Predicting residential indoor concentrations of nitrogen dioxide, fine particulate matter, and elemental carbon using questionnaire and geographic information system based data

Lisa K. Baxter*[1], Jane E. Clougherty[1], Christopher J. Paciorek[2], Rosalind J. Wright[3], and Jonathan I. Levy[1]

*Corresponding Author

Phone: 617-384-8528

FAX: 617-384-8859

Email: lbaxter@hsph.harvard.edu

[1]Harvard School of Public Health, Department of Environmental Health, Landmark Center-4th Floor West, P.O. Box 15677, Boston, MA 02215, USA

[2]Havard School of Public Health, Department of Biostatistics, 655 Huntington Avenue, SPH2-4th Floor, Boston, MA 02115, USA

[3]Channing Laboratory, Brigham and Women's Hospital, Department of Medicine, Harvard Medical School, 181 Longwood Ave., Boston, MA 02115, USA

Abstract

Previous studies have identified associations between traffic-related air pollution and adverse health effects. Most have used measurements from a few central ambient monitors and/or some measure of traffic as indicators of exposure, disregarding spatial variability and/or factors influencing personal exposure-ambient concentration

relationships. This study seeks to utilize publicly available data (i.e., central site monitors, geographic information system (GIS), and property assessment data) and questionnaire responses to predict residential indoor concentrations of traffic-related air pollutants for lower socioeconomic status (SES) urban households.

As part of a prospective birth cohort study in urban Boston, we collected indoor and outdoor 3-4 day samples of nitrogen dioxide (NO_2) and fine particulate matter ($PM_{2.5}$) in 43 low SES residences across multiple seasons from 2003 – 2005. Elemental carbon concentrations were determined via reflectance analysis. Multiple traffic indicators were derived using Massachusetts Highway Department data and traffic counts collected outside sampling homes. Home characteristics and occupant behaviors were collected via a standardized questionnaire. Additional housing information was collected through property tax records, and ambient concentrations were collected from a centrally-located ambient monitor.

The contributions of ambient concentrations, local traffic and indoor sources to indoor concentrations were quantified with regression analyses. $PM_{2.5}$ was influenced less by local traffic but had significant indoor sources, while EC was associated with traffic and NO_2 with both traffic and indoor sources. Comparing models based on covariate selection using p-values or a Bayesian approach yielded similar results, with traffic density within a 50m buffer of a home and distance from a truck route as important contributors to indoor levels of NO_2 and EC, respectively. The Bayesian approach also highlighted the uncertanity in the models. We conclude that by utilizing public databases and focused questionnaire data we can identify important predictors of indoor concentrations for multiple air pollutants in a high-risk population.

Keywords: indoor air; NO_2; $PM_{2.5}$; EC; geographic information system

1. Introduction

Numerous studies have identified associations between traffic-related air pollution and adverse heath effects either by characterizing exposures to specific pollutants using measurements from a few central ambient sites (Dockery et al. 1993; Pope et al. 1995; Studnicka et al. 1997; Laden et al. 2000), or by some measure of traffic (Oosterlee et al. 1996; Garshick et al. 2003; Heinrich et al. 2005; Ryan et al. 2005). Yet, by ignoring the contribution of indoor sources and the effect of residential ventilation, it is difficult to accurately estimate personal exposures, especially in an intraurban epidemiological study. Residential indoor concentrations are a product of ambient-generated pollution that has infiltrated indoors and indoor-generated pollution, and are strongly correlated with personal exposures (Levy et al. 1998; Koistinen et al. 2001; Kousa et al. 2001; Brown 2006). However, it is often impractical to obtain direct indoor measurements (or personal exposure measurements) for all participants in a large epidemiological study, raising the question of how personal exposures can be best estimated. Given the logistical constraints, utilizing public databases and focused questionnaires may be the best approach to reasonably estimate indoor and therefore personal exposures.

In lieu of using home-specific outdoor measurements to determine ambient-generated pollutant exposures (which would be nearly as labor-intensive as indoor monitoring), factors generated from Geographic Information Systems (GIS), such as distance from road, population density, and land use can be used in combination with central site monitoring data to estimate ambient exposures (Briggs et al. 1997; Brauer et al. 2003). Questionnaire (e.g., opening of windows, air conditioning usage) and/or property assessment data on individual building characteristics can then be used to estimate residential ventilation patterns (Long et al. 2001; Setton et al. 2005) that potentially affect the influence of ambient concentrations and indoor sources (Abt et al. 2000). Similarly, questionnaire data on exposure-related activities can be used to predict indoor sources.

The current study seeks to utilize publicly available data (i.e., central site monitors, GIS, and property assessment data) and questionnaire responses to predict residential indoor concentrations of traffic-related air pollutants for lower socioeconomic status (SES) households in an urban area. Lower SES urban residents have been previously identified as a high risk population for asthma (The American Lung Association 2001) and often live in smaller apartments, possibly resulting in greater contributions from indoor sources (given smaller volumes and higher occupant densities), traffic (nearer to busier roads), and different ventilation patterns (given adjoining units and lack of central air conditioning). We will build upon previously developed predictive models identifying important indoor source terms in this population (Baxter et al. in press), and home characteristics and occupant behaviors associated with infiltration (Baxter et al. 2006). We hypothesize that GIS variables addressing traffic volume and composition will be more predictive of indoor levels for pollutants with more spatial heterogeneity and fewer indoor sources, such as elemental carbon (EC), relative to those with less spatial heterogeneity (fine particulate matter, $PM_{2.5}$) or those with indoor sources ($PM_{2.5}$ and nitrogen dioxide, NO_2).

