



# Land use regression modeling of intra-urban residential variability in multiple traffic-related air pollutants

The Harvard community has made this  
article openly available. [Please share](#) how  
this access benefits you. Your story matters

Citation	Clougherty, Jane E., Rosalind J. Wright, Lisa K. Baxter, and Jonathan I. Levy. 2008. Land use regression modeling of intra-urban residential variability in multiple traffic-related air pollutants. <i>Environmental Health</i> 7: 17.
Published Version	<a href="https://doi.org/10.1186/1476-069X-7-17">doi:10.1186/1476-069X-7-17</a>
Citable link	<a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:4892359">http://nrs.harvard.edu/urn-3:HUL.InstRepos:4892359</a>
Terms of Use	This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a>

Research

Open Access

## Land use regression modeling of intra-urban residential variability in multiple traffic-related air pollutants

Jane E Clougherty\*<sup>1</sup>, Rosalind J Wright<sup>2,3</sup>, Lisa K Baxter<sup>4</sup> and Jonathan I Levy<sup>1</sup>

Address: <sup>1</sup>Harvard School of Public Health, Department of Environmental Health, Landmark Center 4th Floor West, P.O. Box 15677, Boston, MA 02215, USA, <sup>2</sup>Channing Laboratory, Brigham & Women's Hospital, Harvard Medical School, Landmark Center, 401 Park Drive, Boston, MA 02215, USA, <sup>3</sup>Harvard School of Public Health; Department of Society, Human Development, and Health, 677 Huntington Avenue, Boston, MA 02215, USA and <sup>4</sup>National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC 27711, USA

Email: Jane E Clougherty\* - jcloughe@hsph.harvard.edu; Rosalind J Wright - rerjw@channing.harvard.edu;

Lisa K Baxter - lbaxter@hsph.harvard.edu; Jonathan I Levy - jilevy@hsph.harvard.edu

\* Corresponding author

Published: 16 May 2008

Received: 17 July 2007

*Environmental Health* 2008, **7**:17 doi:10.1186/1476-069X-7-17

Accepted: 16 May 2008

This article is available from: <http://www.ehjournal.net/content/7/1/17>

© 2008 Clougherty et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** There is a growing body of literature linking GIS-based measures of traffic density to asthma and other respiratory outcomes. However, no consensus exists on which traffic indicators best capture variability in different pollutants or within different settings. As part of a study on childhood asthma etiology, we examined variability in outdoor concentrations of multiple traffic-related air pollutants within urban communities, using a range of GIS-based predictors and land use regression techniques.

**Methods:** We measured fine particulate matter (PM<sub>2.5</sub>), nitrogen dioxide (NO<sub>2</sub>), and elemental carbon (EC) outside 44 homes representing a range of traffic densities and neighborhoods across Boston, Massachusetts and nearby communities. Multiple three to four-day average samples were collected at each home during winters and summers from 2003 to 2005. Traffic indicators were derived using Massachusetts Highway Department data and direct traffic counts. Multivariate regression analyses were performed separately for each pollutant, using traffic indicators, land use, meteorology, site characteristics, and central site concentrations.

**Results:** PM<sub>2.5</sub> was strongly associated with the central site monitor ( $R^2 = 0.68$ ). Additional variability was explained by total roadway length within 100 m of the home, smoking or grilling near the monitor, and block-group population density ( $R^2 = 0.76$ ). EC showed greater spatial variability, especially during winter months, and was predicted by roadway length within 200 m of the home. The influence of traffic was greater under low wind speed conditions, and concentrations were lower during summer ( $R^2 = 0.52$ ). NO<sub>2</sub> showed significant spatial variability, predicted by population density and roadway length within 50 m of the home, modified by site characteristics (obstruction), and with higher concentrations during summer ( $R^2 = 0.56$ ).

**Conclusion:** Each pollutant examined displayed somewhat different spatial patterns within urban neighborhoods, and were differently related to local traffic and meteorology. Our results indicate a need for multi-pollutant exposure modeling to disentangle causal agents in epidemiological studies, and further investigation of site-specific and meteorological modification of the traffic-concentration relationship in urban neighborhoods.

## Background

There is a growing body of literature linking geographic information system (GIS)-based measures of traffic exposure to asthma and other respiratory outcomes. In the U.S. and Europe, children living or attending school near truck routes and highways show greater asthma symptoms [1-3], asthma hospitalizations [4,5], respiratory illness [1], allergic rhinitis [6], and reduced lung function [7]. However, proximity measures can represent a variety of pollutants or other near-roadway exposures (i.e., noise, poverty). There is no consensus on which traffic indicators may best capture variability in different pollutants within different settings, and the specific exhaust components responsible for health effects remain unidentified. For these reasons, there is a need to distinguish the relative spatial patterns of multiple traffic-related air pollutants, and to estimate concentrations using different GIS-based traffic indicators applicable across larger epidemiological studies.

Pollutants of interest include nitrogen dioxide (NO<sub>2</sub>), fine particulate matter (PM<sub>2.5</sub>), and elemental carbon (EC); each has been linked to both respiratory health and vehicular emissions. One recent study distinguished their relative spatial distributions within urban settings using GIS; this study found greater intra-urban variability and stronger traffic influences for NO<sub>2</sub> and EC than for PM<sub>2.5</sub> in European cities [8]. Comparable multi-pollutant analyses in the United States or in other settings have been limited, especially with a focus on residential settings within epidemiological investigations.

Addressing multiple pollutants at residences in large cohort studies is valuable but imposes constraints on the exposure assessment. For example, equipment-intensive multi-pollutant sampling (including the indoor environment) generally limits the number of sites that can be sampled simultaneously, reinforcing the need for models with spatial and temporal components, which can calibrate spatial models over time. For outcomes like asthma etiology, models estimating long-term exposures are needed, implying that measurements need to be taken at multiple points in time and that models need to separate spatial from temporal factors to the extent possible. This necessitates a careful evaluation of the role of meteorology in modifying the relationship between traffic and concentrations, a topic that has received little attention in GIS-based models to date.

In addition, issues regarding choice of traffic indicators and spatial-temporal separation may be exacerbated within urban neighborhoods, as predictors shown to be significant elsewhere (i.e., land use type) lack adequate variability to predict concentrations. Moreover, unlike measurements collected in open spaces near major road-

ways, residential measures may be impacted by near-home sources (e.g., idling cars, home heating, smoking, grilling) and site characteristics altering the traffic-concentration relationship (e.g., building configuration). Monitoring at residences imposes logistical constraints related to site configuration (e.g., availability of power supply, secure space), which may modify the traffic-concentration relationship. Traffic data quality can also be poor in residential areas, as most government and commercially-available datasets rely on estimates for smaller roads, reducing variability and producing significant misclassification within residential areas. Finally, in North America, the diesel fraction of total traffic in residential neighborhoods is generally smaller than in Europe, and is poorly characterized, such that total traffic measures may be less predictive of EC concentrations.

Land-use regression (LUR), a standard approach for predicting pollutant concentrations using concentration measures, GIS-derived spatial parameters, and site characteristics, allows for the characterization of exposure differentials within urban areas, and has been shown to better capture small-scale intra-urban variability than does kriging, integrated meteorological-emission (IME) models, or dispersion models [9]. LUR models of traffic-related pollution have shown stronger relationships with children's respiratory outcomes than have simple distance-to-roadway measures [10]. LUR and other GIS-based methods, however, have shown poor generalizability, as parameters selected using available spatial data and local characteristics in one city may not produce comparable estimates for another. Most LUR studies to date have been based on measurements collected near roadways or in other unobstructed urban locations, often at a fixed height [11,12]. As such, most LUR studies find a clear effect of traffic, potentially over-estimating the influence of traffic on average personal exposures in the urban area, and few LUR studies have attempted to characterize the near-home environment in dense residential areas. A recent LUR study captured a similar geographic region as our analysis, but focused on metropolitan-scale variability for a single pollutant (black carbon) with comparatively little exploration of spatial predictors, traffic terms, or meteorological contributors beyond wind speed [13]. Only one previous LUR study, to our knowledge, has attempted to account for wind speed and direction in detail, but again for a single pollutant without an explicit residential focus [14].

