.16)
Nonmetropolitan	1.10 (1.05, 1.15)
Unidentified	1.06 (1.01, 1.11)
Interaction term: Black*Men	1.36 (1.22, 1.53)
Hispanic*Men	1.53 (1.32, 1.78)
Other*Men	1.36 (1.17, 1.59)
Interaction term: Black*MW	1.21 (0.98, 1.49)
Black*S	0.77 (0.64, 0.92)
Black*W	1.39 (1.08, 1.79)
Hispanic*MW	0.88 (0.67, 1.15)
Hispanic*W	0.50 (0.40, 0.63)
Hispanic*S	0.64 (0.51, 0.79)
Other*MW	1.30 (1.00, 1.70)
Other *S	0.79 (0.60, 1.03)

Table 3. Multilevel regression model of total number of cigarettes smoked daily by home-smokers as a function of geographic and demographic covariates (Model 1b). Reference categories not shown.

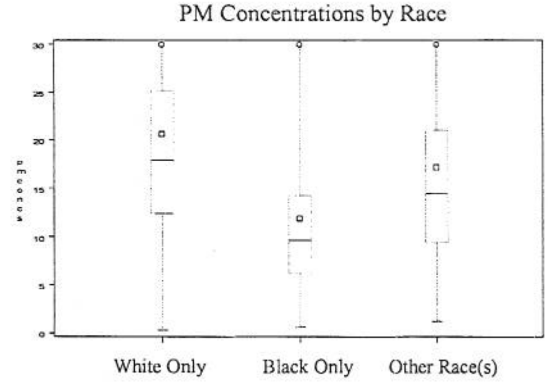
PARAMETER	ESTIMATE (95% CI)
Intercept	19.00
Men	3.93 (3.59, 4.26)
<$30,000/year	-0.001 (-0.38, 0.38)
$60,000-99,000/year	0.05 (-0.46, 0.56)
≥$100,000/year	-0.61 (-1.38, 0.15)
Family income not reported	-0.07 (-0.63, 0.50)
Age 18-24yrs	-5.57 (-6.25, -4.88)
25-34yrs	-2.81 (-3.32, -2.30)
35-44yrs	-1.13 (-1.57, -0.69)
55-64yrs	-0.13 (-0.59, 0.34)
65-74yrs	-1.17 (-1.80, -0.54)
≥75yrs	-3.73 (-4.68, -2.79)
Black Non-Hispanic	-6.00 (-6.72, -5.29)
Hispanic	-4.62 (-5.70, -3.53)
Other Non-Hispanic	-2.26 (-3.25, -1.28)
Not a native U.S. Citizen	-2.36 (-3.15, -1.58)
Formerly married	-0.77 (-1.27, -0.28)
Never married	0.19 (-0.22, 0.61)
Less than High School Edu	0.83 (0.42, 1.25)
Some College	-1.19 (-1.56, -0.82)
College Degree	-3.21 (-3.81, -2.61)
Graduate Degree	-3.96 (-4.96, -2.96)
Living alone	-1.23 (-1.70, -0.76)
Living with 1 other person	-0.36 (-0.73, 0.02)
Northeast Census region (NE)	-1.44 (-1.88, -1.00)
Midwest Census region (MW)	-0.87 (-1.26, -0.49)
West Census region (W)	-1.54 (-1.99, -1.08)
Metropolitan Central City	-0.87 (-1.33, -0.42)
Nonmetropolitan	0.48 (0.08, 0.88)
Unidentified	0.01 (-0.42, 0.45)
Interaction term:	
Black*Men	-2.79 (-3.82, -1.77)
Hispanic*Men	-2.83 (-4.27, -1.39)
Other*Men	-0.43 (-1.80, 0.94)

Figure 1. Distribution of ETS-PM concentrations ($\mu g/m^3$) among households with non-zero exposures, stratified by selected demographic and geographic covariates.
(Upper limit of box represents 75th percentile; values >30 $\mu g/m^3$ not shown.)