2. Methods

2.1 Data Collection

Study design, sampling, analysis, and quality control measures are described in a previous publication (Baxter et al. in press). Briefly, residential indoor and outdoor $PM_{2.5}$ and NO_2 samples and home characteristics/occupant behavior data were collected at 43 homes from 2003 - 2005 in the metropolitan Boston area as part of the Asthma Coalition for Community, Environment, and Social Stress (ACCESS) study, a prospective birth cohort assessing asthma etiology in a lower SES population. Sampling was conducted in two seasons, the non-heating (May – October) and heating season (December – March). When possible, two consecutive 3-4 day measurements were collected in each season; all analyses were based on the average of within-season measurements. $PM_{2.5}$ samples were collected with Harvard Personal Environmental Monitors (PEM) on Teflon filters, and analyzed for EC using reflectance analysis. NO_2 concentrations were measured using Yanagisawa passive filter badges. A standardized questionnaire was administered at the end of each sampling period to gather housing characteristics/occupant behavior data. The study was approved by the Human Studies Committee at the Brigham & Women's Hospital and the Harvard School of Public Health.

Information on housing characteristics was also collected through the City of Boston, Brookline, Cambridge, and Somerville property tax records, and ambient concentrations were collected from an ambient monitor (the Massachusetts Department of Environmental Protection monitor in Dudley Square, Roxbury) located near the center of our monitoring area. Ambient concentrations were averaged over the same sampling period (matching date and time) as when the indoor and outdoor samples were collected. Finally, continuous traffic counts were recorded on the largest road within 100m of the home with a Jamar Trax I Plus traffic counter.

Sample homes were individually geocoded with ArcGIS 9.1 using U.S. Census TIGRE files and City of Boston street parcels data, and combined with road networks and traffic data obtained from the Massachusetts Highway Department (MHD) to create various measures of traffic. Because different aspects of traffic (e.g. density, roadway configuration, vehicle speed) may affect overall emission rates, pollutant mix, and dispersal, we created and examined a number of traffic indicators to capture varying characteristics, including cumulative traffic density scores (unweighted and kernel-weighted) at various radii (50-500m), distance-based measures, total roadway length measures, and characteristics of traffic on the nearest major road to each home. To consider the influence of the nearest major road, we created indicators for its average daily traffic, diesel traffic (using axle length from ACCESS traffic measurements), and weighted each by distance to the road. Lastly, block group-level population and area measures were used to estimate population density (Clougherty 2006).

2.2. Data Analysis

2.2.1 Regression Models

Models utilizing publicly available data and questionnaire responses were developed by regressing ambient concentrations, predetermined indoor source terms, and traffic indicators against indoor concentrations as seen in Equation (1).

$$Cin_{ij} = \beta_{oj} + \beta_{1j} * Cambient_{ij} + \beta_{2j} * Q_{ij} + \beta_{3j} * Traffic_{ij} \quad (1)$$

where Cin_{ij} (ppb, µg/m^3, or m^{-1} x 10^{-5}) is the indoor concentration of pollutant *j* for sampling session *i*, $Cambient_{ij}$ is the concentration collected from the ambient monitors, Q_{ij} is a vector of the various indoor source terms, and $Traffic_{ij}$ represents the different traffic indicators created for each home and then selected by pollutant. The indoor source terms were determined from a previous analysis where home-specific outdoor concentrations and exposure-related activities, collected via questionnaire, were regressed against home-specific indoor concentrations. The indoor source terms were as follows: for $PM_{2.5}$, cooking time (≤ 1/day vs. > 1h.day) and occupant density (people/room); for NO_2, gas stove usage (using an electric stove or a gas stove ≤1 h/day vs. using a gas stove >1h/day); and for EC, no indoor sources were identified (Baxter et al. in press). We restricted our modeling to these terms, for the sake of comparability and to minimize the likelihood of spurious findings. The best model was then selected based on the lowest p-values for the traffic term.

Although many homes had two sampling sessions, conducted in two different seasons (a heating and non-heating season), these were broadly defined and covered a period up to 6 months. Therefore, each sampling session was treated as an independent measurement. In all regression models, outliers were removed that unduly influenced regression results, defined as having an absolute studentized residual greater than four. One outlier was removed for $PM_{2.5}$ and two were removed for EC.

2.2.2. Bayesian Variable Selection

With 24 traffic variables and a small dataset, there may be issues with comparing models using p-values, both because multiple variables may have similar significance levels and because the observed relationships may be due to chance. For a more formal model comparison, a Bayesian approach was used to estimate the probability that a model using a given traffic covariate is the best model. This approach allowed us to weigh the evidence for each traffic term and see the amount of uncertainty in choosing the best model. The posterior model probabilities for each pollutant are shown by Equations (2) – (4) (George and McCulloch 1997; Chipman et al. 2001).

$$P(M_k|Y) \propto l(Y|M_k) * P(M_k) \tag{2}$$

where M_k is the model with traffic term k when all of the other variables (e.g. ambient concentrations, indoor sources) are in the model, Y is the observed indoor concentrations for one of the pollutants, $P(M_k|Y)$ is the posterior model probability of M_k given Y, $l(Y|M_k)$ is the marginal likelihood of Y given M_k, $P(M_k)$ is the prior probability that M_k is the true model. We assumed the same prior probability $P(M_k)$ for all of the traffic terms, equal to $1/N$ (N = the number of traffic terms).