In this study, we used LUR techniques and GIS-derived variables to investigate the varying associations between multiple traffic indicators and outdoor residential concentrations of multiple air pollutants within the urban neighborhoods in and adjacent to Boston, Massachusetts. We evaluated a suite of GIS-based traffic indicators, and

explored meteorology and residential site characteristics as potential modifiers of the traffic-concentration relationship, with the goal of understanding the extent to which traffic-concentration relationships may be different by pollutant, and ultimately to inform exposure modeling for future epidemiological analyses focused on urban cohorts.

## Methods

### Site selection

This exposure modeling effort was nested within the Asthma Coalition on Community, Environment and Social Stress (ACCESS) birth cohort study. Sample homes were selected to represent variability in traffic densities across Boston and other proximate neighborhoods. Candidate homes were geocoded using U.S. Census TIGRE files and City of Boston street parcel data, and initial traffic scores for each home were assigned using Massachusetts Highway Department (MHD) traffic volume data. As we anticipated first-order (Gaussian) decay of key pollutants in the first 100–300 meters near major roadways [15], we opted to create initial traffic scores for site selection by applying a kernel weighting function to total traffic counts for all road segments within 100 meters of the home. The kernel function approximates concentration gradients expected under Gaussian decay, assigning higher weights to road segments nearer to the home. Resultant traffic scores were divided into tertiles, and sampling homes were selected to represent the observed range of traffic scores and neighborhoods. Due to unbalanced cohort recruitment in the study's early stages, additional non-cohort participants were recruited to capture a wider range of traffic scores, and neighborhoods where further recruitment was anticipated were over-sampled. The spatial distribution of our final sampling cohort is shown in Figure 1, where homes are shaded by 100-meter kernel-weighted traffic score, against a surface of the same measure for each 50-meter cell.

### Sampling methods

We measured indoor and outdoor concentrations of  $PM_{2.5}$ ,  $NO_2$ , and EC in two seasons (summer: May through early October, winter: December through March) at 44 homes across Boston and nearby urban communities, though only outdoor measures are included in this analysis.  $PM_{2.5}$  was measured using the Harvard Personal Environmental Monitor (PEM) [16], EC using reflectance analysis of  $PM_{2.5}$  filters, using the M43D Smokestain Reflectometer (Diffusion Systems Ltd., London UK) and with the absorption coefficient calculated in accordance with ISO 9835, as described in [17].  $NO_2$  was analyzed using Yanagisawa passive filter badges [18], analyzed by spectrophotometry. Integrated measures for each pollutant were collected for one week per season per home wherever feasible, in two sessions of 3 to 4 days duration,

averaged to one mean concentration per home per season for our LUR analysis. 24-hour traffic counts were collected using the Trax I Plus (JAMAR Technologies, Horsham, PA), on the highest-density road within 100 m of the home, during each sampling period when this was feasible (i.e., not during periods with snow/ice on the ground). Questionnaires were administered to identify nearby sources and sampling week activities that may influence concentrations, as detailed elsewhere [19].

### Additional data sources

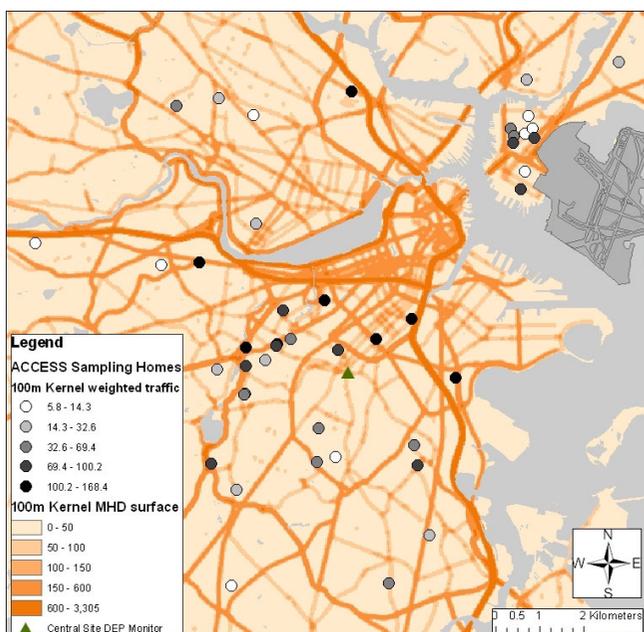
#### Traffic Data

Road networks and traffic data were obtained from MHD. Because different aspects of traffic including density, roadway configuration, and average vehicle speed may affect emission rates, pollutant mix, and dispersion, we opted to create a suite of 25 traffic indicators (Table 1) capturing varying aspects of traffic. We built raster-based cumulative density scores for average daily traffic (ADT) counts within radii of 50 to 500 meters around each home. Because roadway segments nearer to the home may have greater influence on concentrations, we also explored inverse-distance quadratic functions (kernel-weighted buffers) for the same radii. As traffic counts on smaller residential roads were sparse, we created cumulative density scores including only larger roads (above 8,500 cars/day), summary measures of total roadway length within radii of 50 to 500 meters, and the product of roadway length and average daily traffic counts within 200 meters. We considered distance to various roadway types, including the nearest larger road (greater than 8,500 cars/day), major road (13,000 cars/day), highway, and designated truck route. Lastly, to explore the influence of major roads on nearby neighborhoods, we created indicators of its average daily traffic, diesel traffic (estimated using axle length by the Trax I Plus), and weighted each by the home's distance to the road.

We considered other GIS covariates that may be associated with traffic, represent other pollutant sources, or modify the observed traffic-concentration relationship. Block group-level population and area measures were used to estimate population density. NCLD-50 land use categories and elevation data were downloaded from the USGS National Land Cover Dataset (NLCD) and National Elevation Dataset (NED), respectively.

#### Temporal variability: Background concentrations and meteorology

With a residential multi-pollutant approach, we were able to sample at a maximum of three homes per week, creating the need to account for temporal variability in background concentrations and meteorology. We estimated the influence of temporal heterogeneity in our data by regressing measured concentrations against mean central site concentrations for specific hours that each sample was



**Figure 1**  
**100-meter kernel-weighted traffic scores for urban area and sampling homes (Vehicle-miles per day/km<sup>2</sup>).**

collected. This temporal correction method is similar to that used elsewhere [20], though annual averages were not calculated.

Our primary central site concentration data were collected from the Massachusetts Department of Environmental Protection (DEP) monitor in the central Roxbury neighborhood (Figure 1). Hourly NO<sub>2</sub> is measured at the DEP monitor using the TECO42c chemiluminescence method. Hourly PM<sub>2.5</sub> is measured using the Met-One BAM with a PM<sub>2.5</sub> SCC beta attenuation method. Notably, EC is measured using the AE22ER aethelometer for optical absorption (Magee Scientific, Berkeley CA), as compared with the reflectance analysis used at our homes. The relationship between aethelometer and reflectance measures of EC has previously been found to differ by season in Boston, with aethelometers reading high in the summer (biased upward by 30% or more) and lower in winter (G. Allen, personal communication), and recent studies have shown similar seasonal biases with the aethelometer in other cold weather settings [21]. Although hourly data NO<sub>2</sub> were available at two additional nearby sites, with hourly data for EC and PM<sub>2.5</sub> available at one additional site each within the city, we used the Roxbury central site monitor in our main model given the availability of all three pollutants. We regressed outdoor PM<sub>2.5</sub>, EC, and NO<sub>2</sub> concentrations against mean DEP concentrations for the specific hours that each residential sample was col-

lected. Temporally-adjusted residuals were used for selection of spatial covariates, and final models were sensitivity tested against the use of data from other DEP monitors.

Meteorological data were collected from the same central site, because windspeed and direction could not be measured at each home during the sampling period. Mean windspeed and direction were calculated for daytime hours (6am–9pm) within each sampling period, when we anticipate significant traffic, our main source of interest. Further, several wind parameters were created in relation to traffic sources (i.e., percent of sampling hours when home is downwind from the nearest road), such that significance of the wind term implies source significance. Lastly, although meteorological texts define 'still winds' as below 1 m/s [22], we used 2.0 m/s to better dichotomize our high-windspeed dataset (median = 4.9 m/s). Meteorological factors and other covariates considered as effect modifiers of the traffic-pollution relationship are summarized in Table 2.

#### *Analytic methods and model-building*

We built models separately by pollutant, allowing different aspects of traffic, meteorology, and site-specific factors to predict concentrations of different pollutants. We selected candidate traffic indicators and modifiers against the temporally-corrected residuals, using nonparametric univariate correlations (Spearman correlations,  $p < 0.3$ ) of concentrations against traffic indicators as our primary selection method.