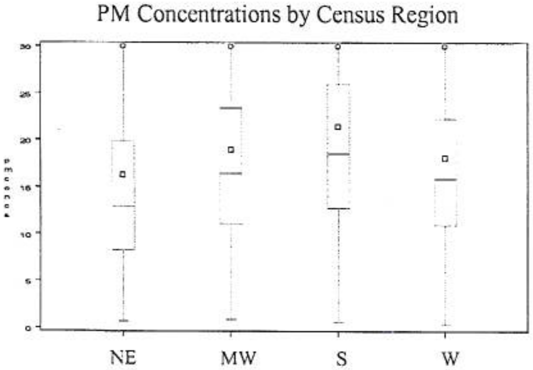
PM Concentrations by Annual Household Income ($)



PM Concentrations by Race



PM Concentrations by Census Region



PM Concentrations by Urban Status

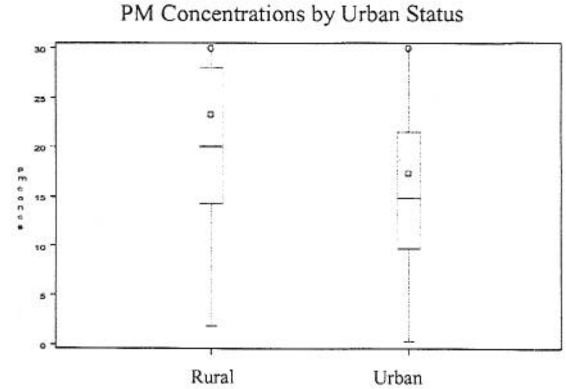


Table 4. Results from multiple logistic regression models of residential ETS exposure as binary outcome (N=1000). Mean parameter estimates (intercept and coefficients, denoted by β), odds ratios, and 95% confidence intervals are presented.

Model 2a.	**Variable**	**Mean β**	**SD**	**OR**	**95% CI**	
Intercept		-1.93	0.03			
REGION	NE	0.45	0.05	1.57	1.43	1.72
	MW	0.63	0.04	1.88	1.72	2.05
	S	0.50	0.04	1.65	1.52	1.79
	W	0				
Householder's Age	<25	-0.43	0.07	0.65	0.57	0.74
	25-34	-0.24	0.05	0.79	0.72	0.86
	35-44	0				
	45-54	0.25	0.04	1.28	1.19	1.39
	55-64	-0.04	0.04	0.96	0.89	1.04
	65-74	-0.62	0.05	0.54	0.48	0.59
	75+	-1.46	0.06	0.23	0.21	0.26
Householder's Race	White Only	0				
	Black Only	-0.10	0.04	0.91	0.84	0.99
	Other Race(s)	-0.08	0.06	0.92	0.82	1.04
Urban Status	Urban	0				
	Rural	0.23	0.03	1.25	1.18	1.33
Household Income	<30K	0				
	30<60K	-0.30	0.03	0.74	0.70	0.79
	60<100K	-0.76	0.04	0.47	0.43	0.50
	≥100K	-1.37	0.04	0.26	0.23	0.28

Table 5.
Results from multiple log-linear regression models (N=1000) of ETS-PM concentrations ($\mu g/m^3$) in exposed households. Mean parameter estimates (intercept and coefficients, denoted by β), and 95% confidence intervals are presented.

Model 2b.	Variable	Mean β	SD	95% CI	
Intercept		2.87	0.02	2.83	2.91
REGION	NE	-0.20	0.02	-0.25	-0.16
	MW	0.01	0.02	-0.03	0.05
	S	0.16	0.02	0.13	0.20
	W	0			
Householder's Age	<25	-0.35	0.03	-0.41	-0.29
	25-34	-0.12	0.02	-0.16	-0.08
	35-44	0			
	45-54	0.06	0.02	0.02	0.09
	55-64	0.05	0.02	0.02	0.08
	65-74	0.02	0.02	-0.03	0.06
	75+	-0.07	0.03	-0.13	-0.02
Householder's Race	White Only	0			
	Black Only	-0.57	0.02	-0.61	-0.54
	Other Race(s)	-0.14	0.03	-0.19	-0.09
Urban Status	Urban	0			
	Rural	0.21	0.01	0.19	0.24
Household Income	<30K	-0.18	0.01	-0.21	-0.16
	30<60K	0			
	60<100K	-0.04	0.02	-0.07	-0.01
	≥100K	-0.22	0.02	-0.26	-0.18

REFERENCES

Bohac, D., Hewett, M.J., Fitzgerald, J.E., Grimsrud, D. (2007) Measured change in multifamily unit air leakage and airflow due to air sealing and ventilation treatments. Thermal Performance of the Exterior Envelopes of Whole Buildings XI International Conference, December 2-7 2007

California EPA (1997) Health Effect of Exposure to Environmental Tobacco Smoke. Office of Environmental Hazard Assessment, California Environmental Protection Agency.