The marginal likelihood is the likelihood of the observed data under M_k accounting for the uncertainty in the regression coefficients as shown in Equation (3).

$$l(Y|M_k) = \sqrt{\frac{1}{c+1}} * \frac{1}{\left(\sum_{i=1}^{k} Y_i^2 - \frac{\left(\sum_{i=1}^{n} X_{ik} Y_i \right)^2}{\left(1 + \frac{1}{c} \right) * \sum_{i=1}^{n} X_{ik}^2} \right)^{\frac{n}{2}}} \tag{3}$$

where Y_i is the residual from sampling session i from regressing indoor concentrations on ambient concentrations and indoor source terms, X_{ik} is the residual from regressing traffic term k on ambient concentrations and indoor source terms, n is the number of observations, and c reflects our prior uncertainty on the regression coefficients of the traffic terms in $Y_j|M_k$. We used $c = n$, making c large enough to acknowledge reasonable uncertainty in the effect estimates while still giving very unlikely effect estimates low prior probability. We also conducted sensitivity analysis by calculating the posterior probabilities with a range of c 's (5 -100) (Chipman et al. 2001).

The probabilities then need to be normalized as shown in Equation (4) (multiplied by 100 to calculate a percentage).

$$P(M_k|Y) = \frac{P(M_k|Y)}{\sum_{i=1}^{N} P(M_k|Y)_i} * 100 \tag{4}$$

In a sensitivity analysis, we considered another model where M_0 is the model without a traffic term. We assumed a $P(M_k)$ of $\frac{1}{2}$ and $\frac{1}{2N}$ for M_0 and M_k (models with the traffic term), respectively. This assumed an equal chance of traffic affecting indoor concentrations as not. Using the $\frac{1}{2N}$ weights in the model selection inherently penalized for testing many traffic terms in a small dataset. The posterior probabilities of M_0 for each pollutant were calculated as shown by Equation (5) and normalized utilizing Equation (4).

$$P(M_k|Y) \propto l(Y|M_k) * P(M_k)$$
$$\propto \frac{1}{\left(\sum_{i=1}^{k} Y_i^2\right)^{\frac{n}{2}}} * P(M_k) \qquad (5)$$

2.2.3 Effect Modification by Ventilation Characteristics

The model expressed in Equation (1) does not account for variations in home ventilation patterns which may influence the effect of indoor sources, local traffic, and ambient concentrations. In this study there are no direct measurements of air exchange rates (AERs), so we relied on other methods to capture the effects of ventilation. Prior studies conducted in Boston area homes observed a strong relationship between the infiltration factor (F_{INF}) and AER (Sarnat et al. 2002; Long and Sarnat 2004). In a previous analysis, we described home ventilation characteristics using F_{INF} estimated by the indoor-outdoor sulfur ratio, and then estimated the contribution of season, home characteristics (e.g. year of construction, apartment vs. multi-family home, and floor level), and occupant behaviors (e.g. open windows and air conditioner use). We predicted F_{INF} using logistic regression, dichotomizing F_{INF} at the median into high and low categories, and found open windows to be the most significant contributor in our dataset (Baxter et al. 2006).

The variable of open windows (no vs. yes) was therefore used as a readily available proxy for the infiltration factor and was incorporated as an interaction term into the model illustrated in Equation (1). This can be expressed as:

$$Cin_{ij} = \beta_{oj} + \beta_{1j} * Cambient_{ij} * Openwindows_i + \beta_{2j} * Q_{ij} * Openwindows_i + \beta_{3j} * Traffic_i * Openwindows_i \quad (6)$$

where $Openwindows_i$ indicates whether during the sampling period the occupant had their windows open or closed. Adhering to the mass balance framework, the opening of windows should theoretically increase the influence of ambient concentrations and traffic while decreasing the influence of indoor sources. All analyses were done using SAS version 8.

3. Results and Discussion

3.1 Data Analysis

3.1.1. General Characteristics

A total of 66 sampling sessions were conducted. The 43 sites (shown in Figure 1) were distributed among 39 households throughout urban Boston, with 4 participants moving and allowing us to sample in their new home. Summary statistics of NO_2, $PM_{2.5}$, and EC for indoor, outdoor, and ambient concentrations (collected from a centrally located monitor) are presented in Table 1 and are comparable to those seen in other studies (Zipprich et al. 2002; Brunekreef et al. 2005; Meng et al. 2005; Brown 2006). Average indoor concentrations of NO_2 and $PM_{2.5}$ are greater than both home-specific outdoor and ambient concentrations while indoor concentrations of EC were less than both outdoor and ambient concentrations. For EC, ambient concentrations are in mass-based units while the absorption coefficient is used for the indoor and outdoor concentrations. For the sake of comparison, a conversion factor of 0.83 $\mu g/m^3$ per m^{-1} x

10^{-5} (Kinney et al. 2000) was used on the indoor and home-specific outdoor concentrations.

We regressed indoor concentrations on outdoor concentrations, indoor on ambient, and outdoor on ambient, to help determine the likely predictors of indoor concentrations (Table 2). For our outdoor concentrations, the ambient monitor was strongly predictive for $PM_{2.5}$, but not for NO_2 or EC. This indicates that temporal rather than small-scale spatial variability was dominant for $PM_{2.5}$, whereas for NO_2 and EC, there was more pronounced spatial variability and more influential local sources, such as local traffic conditions. The coefficients of determination (R^2) for indoor vs. outdoor and indoor vs. ambient are similar to one another for NO_2 and $PM_{2.5}$, however, outdoor and ambient concentrations did not explain the majority of variability seen in indoor concentrations, possibly due to the influences of indoor sources. For EC, the R^2s were quite different, with outdoor concentrations explaining a large portion of the variability whereas ambient concentrations did not due to the influence of local traffic.

3.1.2 Regression Models

Variables and regression coefficients of the regression models with the most significant traffic terms are shown in Table 3. The unweighted cumulative density score within 50 m of the home was associated with an increase in indoor NO_2 levels. For EC, a proxy for diesel traffic appeared to be predictive of indoor concentrations, with levels decreasing as the distance a home is from a designated truck route increases. No traffic variable was significantly associated with indoor $PM_{2.5}$ concentrations.