Because traffic indicators are highly correlated, however, we considered cluster analysis as a secondary selection method; the tree command in R groups observed concentrations by applying an impurity criterion to minimize within-group variances while maximizing between-group differences. The command compared concentration groups created using the 25 examined traffic indicators as predictors, and returned the indicators which best distinguished, as a group, high and low pollution locations, and the most effective binary cut-point for each indicator. Multivariate models were built using those traffic indicators selected by both correlation and clustering methods.

Using a stepwise forward regression process, we first included central site data, then traffic indicators, meteorological and site-specific modifiers as interaction terms with traffic indicators. Finally, we examined the effect of additional sources (e.g., grilling or smoking noted near outdoor monitor, block group population density, land use type, proximity to industry, season). We note that several of these indicators may be associated with traffic, capturing some traffic effect. We used the general form of Equation 1, and a maximum p-value of 0.1 to retain variables at each stage.

**Table 1: Traffic indicators examined for GIS-based LUR models.**

Indicator type	Indicator	Units
<b>Cumulative density scores:</b>	Unweighted density within: 50, 100, 200, 300, 500 m buffers	Vehicle-meters/per day/m <sup>2</sup>
	Kernel-weighted density: 50, 100, 200, 300, 500 m buffers	Vehicle-meters/per day/m <sup>2</sup>
	Density of urban roads (> 8500 cars/day) within 200 m	Vehicle-meters/per day/m <sup>2</sup>
<b>Summary measures:</b>	Total roadway length within: 50, 100, 200, 300, 500 m	Meters
	Total ADT*Length (VMT) within 200 m	Vehicle-meters per day
<b>Distance-based measures:</b>	Distance to nearest urban road (>8500 cars/day)	Meters
	To nearest major road (>13,000 cars/day)	Meters
	To nearest highway (>19,000 cars/day)	Meters
	To nearest MHD-designated truck route	Meters
<b>Characteristics of nearest major road:</b>	Average daily traffic (ADT)	Vehicles/day
	ADT/Distance to major road	(Vehicles/day)/m
	Diesel fraction	Percent (%)
	Trucks per day	Vehicles/day
	Trucks/Distance to major road	(Vehicles/day)/m

**Table 2: Potential effect modifiers of traffic-concentration relationship**

		Units
<b>Home characteristics:</b>	Obstructed from road	Yes/no
	Obstructed from major road	Yes/no
<b>Sampling period characteristics:</b>	Percent of hours downwind from major road	Percent (%)
	Average windspeed during daytime sampling hours (6am–9pm)	m/s
	Percent daytime hours with still winds (< 2 m/s)	Percent (%)
	Percent of weekend sampling days	Percent (%)
	Floor (monitor height)	(Categorical: 1, 2, 3+)
	Snow during sampling period	Yes/no

$$Concentration_{ijt} = \beta_{0j} + \beta_{1j} * DEP_{jt} + \beta_{2j} * Traffic_i + \beta_{3j} * Traffic_i * Modifier_{it} + \beta_{4j} * Other\ sources_{it} + e_{ijt} \quad (1)$$

Where  $Concentration_{ijt}$  is the measured concentration of pollutant  $j$  at location (home)  $i$  during sampling period (time)  $t$ .  $DEP_{jt}$  is the mean concentration of pollutant  $j$  at the central site during sampling period  $t$ .  $Traffic_i$  is the value of each traffic indicator listed in Table 1, tested separately in prediction models, at location  $i$ .  $Modifier_{it}$  is the value of meteorological or site characteristics altering the association between traffic indicators and  $Concentration_{ijt}$ .  $PM_{2.5}$ , EC, and  $DEP_{jt}$  values for  $PM_{2.5}$  and EC were log-distributed, and thus transformed prior to covariate selection and model building in our primary model.  $NO_2$  values were normally distributed, and not transformed.

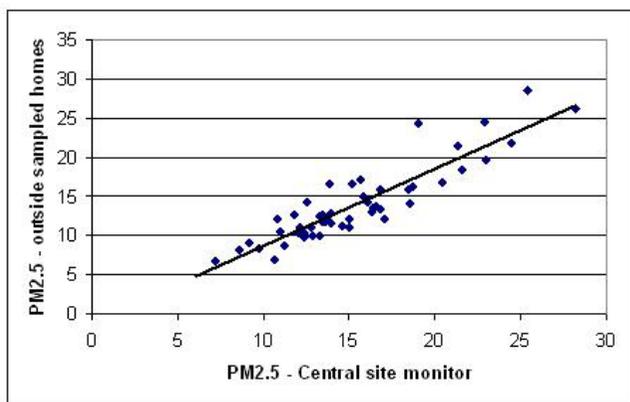
For residential EC, reflectance values are indicated by filter absorbance (units of  $m^{-1} * 10^{-5}$ ). To facilitate interpretation of our findings, we approximate these values to as mass units using  $0.83 \mu g/m^3/m^{-1} * 10^{-5}$ , derived from side-by-side reflectance and quartz fiber samples collected during summer in the urban northeastern U.S. [23]. This relationship may vary by location and season; because we anticipated that residential EC may display a different relationship with central site EC by season, we built our

models using non-converted (reflectance) units, and allowed for season-specific slopes in each model. Along with the aforementioned concerns about biases with aethalometer data and seasonal variability in the mass-absorption relationship, hypothesized sources of EC (i.e., wood smoke, home heating fuel) may display greater spatial variability during winter, when lower atmospheric mixing height may increase their influence.

**Sensitivity Analyses**

Extensive sensitivity tests were performed on the final model for each pollutant. Models were examined for sensitivity to the selection of traffic indicator by individually substituting each traffic indicator from Table 1. Likewise, we examined the selection of meteorological and site-specific modifiers by individually substituting other candidates. In each case, the final model was retained based upon overall model fit ( $R^2$ ).

To examine the quality of resolution in our area-level (raster) GIS data, we considered a range of base cell sizes (the smallest spatial unit employed in variable creation), varying in width from 10 to 50 meters. To test the quality and robustness of our road traffic data, we compared our traffic indicators derived from MHD data to comparable



**Figure 2**  
**Scatter plots of outdoor concentrations vs. central site concentration averages during sampling periods.**  $PM_{2.5}$  at homes vs. central site ( $\mu g/m^3$ ).

indicators using three other data sources. We initially log-transformed  $PM_{2.5}$  and EC data due to its lognormal concentration distribution, and tested the effect of this transformation. To examine the potential for residual spatial confounding using our central-site monitor data, we evaluated the use of other Boston-area DEP monitors. To assess residual seasonality not captured by DEP data, we tested a categorical season term and seasonally-varying slopes associating home data to the central site for all pollutants. Finally, we examined the robustness of each model to within-site autocorrelation owing to multiple measures at each site, using random effects by household.

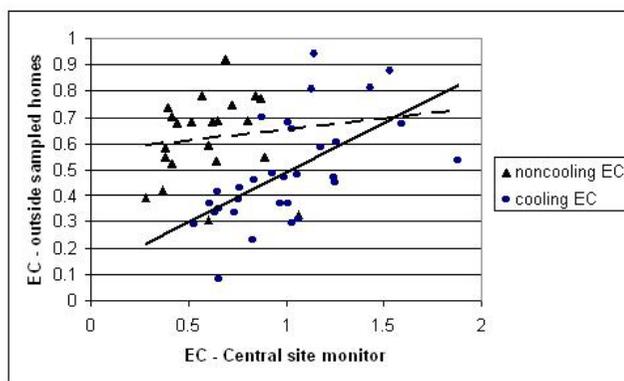
All traffic and land use variables were created in ArcGIS 9, clustering analyses were performed using the tree command in R version 2.2.0, and model-building in SAS version 9.1.

**Results**

We conducted 66 sampling sessions in total, consisting of 86 three-to-four day measurements in 44 homes. Fifty-one measurements were taken in 36 homes during summer months, and 35 measurements were taken in 25 homes during winter months. Table 3 summarizes the within-season average concentrations by sampling session for each pollutant.  $PM_{2.5}$  and EC were significantly correlated during winter and summer ( $p < 0.05$ ), while EC and  $NO_2$  were marginally correlated in both seasons, and  $PM_{2.5}$  and  $NO_2$  were not.