Chan, W.R., Nazaroff, W.W., Price, P.N., Sohn, W.D., Gadgil, A.J. (2005) Analyzing a database of residential air leakage in the United States. *Atmospheric Environment* **39**: 3445-3455.

Centers for Disease Control and Prevention (2006) The health consequences of involuntary exposure to tobacco smoke: a report of the Surgeon General. U.S. Dept. of Health and Human Services, Centers for Disease Control and Prevention, Coordinating Center for Health Promotion, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Atlanta, GA.

Centers for Disease Control and Prevention (2010). Youth Risk Behavior Surveillance – United States, 2009. Surveillance Summaries, June 4, 2010. MMWR 2010;59(No. SS-5). Available: http://www.cdc.gov/HealthyYouth/tobacco/index.htm

Chahine, T. (2010) Modeling Variability in Environmental Exposures and Health Risks for Community-based Risk Assessment. Doctoral Dissertation, Harvard University, Boston, MA

Datta, G.D., Subramanian, S.V., Colditz, G.A., Kawachi, I., Palmer, J.R., Rosenburg, L. (2006). Individual, neighborhood, and state-level predictors of smoking among U.S. Black women: A multilevel analysis. *Social Science & Medicine* **63**: 2034-1044

Diamond, R., Feustel, H., Matston, N. (1996) Energy Efficient Ventilation for Apartment Buildings. Lawrence Berkeley National Laboratory and U.S. Department of Energy, Available: http://epb.lbl.gov/publications/

Dockery, D.W., Spengler J.D. (1981) Personal exposure to respirable particulates and sulfates. *Journal of the Air Pollution Control Association* **31**(2): 153-159.

Jenkins, R.A., Maskarinec, M.P., Counts, R.W., Caton, J.E., Tomkins, B.A., Ilgner, R.H. (2001) Environmental tobacco smoke in an unrestricted smoking workplace: area and personal exposure monitoring. *Journal of Exposure Analysis and Environmental Epidemiology* **5**:369-80.

Jenkins, R.A., Palausky, A., Counts, R.W., Bayne, C.K., Dindal, A.B., Guerin, M.R. (1996) Exposure to environmental tobacco smoke in sixteen cities in the United States as determined by personal breathing zone air sampling. *Journal of Exposure Analysis and Environmental Epidemiology* 4: 473-502.

Klepeis, N.E. (1999) An introduction to the indirect exposure assessment approach: modeling human exposure using microenvironmental measurements and the recent national human activity pattern survey. *Environmental Health Perspectives* **107**(S2): 365-374.

Klepeis, N.E., Apte, M.G., Gundel, L.A., Sextro, R.G., Nazaroff, W.W. (2003) Determining size-specific emission factors for environmental tobacco smoke particles. *Aerosol Science and Technology* **37** (1), 780-790.

Klepeis, N.E., Nazaroff, W.W. (2006) Modeling residential exposure to secondhand tobacco smoke. *Atmospheric Environment* **40**:4393-4407.

Leaderer, B.P. (1990) Assessing exposures to environmental tobacco smoke. *Risk Analysis* **10**(1):19-26

Mannino, D.M., Caraballo, R., Benowitz, N., Repace, J. (2001) Predictors of Cotinine Levels in Children: Data from the Third National Health and Nutrition Examination Survey. *Chest* **120:**718-724.

Marano, C., Schober, S., Brody, D., Zhang, C. (2009) Secondhand Tobacco Smoke Exposure Among Children and Adolescents: United States, 2003–2006. *Pediatrics* **124**(5): 1299-1305.

Martin, P., Heavner, D.L., Nelson, P.R., Maiolo, K.C., Risner, C.H., Simmons, P.S., Morgan, W.T., Ogden, M.W. (1997) Environmental tobacco smoke (ETS): a market cigarette study. *Environment International* **23** (1), 75-90.