3.1.3. Bayesian Variable Selection

For each pollutant, the posterior probabilities of models using the different traffic variables were calculated and grouped based on the GIS algorithm used to create them (Table 4). Posterior probabilities greater than three times the prior probability (4.2%) included the unweighted cumulative density score within a 50m buffer, which yielded the highest probability (26.5%) for NO_2, and distance from a designated truck route (14.3%) for EC. Average daily traffic (ADT) had the highest posterior probability in the $PM_{2.5}$ models (8.3%), but was less than twice the prior probability, and multiple additional measures had comparable probabilities. We calculated these posterior probabilities using a range of *c*'s (5-100) and the results were similar (not shown).

Within the Bayesian analysis, all posterior probabilities were under 30%, emphasizing the difficulty in choosing the correct model with a small dataset and many correlated predictors. For NO_2, models describing traffic closer to the home (50 -100m buffers) generally had the highest probabilities. This agrees with previous studies showing outdoor NO_2 levels decreasing significantly with increasing logarithmic distance from the road (Roorda-Knape et al. 1999; Gilbert et al. 2003), and the majority of air pollution from the road occurring within 50-75m (Van Roosbroeck et al. 2006). Therefore roadways within 50m of the home may be the largest contributor to the total NO_2 concentration.

For EC, the highest probability traffic terms were related to truck traffic. EC has commonly been used as a marker for diesel particles (Gotschi et al. 2002) and since almost all heavy-duty trucks have diesel engines, it is expected that a traffic indicator summarizing truck traffic would be important, especially in the United States where relatively few passenger vehicles use diesel fuel. In contrast to the other pollutants, the traffic model with the highest probability (ADT) was not significant in the indoor $PM_{2.5}$ model. None of the models yielded probabilities over 10%, suggesting little differential information value across covariates and therefore that a traffic variable may not be necessary in the model. This was not entirely unexpected given that $PM_{2.5}$ exhibits less spatial heterogeneity than the other pollutants (Roorda -Knape et al. 1998).

To address the issue of multiple testing, sensitivity analyses calculated the posterior probabilities for pollutant models with (M_k) and without a traffic term (M_0) assuming an equal chance of traffic affecting indoor pollutant concentrations as not. For all of the pollutants, the models without the traffic term had high probabilities, with 77.3% for NO_2, 84.3% for $PM_{2.5}$, and 84.6% for EC, reflecting both the presumed prior probabilities and the relatively small amount of variability explained by the traffic terms. The highest probabilities for those models with the traffic term were 6.02% (unweighted cumulative density score within a 50m buffer) for NO_2, 1.31% (ADT) for $PM_{2.5}$, and 2.21% (distance from a designated truck route) for EC. This suggests the difficulty in relating traffic variables to indoor concentrations given less spatial variation across an urban area as opposed to comparing an urban vs. suburban/rural area, as well as the contribution of indoor sources and ventilation. The small sample sizes and multiple testing also contribute to the difficulty of definitively demonstrating that traffic terms should be in the model.

3.1.4 Effect Modification by Ventilation Characteristics

The use of open windows as a ventilation proxy agrees with a similar study conducted in Boston which found air exchange rates (AER) higher in homes with open windows, and that an open windows covariate may be a better estimate of air exchange with outdoors than measured AERs for multi-unit buildings, such as those seen in the current study. This is because measured AERs cannot distinguish between make-up air from adjacent apartments and the air from the outdoors (Brown 2006). The term *openwindows* served as a proxy for 'high' and 'low' infiltration factors and is used as an effect modifier as described by Equation (5). This was done without modifying the effect of indoor sources due to the limited statistical power and resulting statistical instability when effect modification of indoor sources was included (related in part to the use of categorical variables for many indoor source terms). The final models, including only the significant ($p < 0.2$) interaction terms, are shown in Table 5. For NO_2 and EC, the traffic variables were significantly modified by the open windows variable, with their effects on indoor levels more pronounced in homes where windows were opened. For $PM_{2.5}$, the effect of ambient concentrations was significantly greater in home where windows were opened compared to those where windows were kept closed. The inclusion of this term increased the R^2 from 0.02 to 0.25 for NO_2, 0.20 to 0.40 for $PM_{2.5}$, and 0.16 to 0.32 for EC.

3.2. Contribution of indoor and outdoor sources to indoor concentrations

It is also important to understand whether indoor or outdoor sources appear to contribute more to indoor concentrations. We therefore calculated the contributions due to local traffic and indoor sources for NO_2, of traffic on EC, and of ambient concentrations and indoor sources on $PM_{2.5}$. For NO_2, the contribution of local traffic, given a range of cumulative unweighted density traffic scores (within 50m buffer) from 4.1-198 vehicles*m, was approximately 0.29 ppb – 14 ppb for homes with open windows, with no significant contribution to homes with closed windows. This is comparable to a study conducted in the Netherlands which reported a difference of about 7 ppb in average classroom concentrations comparing schools in high urbanization areas to schools in low urbanization areas (Rjinders et al. 2001). Gas stove usage contributed on average 7 ppb to indoor NO_2 levels, similar in magnitude as observed in previous studies (Lee et al. 1998; Levy et al. 1998). Thus, local traffic is a larger contributor to indoor NO_2 where traffic density is high and windows are opened, whereas indoor sources are a larger contributor when traffic density is low or windows are closed.