**Pollutant-specific modeling results**

Outdoor  $PM_{2.5}$  was highly correlated with central-site  $PM_{2.5}$  ( $R^2 = 0.68$ ), as suggested in Figure 2, indicating a predominance of temporal variability and relative spatial homogeneity in  $PM_{2.5}$  across the urban area. In multivari-



**Figure 3**  
**Scatter plots of outdoor concentrations vs. central site concentration averages during sampling periods.** EC at homes vs. central site ( $\mu g/m^3$ ); one influential point removed each season.

ate regressions including central site data, the best traffic indicator was total roadway length within 100 meters of the home (Table 4). Final multivariate model results indicate that the traffic- $PM_{2.5}$  relationship was not significantly altered by any of our candidate modifiers. Other combustion sources (smoking or grilling) and population density significantly contributed to concentrations (overall  $R^2 = 0.76$ ).

EC shows relatively poor associations with central site data overall ( $R^2 = 0.08$ ), though this is partly attributable to seasonal differences in the relationship (Figure 3), with varying slopes and stronger correlations during summer (Spearman  $r = 0.66$ ) than winter ( $r = 0.37$ ). In the final multivariate model ( $R^2 = 0.52$ ), EC was best predicted by total roadway length within 200 meters of the home, and the association between EC and traffic was increased under low wind speed conditions. During summer months, residential EC concentrations were somewhat lower and displayed stronger associations with central site data. Approximately 30% of the variability in EC was explained by temporal terms, and 14% by the traffic term (spatial component). The interaction of traffic with hours of low wind speed, incorporating both spatial and temporal variance, accounted for an additional 8%.

$NO_2$  was weakly associated with central site concentrations ( $R^2 = 0.21$ ), suggesting significant spatial heterogeneity within urban residential areas (Figure 4). The final multivariate model ( $R^2 = 0.56$ ) includes total roadway length within 50 meters of the home, significantly attenuated by an obstruction (i.e., building) between the monitor and nearest major road. Residential  $NO_2$  concentrations were higher during summer months, and positively associated with population density (Table 4).

Spatial terms (traffic, obstruction between the monitor and nearest major road, and population density) together account for approximately 23% of NO<sub>2</sub> variability. Temporal terms (central site, summer months) account for about 34%.

**Sensitivity analyses**

*Selection of traffic indicator*

Sensitivity analyses indicate that other traffic indicators could not be substituted to create a comparable model for PM<sub>2.5</sub> (Table 5) For EC (Table 6), diesel-based measures can explain more variability, with R<sup>2</sup> values of approximately 0.54, but were available for only a subset of locations (n = 34) and thus were not considered robust for the primary model. For the full cohort, no indicator was exchangeable with roadway length within 200 meters of the home. In addition, the interaction term of traffic modified by low wind speeds remained significant in several cases where the main effect of traffic did not maintain significance. For NO<sub>2</sub>, sensitivity tests (Table 7) support the finding that shorter buffer lengths were most effective. Larger buffer lengths did not produce a comparable model, but kernel-weighted traffic density within 50 meters of the home could be substituted effectively, as could unweighted cumulative density within 100 meters.

*Accuracy of traffic data*

To validate raster-based traffic indicators, we considered a range of base cell sizes from 10 to 50 meters square, bearing little difference on traffic indicator values compared to our default 25 meter cell size. Given concerns about data quality, where possible we verified MHD counts against traffic data obtained from the Massachusetts Executive Office of Transportation, ESRI Business Analyst, and our traffic counts collected outside cohort homes using the Jamar Trax I device. Correlations across traffic sources were generally above 0.7.

*Selection of central site monitor*

We considered several alternatives to the use of the Roxbury central site monitor concentrations for temporal correction, including using the other available urban monitors individually, the average concentration from all urban monitors available during each sampling period,

and the mean concentration at a background monitor south of Boston (available for summer months only). No alternative to the Roxbury central site sampling period mean explained greater variability in concentrations or significantly altered traffic-pollution relationships in multivariate models.

*Selection of meteorological and site-specific modifiers for EC and NO<sub>2</sub>s*

All EC models showed a significant, positive effect of low wind speeds on the traffic-pollution relationship. Sensitivity analyses indicated that other wind variables (mean daytime windspeed, percent of day downwind from road) were significant and may be substituted for percent of low wind speed hours, losing only marginal explanatory power (R<sup>2</sup> = 0.52 and 0.49, respectively). The similar findings for windspeed and direction may be expected, as windspeed and direction at our central site were highly correlated, with higher windspeeds from the west (data not shown).

For NO<sub>2</sub>, no other modifier could replace obstruction between the monitor and nearest major road in the final model. Because presence of an obstruction could theoretically proxy for distance to nearest major road, we replaced the term with distance to major road, and found highly non-significant results, indicating that this was not likely the case.

*Log-transformation of PM<sub>2.5</sub> and EC data*

The selection of the 100-meter roadway length term and other predictors for the PM<sub>2.5</sub> model was not dependent on log transformation. Using un-transformed PM<sub>2.5</sub>, we achieve an R<sup>2</sup> of 0.73, and retain significance in all predictors. For un-transformed EC, the same traffic term and all other predictors retained significance, with an R<sup>2</sup> of 0.51.

*Inclusion of a categorical variable for season*

Because the season term may be extraneous in models including temporal data from a central site, we explored the effect of removing this term from the final EC and NO<sub>2</sub> models. For EC, removing the season term caused the central site monitor estimate to drop by half and fall out of significance, while the effect of low wind speed

**Table 3: Within-season average outdoor concentrations at homes and central-site monitor**

Pollutant	Location	Overall			Summer			Winter		
		N	Mean (SD)	Median	N	Mean (SD)	Median	N	Mean (SD)	Median
PM <sub>2.5</sub> (µg/m <sup>3</sup> )	Outdoor	59	13.9 (5.0)	12.5	35	15.1 (5.7)	13.8	24	12.2 (3.1)	13.8
	Central Site	59	15.4 (6.1)	14.6	35	17.0 (7.1)	14.9	24	13.1 (3.4)	12.9
EC (µg/m <sup>3</sup> )	Outdoor	58	0.52 (0.41)	0.46	34	0.51 (0.51)	0.39	24	0.54 (0.18)	0.56
	Central Site	58	0.86 (0.34)	0.83	34	1.01 (0.31)	1.01	24	0.65 (0.27)	0.61
NO <sub>2</sub> (ppb)	Outdoor	52	17.2 (6.0)	16.8	36	17.7 (6.4)	17.3	16	16.3 (4.9)	15.5
	Central Site	52	18.4 (3.9)	17.9	36	17.3 (3.5)	18.0	16	21.0 (3.4)	21.3

increased by almost 50%, and overall model fit declined. Because of this decline in overall explanatory power when removed and the interpretability of the term given methodological aspects of EC concentration estimation, we opted to maintain the season term and season-specific slopes on the central site monitor term in the final model.

For NO<sub>2</sub>, dropping the season term decreased the effect of the central site monitor by approximately 50%, but did not affect overall model fit or other parameters. Thus, although NO<sub>2</sub> is higher at our residences during summer months, the effect is captured in part by the central site monitor; because the term did not significantly alter other parameters, we opted to leave it in the final model. Finally, we tested the addition of a season term to the PM<sub>2.5</sub> model, and found no effect on the central site term or overall fit, although the influence of other combustion sources (i.e., smoking or grilling) was increased by approximately 35%, indicating possible seasonal differences in source activities. Because traffic is our main source of interest, however, and because this result did not improve overall fit or alter the observed influence of traffic, we opted to maintain the more parsimonious original PM<sub>2.5</sub> model.

*Robustness to within-site autocorrelation*

For all pollutants, because the majority of homes were monitored in two seasons, we examined the effect of within-site autocorrelation using random effects by

household. Autocorrelation by site did not influence model fit or parameter estimates for any model.

Finally, a one-at-a-time exclusion cross-validation was performed to assess the internal consistency of model results. The Spearman correlation between estimated and measured log PM<sub>2.5</sub> was 0.80, 0.66 for log EC, and 0.66 for NO<sub>2</sub> (p < .0001 in all cases), indicating strong associations between predicted and actual values, indicating acceptable internal validity.

**Discussion**

Working strictly within urban neighborhoods and employing a multi-pollutant approach, our study offers several findings useful to future research exploring and modeling air pollution exposures for epidemiological purposes. These observations broadly apply to four areas: (1) urban residential variability in traffic densities and pollution concentrations, (2) fraction of urban residential pollution that is attributable to traffic, (3) selection of traffic indicators for residential exposure estimation, and (4) modification of traffic-concentration relationships by site characteristics and meteorology.