Max, W., Sung, H.Y., Shi, Y.L. (2009) Who is Exposed to Secondhand Smoke? Self-Reported and Serum Cotinine Measured Exposure in the U.S., 1999-2006. *International Journal of Environmental Research and Public Health.* **6**(5): 1633-1648.

Myatt, T.A., Minegishi, T., Allen, J.G., MacIntosh, D.L. et al (2008) Control of Asthma Triggers in Indoor Air with air cleaners: a modeling analysis. Environ Health 7:43-56.

National Cancer Insitute, U.S. National Institutes of Health (2006/7) Tobacco Use Supplement of the Current Population Survey, http://riskfactor.cancer.gov/studies/tus-cps/

Nazaroff, W.W., Singer, B.C. (2004) Inhalation of hazardous air pollutants from environmental tobacco smoke in U.S. residences. *Journal of Exposure Analysis and Environmental Epidemiology* **14**: S71-S77.

Osypuk, T.L., Kawachi, I., Subramanian, S.V., Acevedo-Garcia, D. (2006) Are State Patterns of Smoking Different for Different Racial/Ethnic Groups? An Application of Multilevel Analysis. *Public Health Reports* **121**:563-576.

Ozkaynak, H., Xue, J., Spengler, J., Wallace, L., Pellizzari, E., Jenkins, P. (1996) Personal exposure to airborne particles and metals: results from the Particle TEAM study in Riverside, California. *Journal of Exposure Analysis and Environmental Epidemiology* **6**(1); 57-78.

Price, P.N., Shehabi, A., Chan, R. 2006. *Indoor-Outdoor Air Leakage of Apartments and Commercial Buildings.* California Energy Commission, PIER Energy-Related Environmental Research Program. CEC-500-2006-111.

Rasbash, J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., et al. (2002). A user's guide to MLwiN. London: Center for Multilevel Modelling, Institute of Education, University of London.

Repace, J. (2004) Respirable particles and carcinogens in the air of Delaware hospitality venues before and after a smoking ban. *Journal of Occupational and Environmental Medicine* **46**(9): 887-905.

Repace, J.L. (2007) Exposure to Secondhand Smoke. In: Exposure Analysis. Ott, Steinemann, Wallace. Taylor and Francis Group, Boca Raton, FL. pp. 201-235.

Repace, J.L., Hyde, J.N., Brugge, D. (2006) Air Pollution in Boston bars before and after a smoking ban. *BioMed Central Public Health* **6**:266. doi:10.1186/1471-2458-6-266

Shavers, V.L., Faga, P., Jouridine Alexander, L.A., Clayton, R., Doucet, J., Bauzconde-Garbanati, L. (2006) Workplace and home smoking restrictions and racial/ethnic variation in the prevalence and intensity of current cigarette smoking among women by poverty status, TUS-CPS 1998-1999 and 2001-2002. *Journal of Epidemiology in Community Health* **60**(SupplII)ii34-ii43.

Shavers, V.L., Lawrence, D., Fagan, P., Gibson, J.T. (2005) Racial/ethnic variation in cigarette smoking among the civilian U.S. population by occupation and industry, TUS-CPS 1998-1999. *Preventive Medicine* **41**:597-606

U.S. Census Bureau (2000) Geographic Areas Reference Manual. Available: http://www.census.gov/geo/www/cen_tract.html

U.S. EPA (1992) Respiratory health effects of passive smoking lung cancer and other disorders. Office of Research and Development, Office of Health and Environmental Assessment, [EPA/600/6-90/006F]

U.S. EPA, Office of Air and Radiation, Air Quality Trends (2008). Available: http://www.epa.gov/airtrends/pm.html

U.S. Department of Housing and Urban Development, Office of Policy Development and Research (2009) Codebook for the American Housing Survey, Public Use File: 1997 and later. Available: http://www.huduser.org/Datasets/ahs/AHS_Codebook.pdf

Xu, M. D., Nematollahi, M., Sextro, R. G., Gadgil, A. J., Nazaroff, W. W. (1994) Deposition of tobacco smoke particles in a low ventilation room. *Aerosol Science and Technology* **20** (2), 194.206.

Zartarian, V.G., Schultz, B.D. (2010) EPA Human Exposure Research on Assessing Cumulative Risk in Communities, *Journal of Exposure Science and Environmental Epidemiology*. **20**(4): 351-358.