Similarly, traffic contributed up to 0.2 $\mu g/m^3$ to indoor EC for homes with open windows, with an insignificant contribution for homes where windows were closed. Previous studies have found EC concentrations to be 50% higher in homes located on high intensity streets compared to low traffic homes (Fischer et al. 2000). In addition, indoor EC increased 1.91 $\mu g/m^3$ with increasing truck traffic density (Janssen et al. 2001), although in a European setting with greater prevalence of diesel vehicles.

Ambient concentrations contributed an average of 15 $\mu g/m^3$ to indoor $PM_{2.5}$ for homes with open windows, and 10 $\mu g/m^3$ for homes where windows were closed. Additionally, cooking for more than an hour per day contributed 6.2 $\mu g/m^3$ and average occupant density contributed 6.5 $\mu g/m^3$. The effect of cooking is comparable to results from prior studies (Ozkaynak et al. 1994; Brunekreef et al. 2005). Occupant density is likely a proxy for multiple factors, including resuspension activities. Resuspension has not been as substantial of a contributor in previous studies, although the smaller volumes and greater crowding of our study homes may increase the relative source strength.

Finally, in a previous paper we predicted indoor concentrations using home-specific outdoor concentrations and indoor sources (Baxter et al. in press). For $PM_{2.5}$ and NO_2 the predictive power of the models (R^2 of 0.37 and 0.16, respectively) are similar to those seen in the current analysis. This was expected given the large influence of indoor sources to indoor levels of these pollutants. In contrast, for EC, the predictive power of the model from the current analysis ($R^2 = 0.32$) was weaker than seen in the previous analysis ($R^2 = 0.49$). EC tends to be dominated by outdoor sources; it is therefore more important to accurately capture its outdoor spatial pattern wherein our traffic indicators may not be adequate.

3.3 Limitations

The ambient monitor is located within the city and may be influenced by local traffic. It also uses different measurement methods for EC, possibly explaining both model performance and the higher ambient concentrations relative to outdoor. However, the Dudley Square monitor includes all three pollutants, is at the center of our monitoring region, and is well correlated with other ambient monitors in and around Boston. The sample size also limited our ability to explore a larger range of potential indoor source terms and traffic variables. Deficiencies in the underlying data, with traffic counts on smaller residential roads sparse, led to increased uncertainties for these variables in that they may be imperfect proxies of traffic volume/composition. In addition, many of these indicators do not capture the characteristics of traffic that are relevant to concentrations of different pollutants. For example, dense stop-and-go traffic may create more emissions per vehicle-mile, and total traffic counts fail to capture such aspects. For this reason a variety of traffic indicators were created to capture these different effects as well as those not dependent on total traffic counts (e.g. road segment lengths).

Additionally, the open windows variable may not effectively capture a home's ventilation characteristics in that it is used as proxy for the sulfur indoor/outdoor ratio which itself is a proxy of the infiltration factor. Similarly, the indoor source terms are developed from questionnaires which are surrogates for the source emissions rate and may represent a variety of occupant activities. However, these limitations are inherent in developing exposure estimates based on publicly available or questionnaire data.

Due to limited statistical power we also were not able to incorporate the interaction term on the indoor sources, omitting the effect of ventilation on the indoor source contribution. Finally, while it may have been desirable to develop season-specific models given the inherent seasonality in many factors, we did not have adequate power to construct those models. While it is apparent that many limitations are related to statistical power, it is often difficult to generate a large exposure dataset in an epidemiological context, so many of these issues would need to be confronted by other investigators. More importantly, despite the aforementioned limitations and sample size issues, the models are generally interpretable and in agreement with the literature.

4. Summary and Conclusions

The current paper identified important predictors of indoor concentrations for multiple air pollutants in a high-risk population, by utilizing public databases (e.g.

ambient monitor, GIS, tax assessment databases) and focused questionnaire data. Given the numerous ways to characterize traffic, the use of a Bayesian variable selection approach helped us better determine the appropriate traffic measures for each pollutant. Our regression models indicate that $PM_{2.5}$ was influenced less by local traffic but had significant indoor sources, while EC was associated with local traffic and NO_2 was associated with both traffic and indoor sources. Comparing models based on p-values and using a Bayesian approach yielded similar results, with traffic density/volume within a 50m buffer of a home and distance from a designated truck route as important contributors to indoor levels of NO_2 and EC, respectively. However, results from the Bayesian approach also suggested a high degree of uncertainty in selecting the best model. We also found additional information value in the variable capturing the opening of windows, previously shown to be associated with ventilation, which allowed our model to keep with the principles of the mass balance model.

In general, our study provides some direction regarding how publicly available data can be utilized in population studies, in order to predict residential indoor (and therefore personal) exposures in the absence of measurements. We have demonstrated that information on traffic applied in GIS framework in combination with ambient monitoring data can be used as an effective substitute for home-specific outdoor measurements. Along with some type of evaluation of the ventilation characteristics of the home, the aforementioned information can be used to estimate indoor exposures of outdoor dominated pollutants (e.g., EC). For those pollutants with significant indoor sources (e.g. NO_2 and $PM_{2.5}$) questionnaire data capturing these sources is also needed.

Acknowledgments

This research was supported by HEI 4727-RFA04-5/05-1, NIH U01 HL072494, NIH R03 ES013988, and PHS 5 T42 CCT122961-02. We gratefully acknowledge the hard work of all the technicians associated with the ACCESS project and the hospitality of the ACCESS and other study participants. In addition, we thank Francine Laden from the Department of Environmental Health at Harvard School of Public Health and Channing Laboratory at Brigham and Women's Hospital, and Helen Suh from the Department of Environmental Health at Harvard School of Public Health for providing guidance; Prashant Dilwali, Robin Dodson, Shakira Franco, Lu-wei Lee, Rebecca Schildkret, and Leonard Zwack for their sampling assistance; and Monique Perron for both her sampling and laboratory assistance.