**(1) Urban residential variability in traffic densities and pollution concentrations**

We found significantly greater variability and stronger relationships with local traffic for EC and NO<sub>2</sub> than for PM<sub>2.5</sub>, consistent with prior literature [24,25], and which

**Table 4: Final model results for three pollutants**

Predictor Type	Model	ln(PM <sub>2.5</sub> ) (µg/m <sup>3</sup> )		Model	ln(EC) (m <sup>-1</sup> *10 <sup>-5</sup> )		Model	NO <sub>2</sub> (ppb)	
		β (p-value)	Sequential R <sup>2</sup>		β (p-value)	Sequential R <sup>2</sup>		β (p-value)	Sequential R <sup>2</sup>
<b>Intercept</b>		0.205 (.32)	--		-0.907 (<.0001)	--		-12.50 (.009)	--
<b>Central site Concentration</b>	ln (Central Site [PM <sub>2.5</sub> ])	0.776 (<.0001)	.68	ln (Central site [EC])	0.103 (.59)	.03	Central site [NO <sub>2</sub> ]	1.06 (<.0001)	.21
				ln (Central site [EC]) * warmer season	0.82 (.004)	.26			
<b>Traffic Indicator</b>	Roadway Length in 100 m	1.48*10 <sup>-4</sup> (.02)	.70	Roadway Length in 200 m	1.10 * 10 <sup>-4</sup> (.01)	.40	Roadway Length in 50 m	0.0144 (.002)	.22
<b>Traffic Indicator* Modifier</b>	N/A	N/A	N/A	Roadway Length in 200 m × % Hours of Still Winds	4.38 *10 <sup>-4</sup> (.02)	.48	Roadway Length in 50 m × Obstructed Major Rd	-0.0094 (.005)	.31
<b>Other Sources/ Land Use</b>	Smoking or grilling	0.156 (.01)	.73	Warmer Season	-0.268 (.057)	.52	Warmer Season	4.93 (.001)	.44
	Population Density	9.24*10 <sup>-6</sup> (.01)	.76				Population Density	4.01*10 <sup>-4</sup> (.001)	.56

corroborates evidence that PM<sub>2.5</sub> patterns are largely regional in nature for the Eastern U.S. [26,27]. We found somewhat weaker correlations across the three pollutants than have been shown in prior European studies [1], potentially because of our focus within urban neighborhoods relatively proximate to one another, while most prior intra-urban studies have actually examined metropolitan regions.

Though we found significant variability in traffic density across cohort homes, traffic varied somewhat less within residential neighborhoods than across the entire urban core, as shown in Figure 1. Across the entire urban core, 100-meter kernel-weighted traffic scores ranged from 0 to 3,305 vehicle miles traveled (VMT) per m<sup>2</sup> per day; at cohort homes, this measure ranged from 5.8 to 168 VMT/m<sup>2</sup>-day. This difference is driven largely by major highways, alongside which relatively few homes are located, but this observation may be important for exposure estimation; many models are derived from concentration data near major roads, which may inaccurately reflect traffic-concentration associations at the lower end.

**(2) Fraction of urban residential pollution that is attributable to traffic**

Using the traffic-pollution relationship observed across our sampling sites, we can estimate the portion of residential concentrations that are attributable to traffic. For PM<sub>2.5</sub>, the mean 100-meter summary roadway length of 1,110 meters accounted for a marginal contribution of 1.2 µg/m<sup>3</sup>, or 9.7% of predicted PM<sub>2.5</sub>. Applying our predictive models with mean values for all terms, mean predicted concentration is 13.2 µg/m<sup>3</sup>; an increase of one standard deviation in roadway length alone (371 meters) increases concentrations to 13.9 µg/m<sup>3</sup>. Population density, which likely captures some traffic-related influence, adds 1.1 µg/m<sup>3</sup> on average. Using non-transformed PM<sub>2.5</sub>, the predicted combined traffic contribution is somewhat larger (2.6 µg/m<sup>3</sup>).

For EC, the mean 200-meter buffer roadway length of 3,560 meters accounted for approximately 0.17 µg/m<sup>3</sup>, or 36% of total predicted EC. Increasing roadway length alone by one standard deviation (1,156 meters), with all other parameters at their mean values, increases predictions from 0.47 to 0.54 µg/m<sup>3</sup>. Using non-transformed EC concentrations, total predicted traffic contributions are somewhat larger (0.39 µg/m<sup>3</sup>). We observed a gradient of almost 1 µg/m<sup>3</sup> in EC across sampled homes before correcting for temporal variability, which is relatively small compared to that observed in European studies (approximately 10 µg/m<sup>3</sup>), as expected given the prevalence of diesel passenger vehicles in Europe [28].

Modeled traffic terms accounted for approximately 2.8 ppb, or 21% of modeled NO<sub>2</sub>. A mean 50-meter buffer roadway length of 441 meters, with mean values for other terms, predicts a concentration of 17.8 ppb. A one standard deviation increase (179 meters) increases concentrations to 18.9 ppb. The range of 50-meter roadway lengths observed predicts a NO<sub>2</sub> range of 15.4 to 20.9 ppb. Population density, likely capturing some traffic effect, and accounts for 4.4 ppb on average. The full range of NO<sub>2</sub> variability observed (including spatial and temporal components) was approximately 28 ppb (53 µg/m<sup>3</sup>), on the order of exposures associated elsewhere with increased wheeze and cough among children [29].

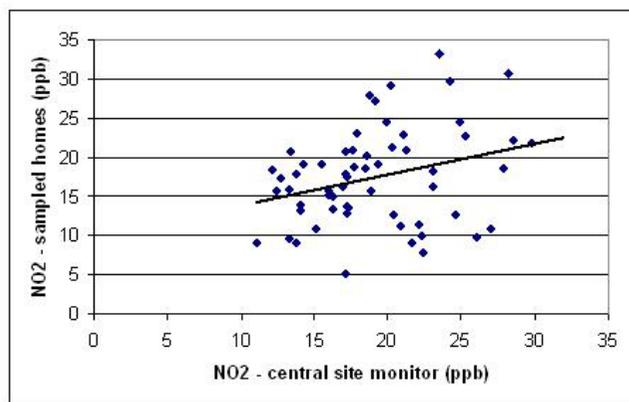
**(3) Selection of traffic indicators**

Total roadway length within varying buffer radii provided useful traffic indicators for each pollutant. Because actual traffic counts for residential roads are generally sparse, length measures provided more stable traffic indicators in these areas than do ADT-based estimates. There is also differential bias in traffic count accuracy by roadway size in the traffic data available for many cities; actual traffic counts are generally collected on a regular basis for highways and major roads, and rough estimates are created for smaller residential roads.

**Table 5: Sensitivity analyses including best predictor in each category, all summary measures.**

Indicator Type	Traffic Indicator	Estimate(s), p-value(s) from multivariate model	Model R <sup>2</sup>
<b>Base model (w/out traffic)</b>			R <sup>2</sup> = .74
<b>Cumulative Density scores</b>	Unweighted 500 m traffic density (n = 57)	$\beta_1 = 4.74 \times 10^{-4}$ (.17)	R <sup>2</sup> = .75
<b>Summary Measures</b>	Total roadway length:		
	50 meters (n = 57)	$\beta_1 = 1.31 \times 10^{-4}$ (.35)	R <sup>2</sup> = .74
	<b>100 meters (n = 57)</b>	$\beta_1 = 1.48 \times 10^{-4}$ (.02)	<b>R<sup>2</sup> = .76</b>
	200 meters (n = 57)	$\beta_1 = 1.42 \times 10^{-5}$ (.56)	R <sup>2</sup> = .74
	300 meters (n = 57)	$\beta_1 = 5.54 \times 10^{-6}$ (.57)	R <sup>2</sup> = .74
<b>Distance-based measures</b>	Distance to nearest designated truck route (n = 57)	$\beta_1 = 1.80 \times 10^{-5}$ (.62)	R <sup>2</sup> = .74
<b>Characteristics of nearest major road</b>	Average daily traffic (ADT) (n = 57)	$\beta_1 = -3.04 \times 10^{-6}$ (.25)	R <sup>2</sup> = .74

Final model bolded; models with diesel-based terms italicized. **Traffic indicators for PM<sub>2.5</sub>**



**Figure 4**  
**Scatter plots of outdoor concentrations vs. central site concentration averages during sampling periods.**  
 NO<sub>2</sub> at homes vs. central site (ppb).