Abt E., Suh H.H., Allen G. and Koutrakis P., 2000. Characterization of indoor particle sources: a study conducted in the metropolitan Boston area. Environmental Health Perspectives 108, 35-44.

Baxter L.K., Clougherty J.E., Laden F. and Levy J.I., in press. Predictors of concentrations of nitrogen dioxide, fine particulate matter, and particle constituents inside of lower socioeconomic status urban homes. Journal of Exposure Science and Environmental Epidemiology.

Baxter L.K., Suh H.H., Paciorek C.J., Clougherty J.E. and Levy J.I. (2006). Predicting infiltration factors in urban residences for a cohort study. Healthy Building 2006, Lisboa, International Society of Indoor Air Quality and Climate.

Brauer M., Hoek G., van Vliet P., Meliefste K., Fischer P., Gehring U., Heinrich J., Cyrys J., Bellander T., Lewne M. and Brunekreef B., 2003. Estimating long-term average particulate air pollution concentrations: application of traffic indicators and geographic information systems. Epidemiology 14, 228-239.

Briggs D.J., Collins S., Elliot P., Fischer P., Kingham S., Lebret E., Pryl K., van Reeuwijk H., Smallbone K. and Van der veen A., 1997. Mapping urban pollution using GIS: a regression based approach. International Journal of Geographical Information Science 11, 699-718.

Brown K.W., 2006. Characterization of particulate and gaseous exposure of sensitive populations living in Baltimore and Boston. Department of Environmental Health. Boston, MA, Harvard School of Public Health.

Brunekreef B., Janssen N.A.H., de Hartog J.J., Oldenwening M., Meliefste K., Hoek G., Lanki T., Timonen K.L., Vallius M., Pekkanen J. and Van Grieken R., 2005. Personal, indoor, and outdoor exposures to $PM_{2.5}$ and its components for groups of cardiovascular patients in Amsterdam and Helsinki. Boston, MA, Health Effects Institute.

Chipman H., George E.I. and McCulloch R.E.,2001. The practical implementation of Bayesian model selection. P. Lahiri (Eds.),Model Selection, Institute of Mathematical Statistics Lecture Notes, 65-116.

Clougherty J.E., 2006. Environmental and social determinants of childhood asthma. Department of Environmental Health. Boston, MA, Harvard School of Public Health.

Dockery D.W., Pope C.A., Xu X., Spengler J.D., Ware J.H., Fay M.E., Ferris B.G. and Speizer F.A., 1993. An association between air pollution and mortality in six U.S. cities. New England Journal of Medicine 329, 1753-1759.

Fischer P.H., Hoek G., van Reeuwijk H., Briggs D.J., Lebret E., van Wijnen J.H., Kingham S. and Elliott P.E., 2000. Traffic-related differences in outdoor and indoor concentrations of particles and volatile organic compounds in Amsterdam. Atmospheric Environment 34, 3713-3722.

Garshick E., Laden F., Hart J.E. and Caron A., 2003. Residence near a major road and respiratory symptoms in U.S. veterans. Epidemiology 14, 728-736.

George E.I. and McCulloch R.E., 1997. Approaches for Bayesian Variable Selection. Statistica Sinica 7, 339-374.

Gilbert N.L., Woodhouse S., Stieb D.M. and Brook J.R., 2003. Ambient nitrogen dioxide and distance from a major highway. The Science of the Total Environment 312, 43-46.

Gotschi T., Oglesby L., Mathys P., Monn C., Manalis N., Koistinen K., Jantunen M., Hanninen O., Polanska L. and Kunzli N., 2002. Comparison of black smoke and $PM_{2.5}$ levels in indoor and outdoor environments of four European cities. Environmental Science and Technology 36, 1191-1197.

Heinrich J., Topp R., Gehring U. and Thefeld W., 2005. Traffic at residential address, respiratory health, and atopy in adults: the National German Health Survey 1998. Environmental Research 98, 240-249.

Janssen N.A.H., van Vliet P.H.N., Aarts F., Harssema H. and Brunekreef B., 2001. Assessment of exposure to traffic related air pollution of children attending schools near motorways. Atmospheric Environment 35, 3875-3884.

Kinney P.L., Aggarwal M., Northridge M.E., Janssen N.A.H. and Shepard P., 2000. Airborne concentrations of $PM_{2.5}$ and diesel exhaust particles on Harlem sidewalks: a community-based pilot study. Environmental Health Perspectives 108, 213-218.

Koistinen K.J., Hanninen O., Rotko T., Edwards R.D., Moschandreas D. and Jantunen M.J., 2001. Behavioral and environmental determinants of personal exposures to $PM_{2.5}$ in EXPOLIS - Helsinki, Finland. Atmospheric Environment 35, 2473-2481.

Kousa A., Monn C., Rotko T., Alm S., Oglesby L. and Jantunen M.J., 2001. Personal exposures to NO_2 in the EXPOLIS-study: relation to residential indoor, outdoor, and workplace concentrations in Basel, Helsinki, and Prague. Atmospheric Environment 35, 3405-3412.

Laden F., Neas L.M., Dockery D.W. and Schwartz J., 2000. Association of fine particulate matter from different sources with daily mortality in six U.S. cities. Environmental Health Perspectives 108, 941-947.

Lee K., Levy J.I., Yanagisawa Y. and Spengler J.D., 1998. The Boston residential nitrogen dioxide characterization study: classification of and prediction of on indoor NO_2 exposure. Journal of the Air and Waste Management Association 48, 739-742.

Levy J.I., Lee K., Spengler J.D. and Yanagisawa Y., 1998. Impact of residential nitrogen dioxide exposure on personal exposure: an international study. Journal of the Air and Waste Management Association 48, 553-560.

Long C.M. and Sarnat J.A., 2004. Indoor-outdoor relationships and infiltration behavior of elemental components of outdoor $PM_{2.5}$ for Boston-area homes. Aerosol Science and Technology 38, 91-104.

Long C.M., Suh H.H., Catalano P.J. and Koutrakis P., 2001. Using time- and size-resolved particulate data to quantify indoor penetration and deposition behavior. Environmental Science and Technology 35, 2089-2099.

Meng Q.Y., Turpin B.J., Korn L., Weisel C.P., Morandi M., Colome S., Zhang J., Stock T., Spektor D., Winer A., Zhang L., Lee J.H., Giovanetti R., Cui W., Kwon J., Alimokhtari S., Shendell D., Jones J., Farrar C. and Maberti S., 2005. Influence of ambient (outdoor) sources on residential indoor and personal $PM_{2.5}$ concentrations: analysis of RIOPA data. Journal of Exposure Analysis and Environmental Epidemiology 15.

Oosterlee A., Drijver M., Lebret E. and Brunekreef B., 1996. Chronic symptoms in children and adults living along streets with high traffic density. Occupational and Environmental Medicine 53, 241-247.

Ozkaynak H., Xue J., Weker R., Butler D., Koutrakis P. and Spengler J., 1994. The particle TEAM (PTEAM) study: analysis of the data. final report. Vol. III. Research Triangle Park, NC, U.S. EPA.

Pope C.A., Thun M.J., Namboordiri M.M., Dockery D.W., Evans J.S., Speizer F.E. and Heath C.W., 1995. Particulate air pollution as a predictor of mortality in a prospective study of U.S. adults. American Journal of Respiratory and Critical Care Medicine 151, 669-674.

Rjinders E., Janssen N.A.H., van Vilet P.H.N. and Brunekreef B., 2001. Personal and outdoor nitrogen dioxide concentrations in relation to degree of urbanization and traffic density. Environmental Health Perspectives 109, 411-417.

Roorda -Knape M.C., Janssen N.A.H., de Hartog J.J., Van Vliet P.H.N., Harssema H. and Brunkreef B., 1998. Air pollution from traffic in city districts near major motorways. Atmospheric Environment 32, 1921-1930.

Roorda-Knape M.C., Janssen N.A.H., de Hartog J., Van Vliet P.H.N., Harssema H. and Brunekreef B., 1999. Traffic related air pollution in city districts near motorways. The Science of the Total Environment 235, 339-341.

Ryan P.H., LeMasters G., Biagini J., Bernstein D., Grinshpun S.A., Shukla R., Wilson K., Villareal M., Burkle J. and Lockey J., 2005. Is it traffic type, volume, or distance? Wheezing in infants living near truck and bus traffic. Journal of Allergy and Clinical Immunology 116, 279-284.

Sarnat J.A., Long C.M., Koutrakis P., Coull B.A., Schwartz J. and Suh H.H., 2002. Using sulfur as a tracer of outdoor fine particulate matter. Environmental Science and Technology 36, 5305-5314.

Setton E., Hystad P. and Keller C.P., 2005. Opportunities for using spatial property assessment data in air pollution exposure assessments. International Journal of Health Geographics 4, 26.

Studnicka M., Hackl E., Pischinger J., Fangmeyer C., Haschke N., Kuhr J., Urbanek R., Neumann M. and Frischer T., 1997. Traffic-related NO_2 and the prevalence of asthma and respiratory symptoms in seven year olds. The European Respiratory Journal 10, 2275-2278.

The American Lung Association, 2001. Urban air pollution and health inequities: a workshop report. Environmental Health Perspectives 109, 357-374.

Van Roosbroeck S., Wichmann J., Janssen N.A.H., Hoek G., van Wijnen J.H., Lebret E. and Brunekreef B., 2006. Long-term personal exposure to traffic-related air pollution among school children, a validation study. The Science of the Total Environment 368, 565-573.

Zipprich J.L., Harris S.A., Fox C. and Borzelleca J.F., 2002. An analysis of factors that influence personal exposure to nitrogen dioxides in residents of Richmond, Virginia. Journal of Exposure Analysis and Environmental Epidemiology 12, 273-285.

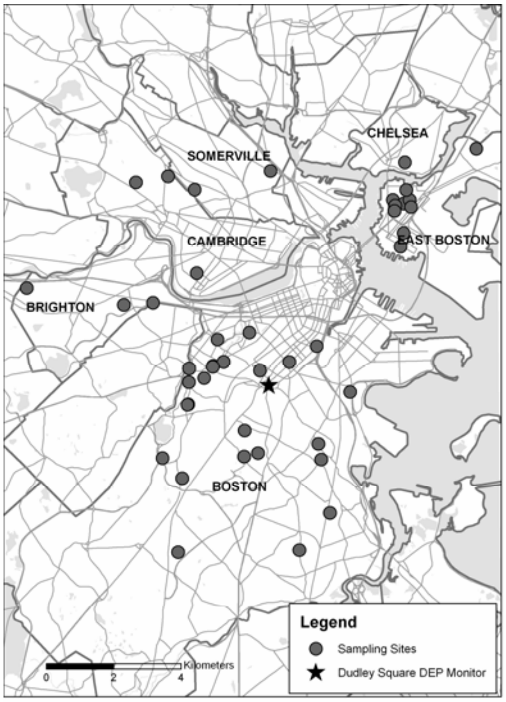

Figure 1. Location of sampling sites and DEP monitor

Table 1. Indoor, home-specific outdoor and ambient (from centrally located monitors) concentrations