While correlated, total roadway lengths within various buffers were not interchangeable in predictive models. The 100-meter buffer length effective for PM<sub>2.5</sub> coincides with our original buffer radius created for site selection, selected because maximum declines in particle concentrations occur in the first 100–300 meters alongside major roadways, dependent on particle size distributions and wind characteristics. We found a larger buffer of 200 m to be most predictive for EC; a meta-analysis of the literature [30] found buffers ranging from 50–250 meters across studies. The slightly higher value for EC than for PM<sub>2.5</sub> could be related to differential contributions of background and related complications, the relative lack of diesel sources and the likelihood of such sources on major roads (indicated by the significance of diesel terms related to nearest major road), or to the smaller particle size distribution anticipated for EC. For NO<sub>2</sub>, shorter buffer lengths of 50 meters were most effective. Secondary formation of NO<sub>2</sub> from NO might suggest longer buffer lengths, with one LUR study showing 300 m buffers to be effective [31] and one meta-analysis [30] indicating values of 200–500 m across studies. Other data, however, suggests that NO<sub>x</sub> concentrations can decline up to 60 percent, and NO<sub>2</sub> by 30 percent, within the first 50 meters from a roadway [12]. More generally, the precise distance-concentration decay relationship for NO<sub>2</sub> varies by setting, as dispersion, dilution, and chemical transformation are affected by the local pollution mix, wind characteristics, and meteorology.

Sensitivity tests suggest that diesel-based indicators may better predict EC and possibly NO<sub>2</sub>. Diesel influence raises a critical difference between traffic modeling in North America and Europe, as diesel vehicles constitute a smaller and poorly measured fraction of traffic in North

American residential areas. Here, we rely on axle-length estimates from our traffic counter on one major road near the home, to approximate vehicle type (i.e., CNG-powered buses, for example, would be categorized here as "diesel"). In Europe, the higher prevalence of diesel passenger vehicles implies that total traffic may itself act as a diesel marker, providing more stable predictors for EC.

#### **(4) Modification of traffic-concentration relationships by site characteristics and meteorology**

We found that accurate exposure modeling near urban residences required some consideration of site characteristics, such as population density and obstructions, which explained significant variability in concentrations and altered traffic-concentration associations. Population density, significant for PM<sub>2.5</sub> and NO<sub>2</sub>, may proxy either for other residential sources or for traffic, and may indicate higher per-mile emission rates from 'stop-and-go' traffic in denser residential neighborhoods. We expected obstruction to modify PM or EC, in keeping with recent findings suggesting that roadside barriers reduce PM concentrations [32]. Though we did not expect to find this effect for a gaseous pollutant such as NO<sub>2</sub>, this result is supported by recent evidence that residential NO<sub>2</sub> concentrations differed significantly depending on whether the home faced onto the courtyard or street, after accounting for distance to road [33]. Our findings may also be in part related to our passive sampling approach (as an obstruction could reduce the face velocity on the sampler).

#### **Limitations**

Although our models are physically interpretable and explain significant variability, there are some limitations to our analysis. Our small sample size may be a limitation in model-building, though sensitivity analyses show our findings to be robust. Similarly, our multi-pollutant approach limited our sampling design to sample at only a few homes simultaneously and distributing sampling sessions over the course of a season, such that samples incorporate both temporal and spatial variability. Measuring at a large number of homes simultaneously, however, is generally infeasible for equipment-intensive multi-pollutant sampling designs, especially given interest in both the indoor and outdoor environment. In addition, long-term residential exposure estimation can benefit from within-season temporal variability, which can not be obtained from short-term simultaneous sampling campaigns. Further, the influence of meteorological covariates may be under-estimated using such approaches, and are needed for long-term exposure estimation models. The central site monitor used to account for temporal heterogeneity may capture some local-source component as well, and for this reason we opted to use only one central site monitor with the most complete data coverage for all three pollutants, to avoid confounding spatial and temporal

**Table 6: Sensitivity analyses including best predictor in each category, all summary measures.**

Indicator Type	Traffic Indicator	Estimate(s), p-value(s) from multivariate model	Model R <sup>2</sup>
<b>Base model</b>			R <sup>2</sup> = .31
<b>Cumulative density scores:</b>	Unweighted 500 m buffer (n = 54)	$\beta_1 = 5.39 \times 10^{-4}$ (.49) $\beta_2 = 5.63 \times 10^{-3}$ (.41)	R <sup>2</sup> = .39
<b>Summary measures:</b>	Total roadway length within:		
	50 meters	$\beta_1 = 3.94 \times 10^{-4}$ (.18)	R <sup>2</sup> = .41
	*Still Winds (n = 54)	$\beta_2 = 4.08 \times 10^{-3}$ (.01)	
	100 meters	$\beta_1 = 2.15 \times 10^{-4}$ (.14)	R <sup>2</sup> = .47
	*Still Winds (n = 54)	$\beta_2 = 1.27 \times 10^{-3}$ (.03)	
	<b>200 meters</b>	<b><math>\beta_1 = 1.10 \times 10^{-4}</math> (.01)</b>	<b>R<sup>2</sup> = .52</b>
	<b>*Still Winds (n = 54)</b>	<b><math>\beta_2 = 4.38 \times 10^{-4}</math> (.02)</b>	
	300 meters	$\beta_1 = 2.99 \times 10^{-5}$ (.11)	R <sup>2</sup> = .48
	*Still Winds (n = 54)	$\beta_2 = 2.01 \times 10^{-4}$ (.04)	
Distance-based measures:	To nearest highway (>19,000 cars/day)	$\beta_1 = 0.452$ (.06)	R <sup>2</sup> = .45
	* Still Winds (n = 54)	$\beta_2 = 0.549$ (.04)	
<b>Characteristics of Nearest major road:</b>	<i>Diesel fraction</i>	<i><math>\beta_1 = -1.06</math> (.02)</i>	R <sup>2</sup> = .54
	* <i>Still Winds (n = 34)</i>	<i><math>\beta_2 = 34.6</math> (.02)</i>	
	<i>Trucks per day</i>	<i><math>\beta_1 = -7.41 \times 10^{-5}</math> (.06)</i>	R <sup>2</sup> = .54
	* <i>Still Winds (n = 34)</i>	<i><math>\beta_2 = 3.51 \times 10^{-3}</math> (.02)</i>	
	<i>Trucks/Distance to major road</i>	<i><math>\beta_1 = -6.31 \times 10^{-3}</math> (.03)</i>	R <sup>2</sup> = .54
	* <i>Still Winds (n = 34)</i>	<i><math>\beta_2 = 0.119</math> (.05)</i>	

Final model bolded; models with diesel-based terms italicized. **Traffic indicators for EC.**

variability during periods when some sites were unavailable. In addition, we chose to maintain the central site term in our final models, rather than predicting the temporally-corrected concentration residual, and included other temporal terms such as season and meteorological parameters.

Ultimately, most GIS-based residential exposure models are intended to allow exposure estimation across large cohorts, and thus we rely on readily-created GIS-based traffic indicators generally available across urban neighborhoods, such as total roadway length measures. Several predictors in our models, however, such as obstruction between the home and nearest major road, are effectively correction factors for the restrictions associated with residential monitoring – i.e. samplers often need be set up behind the buildings, wherever power sources are available, on a porch where smoking or grilling also occurred, or some such non-ideal location. These parameters may not be appropriate for extrapolation, as they may not reflect mean concentrations near the home, but are important to correctly interpreting residential data.

There are a number of issues related to LUR studies which limit model generalizability. First, LUR model results are highly dependent upon the quality of spatial data available. Here, for example, total roadway length produced the strongest concentration estimates in our urban neighborhoods. In areas with better traffic data, however, indicators incorporating traffic density may fare better. Second, spatial variables can have different meaning in different

settings; in rural areas, for example, proximity to major roads may be correlated with proximity to parking lots, industrial areas, and other sources, which is less likely in urban areas. Similarly, overall traffic counts in Europe can provide better estimates for EC modeling due to the higher prevalence of diesel vehicles. Higher emissions of NO<sub>2</sub> and total particle mass from diesel engines suggests that we may expect higher R<sup>2</sup>'s on European LUR models from intra-urban variability in traffic-related pollutants than in the US.