Pollutant	Category	N	Mean (SD)	Median	Range
NO_2 (ppb)	Indoor	54	19.6 (11.0)	17.1	5.67 – 61.1
	Home-Specific Outdoor	52	17.2 (5.67)	16.8	5.21 – 33.3
	Ambient	52	18.4 (3.86)	18.3	12.2 – 27.6
$PM_{2.5}$ (μg/m^3)	Indoor	64	20.3 (12.5)	16.7	6.77 – 74.9
	Home-Specific Outdoor	60	14.2 (5.43)	12.6	6.75 – 31.3
	Ambient	60	15.4 (6.07)	14.6	6.24 – 45.7
EC (μg/m^3)	Indoor[a]	62	0.47 (0.29)	0.41	0.10 – 1.8
	Home-Specific Outdoor[a]	58	0.52 (0.41)	0.46	0.10 – 3.2
	Ambient	58	0.86 (0.34)	0.83	0.28 – 1.9

[a] factor of 0.83 was used to convert from $m^{-1} \times 10^{-5}$ to μg/m^3 (Kinney et al. 2000), to allow for comparison between residential and ambient measurements.

Table 2. Coefficients of determination (R^2) for NO_2, $PM_{2.5}$, and EC concentrations in univariate regression models.

Pollutant	Indoor vs. outdoor	Indoor vs. ambient	Outdoor vs. ambient
NO_2	0.07	0.02	0.21
$PM_{2.5}$	0.23	0.20	0.65
EC	0.49	0.16	0.08

Table 3. Identification of traffic indicators contributing to indoor concentrations after adjusting for ambient concentrations and indoor source terms[a]

Pollutant	R^2	Model	β (SE)	p-value
NO_2 (ppb)	0.20	Ambient Concentrations	0.66 (0.35)	0.06
		Gas Stove Usage	5.0 (3.0)	0.11
		unweighted density at 50m buffer	0.06 (0.03)	0.02
$PM_{2.5}$ (μg/m^3)	0.36	Ambient Concentrations	0.99 (0.25)	<0.01
		Cooking Time	5.1 (2.9)	0.08
		Occupant Density	5.2 (2.2)	0.02
EC ($m^{-1} \times 10^{-5}$)	0.21	Ambient Concentrations	0.26 (0.09)	< 0.01
		Distance to nearest designated truck route	-7.2×10^{-5} (4.2×10^{-5})	0.01

[a] only models with significant ($p < 0.2$) covariates are shown

Table 4. GIS-based variables grouped by algorithm used to create them and their posterior probabilities. Covariates with posterior probabilities three times (12.6%) greater than the prior probability (4.2%) are presented in bold.

	NO_2	$PM_{2.5}$	EC
Cumulative Traffic Scores (number of cars/day)			
density of urban road [a] within 200m	2.39	3.48	3.02
unweighted density within 50m buffer	**26.5**	3.08	2.97
100m buffer	2.15	2.90	2.95
200m buffer	2.23	4.07	3.18
500m buffer	2.46	5.33	3.82
Kernel-weighted densities at 50m buffer	6.64	3.13	3.12
100m buffer	10.3	3.16	3.00
200m buffer	1.93	3.02	3.44
300m buffer	2.25	4.30	3.75
500m buffer	3.25	5.40	3.39
Distance based measures (m)			
Distance to nearest urban road	3.90	5.43	3.57
major road[b]	3.93	6.28	3.65
highway[c]	2.01	2.97	3.72
designated truck route	2.16	4.37	**14.3**
Roadway Segment Length (m)			
Total roadway length contained within 50m	5.76	3.48	3.36
100m	2.31	4.40	3.41
200m	2.30	5.18	2.95
300m	2.42	5.78	4.33
Average Daily Traffic Scores (number of cars/day)			
Average daily traffic (ADT)	2.04	8.34	5.04
ADT/distance to major road	2.27	3.00	3.45
Diesel Measures: based on our traffic counter			
Number of trucks/day on largest roadway within 100m	2.45	2.87	8.63
Diesel fraction on largest roadway within 100m	2.09	2.84	3.77
Trucks per day/distance to major road	4.06	3.16	2.96
Population Density (for census block containing sampling site)			
Population density	2.18	4.06	4.19

[a] urban road defined as > 8500 cars/day
[b] major road defined as > 13,000 cars/day
[c] highway defined as > 19,000 cars/day

Table 5. Regression analyses of contributors to indoor concentrations accounting for the effect modification of open windows[a]

	R^2	Model	β (SE)	p-value
NO_2 (ppb)	0.25	Ambient Concentrations	0.79 (0.35)	0.03
		Gas Stove Usage	6.8 (3.1)	0.04
		unweighted density at 50m buffer*open windows = Yes	0.07 (0.03)	0.01
		unweighted density at 50m buffer*open windows = No	-0.03 (0.06)	0.62
$PM_{2.5}$ (μg/m^3)	0.40	Ambient Concentrations*open windows = Yes	0.98 (0.32)	<0.01
		Ambient Concentrations*open windows = No	0.64 (0.32)	0.05
		Cooking Time	6.2 (2.9)	0.04
		Occupant Density	6.5 (2.3)	0.01
EC ($m^{-1} \times 10^{-5}$)	0.32	Ambient Concentrations	0.38 (0.09)	<0.0001
		Distance to nearest designated truck route* open windows = Yes	-9.2×10^{-5} (4.1×10^{-5})	0.03
		Distance to nearest designated truck route* open windows = No	1.0×10^{-4} (5.9×10^{-5})	0.86

[a] only significant interaction terms ($p < 0.2$) are shown