These issues of generalizability continue to challenge the search for "causal agents" in the association between traffic density and respiratory and cardiovascular illness. Because different traffic indicators have been shown to predict concentrations (and illness) in different regions, it is difficult to identify the specific spatial characteristics of unidentified causal agents. We maintain, however, that because the chemical and physical properties among various pollutants lead to differing rates of decay and deposition near roadways [15,30], predictive models should be built and compared for multiple pollutants in epidemiological studies of the effect of traffic exposure on health. Further, because pollutants more refined than PM<sub>2.5</sub> are still complex (e.g., EC, a subset of PM<sub>2.5</sub>, may have VOCs and metals bound to its surface), there remains a need for spatial models investigating distributions of specific PM<sub>2.5</sub> constituents to evaluate their relationship with health outcomes. Finally, most urban residents of North American cities spend the majority of their time indoors, further complicating efforts to define causal pollutants in the traf-

fic-health relationship. The models presented here do not address infiltration or indoor residential environments directly, but do facilitate estimation of indoor exposures when combined with home characteristics such as building type [19]. Finally, measurement error is differential across the three pollutants modeled here, as evidenced by their varying R<sup>2</sup>'s, which complicates comparisons across predictive models for different pollutants in epidemiological studies. This issue deserves greater attention, as such comparisons across pollutant-specific models will be important in identifying causal agents.

Our results provide some insight to researchers working to elucidate intra-urban residential exposures for long-term epidemiological analyses. First, the issue of distinguishing temporal from spatial variability is a significant difficulty for multi-pollutant sampling designs; when only a small number of homes can be sampled simultaneously, there is not adequate variability observed within any time period to distinguish the temporal from spatial effects. To this end, we suggest, in cities lacking a year-round rural background monitor for all pollutants of interest, a fixed-site monitor (using identical methods to those at the homes) in a less heavily-trafficked area would be useful to capture background concentrations and long-range transport. We also suggest exploration of sampling designs which maximize temporal overlap; for example, a systematically-staggered sampling with a predictable amount of overlap at different homes might allow for a simplified

temporal correction term. This design would retain intra-season variability as well, which is lost in the shorter-term simultaneous sampling campaigns. However, such theoretically optimal sampling designs should be approached realistically, with the recognition that residential sampling will require some amount of visit rescheduling (especially for a cohort such as ours, with women who are pregnant or with young children). Finally, in selecting sampling locations, certain site characteristics potentially impeding accurate sample collection should be considered (e.g., construction, lack of secure outdoor space or power outlets), and incorporated into previous methods for optimizing concentration variability, such as that outlined in [12].

**Conclusion**

Our analysis explored a range of GIS-based traffic indicators to capture small-scale variability in multiple air pollutants within dense urban neighborhoods. Because our measures were collected outside residential homes rather than at roadside locations, our measures likely reflect more realistic exposures to urban residents. The resultant sampling design, however, raised a number of methodological challenges, including the need to account for spatial and temporal variability in measures collected during different weeks, and the need to account for home site characteristics and other residential sources which may obscure the true traffic-concentration relationship.

**Table 7: Sensitivity analyses including best predictor in each category, all summary measures.**

Indicator Type	Traffic Indicator	Estimate(s), p-value(s) from multivariate model	Model R <sup>2</sup>
<b>Null model</b>			R <sup>2</sup> = .39
<b>Cumulative density scores:</b>	Unweighted traffic density in 100 m	$\beta_1 = 0.055 (.004)$	R <sup>2</sup> = .55
	* Obs nearest Major Road (n = 50)	$\beta_2 = -0.051 (.004)$	
	Kernel-weighted density in 50 m	$\beta_1 = 0.034 (.02)$	R <sup>2</sup> = .55
	* Obstructed (n = 50)	$\beta_2 = -0.056 (.002)$	
<b>Summary measures:</b>	Density of urban roads (>8500 cars/day) in 200 m	$\beta_1 = 589.4 (.049)$	R <sup>2</sup> = .52
	* Obstructed (n = 50)	$\beta_2 = -760.9 (.0095)$	
	Total roadway length within:		
	<b>50 meters</b>	$\beta_1 = 0.0144 (<.0001)$	<b>R<sup>2</sup> = .56</b>
* <b>Obstructed (n = 50)</b>	$\beta_2 = -0.0094 (.005)$		
100 meters	$\beta_1 = 0.0022 (.34)$	R <sup>2</sup> = .54	
* Obstructed (n = 50)	$\beta_2 = -0.0042 (.005)$		
200 meters	$\beta_1 = 9.28 \times 10^{-4} (.22)$	R <sup>2</sup> = .52	
* Obstructed (n = 50)	$\beta_2 = -1.25 \times 10^{-3} (.008)$		
300 meters	$\beta_1 = 2.31 \times 10^{-4} (.55)$	R <sup>2</sup> = .50	
* Obstructed (n = 50)	$\beta_2 = -6.65 \times 10^{-4} (.0095)$		
Vehicle Miles Traveled in 200 m	$\beta_1 = 1 \times 10^{-7} (.21)$	R <sup>2</sup> = .47	
* Obstructed (n = 50)	$\beta_2 = -1 \times 10^{-7} (.10)$		
<b>Distance-based measures:</b>	To nearest highway (>19,000/day)	$\beta_1 = 0.0176 (.12)$	R <sup>2</sup> = .50
	* Obstructed (n = 50)	$\beta_2 = -0.0195 (.07)$	
<b>Characteristics of nearest major road:</b>	Trucks per day on nearest major road *	$\beta_1 = .0061 (.01)$	R <sup>2</sup> = .59
	Obstructed (n = 34)	$\beta_2 = .00596 (.01)$	

Final model bolded; models with diesel-based terms italicized. **Traffic indicators for NO<sub>2</sub>.**

We found that traffic indicators not reliant on ADT estimates (i.e. roadway lengths) provided more stable predictors in the residential settings in our study. As shown elsewhere, greater spatial variability was observed in NO<sub>2</sub> and EC than in PM<sub>2.5</sub>, and LUR techniques worked well within the urban setting to capture pollutant variability, although parameters useful at the metropolitan scale (e.g., land use type) displayed low variability and limited predictive power. Finally, the relationship between traffic and pollution concentrations was significantly modified by meteorological factors and site characteristics, indicating the importance of incorporating small-scale spatial and temporal predictors to accurately capture exposure variability in urban residential settings.

### List of abbreviations

ACCESS: Asthma Coalition on Community, Environment, and Social Stress; ADT: average daily traffic; CNG: compressed natural gas; DEP: Massachusetts Department of Environmental Protection; EC: elemental carbon; GIS: geographic information systems; IME: integrated meteorological-emission models; LUR: land use regression; MHD: Massachusetts Highway Department; NED: National Elevation Dataset; NLCD: National Land Cover Dataset; NO<sub>2</sub>: nitrogen dioxide; PEM: Harvard Personal Environmental Monitor; PM<sub>2.5</sub>: fine particulate matter, smaller than 2.5 microns mean diameter; USGS: United States Geological Survey; VMT: vehicle miles traveled.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

JEC participated in exposure study design, sampling design, and data collection, created all spatial and meteorological variables, performed statistical analyses, and drafted the manuscript. RJW oversaw cohort recruitment and overall ACCESS study design. LKB participated in sampling design, data collection, and analysis. JIL oversaw sampling design, data collection, data analysis, and manuscript development. All authors read and approved the final manuscript.

### Acknowledgements

The primary author (JE Clougherty) gratefully acknowledges funding from the American Lung Association, which supported this work. In addition, the ACCESS project is supported by NIH R01 ES10932, U01 HL072494, and R01 HL080674 (RJ Wright, PI), and the analyses were supported by NIH R03 ES013988 and by the Health Effects Institute (HEI 4727-RFA04-5/05-1), JI Levy, PI. Many thanks to our participating families, and to Brent Coull, Chris Paciorek, and Jack Spengler for valuable suggestions and comments on the manuscript.

### References

1. Brauer M, Hoek G, Van Vliet P, Meliefste K, Fischer PH, Wijga A, Koopman LP, Neijens HJ, Gerritsen J, Kerkhof M, Heinrich J, Bellander T, Brunekreef B: **Air Pollution from Traffic and the**

2. **Development of Respiratory Infections and Asthmatic and Allergic Symptoms in Children.** *American Journal of Respiratory and Critical Care Medicine* 2002, **166**:1092-1098.
3. Gordian ME, Haneuse S, Wakefield J: **An investigation of the association between traffic exposure and the diagnosis of asthma in children.** *Journal of Exposure Science and Environmental Epidemiology* 2006, **16**:49-55.
4. Zmirou D, Gauvin S, Pin I, Sahraoui F, Just J, Mouleuc YL, Bremont F, Cassadou S, Reungoat P, Albertini M, Lauvergne N, Chiron M, Labbe A, Investigators V: **Traffic related air pollution and incidence of childhood asthma: results of the Vesta case-control study.** *Journal of Environment and Community Health* 2004, **58**:18-23.
5. Lin S, Munsie P, Hwang SA, Fitzgerald E, Cayo MR: **Childhood Asthma Hospitalization and Residential Exposure to State Route Traffic.** *Environmental Research* 2002, **88(A)**:73-81.
6. Edwards J, Walters S, Griffiths RK: **Hospital admissions for asthma in preschool children: relationship to major roads in Birmingham, United Kingdom.** *Archives of Environmental Health* 1994, **49(4)**:223-227.
7. Duhme H, Weiland SK, Keil U, Kraemer B, Schmid M, Stender M, Chambliss L: **The association between self-reported symptoms of asthma and allergic rhinitis and self-reported traffic density on street of residence in adolescents.** *Epidemiology* 1996, **7(6)**:578-582.
8. Brunekreef B, Janssen NAH, Hartog J, Harssema H, Knape M, van Vliet P: **Air pollution from truck traffic and lung function in children living near motorways.** *Epidemiology* 1997, **8(3)**:298-303.
9. Hochadel M, Heinrich J, Gehring U, Morgenstern V, Kuhlbusch T, Link E, Wichmann HE, Kramer U: **Predicting long-term average concentrations of traffic-related air pollutants using GIS-based information.** *Atmospheric Environment* 2006, **40**:542-553.
10. Jerrett M, Arain A, Kanaroglou P, Beckerman B, Potoglou D, Sahuvaroglu T, Morrison J, Giovis C: **A review and evaluation of intraurban air pollution models.** *Journal of Exposure Analysis and Environmental Epidemiology* 2004, **15(2)**:185-204.
11. Ryan PH, LeMasters GK, Biswas P, Levin L, Hu S, Lindsey M, Bernstein DI, Lockey J, Villareal M, Hershey GKK, Grinshpun SA: **A comparison of Proximity and land use regression traffic exposure models and wheezing in infants.** *Environmental Health Perspectives* 2007, **115(2)**:278-284.
12. Wheeler AJ, Smith-Doiron M, Xu X, Gilbert NL, Brook JR: **Intra-urban variability of air pollution in Windsor, Ontario -- Measurement and modeling for human exposure assessment.** *Environmental Research* 2008, **106**:7-16.
13. Kanaroglou PS, Jerrett M, Morrison J, Beckerman B, Arain MA, Gilbert NL, Brook JR: **Establishing an air pollution monitoring network for intra-urban population exposure assessment: A location-allocation approach.** *Atmospheric Environment* 2005, **39**:2399-2409.
14. Gryparis A, Coull BA, Schwartz J, Suh HH: **Semiparametric latent variable regression models for spatiotemporal modelling of mobile source particles in the greater Boston area.** *Applied Statistics* 2007, **56(2)**:183-209.
15. Arain MA, Blair R, Brook JR, Sahuvaroglu T, Beckerman B, Zhang L, Jerrett M: **The use of wind fields in a land use regression model to predict air pollution concentrations for health exposure studies.** *Atmospheric Environment* 2007, **41**:3453-3464.
16. Zhu Y, Hinds WC, Kim S, Sioutas C: **Concentration and Size Distribution of Ultrafine Particles Near a Major Highway.** *Journal of the Air and Waste Management Association* 2002, **52**:1032-1041.
17. Marple VA, Rukow KL, Turner WV, Spengler JD: **Low flow rate sharp cut impactors for indoor air sampling: Design and calibration.** *Journal of the Air Pollution Control Association* 1987, **37**:1303-1307.
18. Cyrus J, Heinrich J, Hoek G, Meliefste K, Lewne M, Gehring U, Bellander T, Fischer P, Vliet P, Brauer M, Wichmann HE, Brunekreef B: **Comparison between different traffic-related particle indicators: Elemental carbon (EC), PM2.5 mass, and absorbance.** *Journal of Exposure Analysis and Environmental Epidemiology* 2003, **13**:134-143.
19. Yanagisawa Y, Nishimura H: **A badge-type personal sampler for measurement of personal exposure to NO2 and NO in ambient air.** *Environment International* 1982, **8**:235-242.
20. Baxter LK, Clougherty JE, Laden F, Levy JI: **Predictors of concentrations of nitrogen dioxide, fine particulate matter, and par-**

- title constituents inside of lower socioeconomic status urban homes.** *Journal of Exposure Science and Environmental Epidemiology* 2007, **17**:433-444.
20. Brauer M, Hoek G, Vliet PV, Meliefste K, Fischer P, Gehring U: **Estimating long-term average particulate air pollution concentrations: Application of traffic indicators and Geographic Information Systems.** *Epidemiology* 2003, **14**:228-239.
  21. Virkkula A, Ahlquist NC, Covert DS, Arnott WP, Sheridan PJ, Quinn PK, Coffman DJ: **Modification, calibration, and a field test of an instrument for measuring light absorption by particles.** *Aerosol Sci Technol* 2005, **39**(1):68-83.
  22. Hanna S, Briggs GA, Hosker Jr. RP: **Handbook on Atmospheric Diffusion.** Technical Information Center, U.S. Department of Energy; 1982.
  23. Kinney PL, Aggarwal M, Northridge ME, Janssen NAH, Shepard P: **Airborne concentrations of PM2.5 and Diesel Exhaust Particles on Harlem Sidewalks: A Community-Based Pilot Study.** *Environmental Health Perspectives* 2000, **108**:213-218.
  24. Wu J, Lurmann F, Winer A, Lu R, Turco R, Funk T: **Development of an individual exposure model for application to the Southern California Children's Health Study.** *Atmospheric Environment* 2005, **39**:259-273.
  25. Martuzevicius D, Grinshpun SA, Reponen T, Gorny RL, Shukla R, Lockey J, Hu S, McDonald R, Biswas P, Kliucininkas L, LeMasters G: **Spatial and temporal variations of PM2.5 concentration and composition throughout an urban area with high freeway density - the Greater Cincinnati study.** *Atmospheric Environment* 2004, **38**:1091-1105.
  26. Burton RM, Suh HH, Koutrakis P: **Spatial Variation in Particulate Concentrations within Metropolitan Philadelphia.** *Environmental Science and Technology* 1996, **30**(2):400-407.
  27. Suh HH, Nishioka Y, Allen GA, Koutrakis P, Burton RM: **The Metropolitan Acid Aerosol Characterization Study: Results from the Summer 1994 Washington D.C. Field Study.** *Environmental Health Perspectives* 1997, **105**:826-834.
  28. Hoek G, Brunekreef B, Goldbohm S, Fischer P, van den Brandt PA: **Association between mortality and indicators of traffic-related air pollution in the Netherlands: a cohort study.** *Lancet* 2002, **360**:1203-1209.
  29. Nicolai T, Carr D, Weiland SK, Duhme H, Ehrenstein O, Wagner C, Mutius E: **Urban traffic and pollutant exposure related to respiratory outcomes and atopy in a large sample of children.** *Eur Respir J* 2003, **21**:956-963.
  30. Zhou Y, Levy JI: **Factors influencing the spatial extent of mobile source air pollution impacts: a meta-analysis.** *BMC Public Health* 2007, **7**:89.
  31. Jerrett M, Arain MA, Kanaraglou P, Beckerman B, Crouse D, Gilbert NL, Brook JR, Finkelstein N, Finkelstein MM: **Modelling the intra-urban variability of ambient traffic pollution in Toronto, Canada: Rome.** NERAM: Network for Environmental Risk Assessment and Management; 2003.
  32. Chen F, Dimmick F, Williams R, Baldauf R: **Particulate Matter Concentration Measurements from Near Road Studies (NRS) in Raleigh, North Carolina.** In *17th Annual Conference of the International Society for Exposure Analysis* Durham, NC ; 2007.
  33. Reungoat P, Chiron M, Gauvin S, Zmirou-Navier D, Momas I: **Retro-spective assessment of exposure to traffic air pollution using the ExTra index in the VESTA French epidemiological study.** *Journal of Exposure Analysis and Environmental Epidemiology* 2005, **15**:524-533.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

