Supporting Information

The impact of air exchange rate on ambient air pollution exposure and inequalities across all residential parcels in Massachusetts

Anna Rosofsky,*† Jonathan I Levy,† Michael S Breen,‡ Antonella Zanobetti,§ M. Patricia Fabian†

†Department of Environmental Health, Boston University School of Public Health, Boston, MA, USA
‡National Exposure Research Laboratory, U.S. Environmental Protection Agency, Research Triangle Park, NC, USA
§Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA

*Corresponding Author: Anna Rosofsky, email: arosofsk@bu.edu

Total page numbers: 14, including 7 tables (Table S1-S7), 3 figures (Figures S1-S3), and 4 descriptive sections (Section S1-S4)
Table of Contents

- Table S1. Data sources used in infiltration calculation and inequality analysis
- Table S2: Data aggregation and cleaning algorithms
- S1: Parcel imputation procedure
- Table S3: Variables imputed and included in multiple imputation procedure for all Massachusetts parcels
- S2: Normalized Leakage (NL) Estimation
- S3: Stack Coefficient and Wind Coefficient Estimation
- Table S4: Shelter Class $B_{fac}$ and $A_{fac}$ Parameters used in AER calculation
- S4: Extending LBNL Model to Multi-Family Units
- Table S5. Residential parcel frequencies for Equations S1a, S1b, S2 and S3
- Table S6. AER ($h^{-1}$) estimates complete case data
- S5: Indoor PM$_{2.5}$ of ambient origin estimation procedure
- Table S7. $C_{in}$ estimates using upper and lower confidence limits of $P$ and $k$ estimated in Breen et al. 2015
- Figure S1. Sociodemographic characteristics of block groups containing the residential parcels with the lowest air exchange rates (AER) ($<=$10$^{th}$ %tile) in areas with the highest ambient PM$_{2.5}$ ($=>$90$^{th}$ %tile) versus block groups containing parcels with the highest AER and lowest PM$_{2.5}$. Source: US Census 2010.
- Figure S2. Sociodemographic characteristics of block groups containing the residential parcels with the lowest air exchange rates (AER) in areas with the lowest ambient PM$_{2.5}$ (low-exposure, $<=$10$^{th}$ %tile) versus block groups containing parcels with the highest AER and PM$_{2.5}$ (high-exposure, $=>$90$^{th}$ %tile), stratified by urban/rural class. Source: US Census 2010.
- Figure S3. Map of Eastern Massachusetts in 2010 showing distribution of a) summer air exchange rates at parcel level, b) summer PM$_{2.5}$ concentrations at parcel level, c) % Hispanic at block-group, and d) % median annual household income below $20,000 at block group.
| Dataset                              | Variables                                         | Coverage Year | Geographic Resolution | Equation Use                                                                 | Publicly Available |
|-------------------------------------|---------------------------------------------------|---------------|-----------------------|------------------------------------------------------------------------------|---------------------|
| Massachusetts Tax Assessor          | Year built, number of stories, building area (m²) | 2009-2015     | Parcel                | Equations 1 and S1a-b, S2 & S3                                              | X                   |
| Census                              | Racial and ethnic characteristics                  | 2010          | Block group           | Equation S1a & S1b; Inequality analysis                                       | X                   |
| American Community Survey           | Socioeconomic characteristics                      | Average 5-year 2006-2010 | Block group         | Inequality analysis                                                          | X                   |
| Residential Energy Consumption Survey| Indoor temperature                                 | 2009          | N/A                   | Equation 2                                                                  | X                   |
| MassGIS Land Use                    | Categories of land use                             | 2005          | Point                 | Equation 2 and S2                                                            | X                   |
| Automated Surface Observing System  | Average seasonal wind speed (m/s)                  | 2010          | Point                 | Equation 2                                                                  | X                   |
|                                     | Average seasonal temperature (°C)                  | 2010          | Point                 | Equation 2                                                                  | X                   |
| 1 km² gridded surface PM₂.₅ (34)    | Average seasonal ambient PM₂.₅ (µg/m³)             | 2010          | 1 km² grid            | Equation S4; Inequality analysis                                             |                     |
Table S2. Data aggregation and cleaning algorithms

| Dataset                        | Variables                                              | Data Cleaning Process                                                                                                                                 |
|-------------------------------|--------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------|
| Massachusetts Tax Assessor     | Year built, number of stories, building area           | • If year built was before 1900, truncated at 1900; if year built was after 2015, truncated at 2015. This maintains consistency with empirical models from Chan et al. 2013  
• Building area cutoff based on RECS 2009 maximum and minimum building areas by housing type  
• Buildings with >60 stories were truncated at 60 stories |
| US Census                     | Racial and ethnic characteristics                     | • Not modified                                                                                                                                         |
| American Community Survey     | Socioeconomic characteristics                          | • Not modified                                                                                                                                         |
| Residential Energy Consumption Survey | Indoor temperature                                      | • Averaged reported indoor temperature in warming months (20°C) and cooling months (22°C) for Massachusetts participants in the Residential Energy Consumption Survey to estimate indoor temperature |
| MassGIS Land Use              | Categories of land use                                  | • Not modified                                                                                                                                         |
| Automated Surface Observing System | Average seasonal wind speed (m/s)  
Average seasonal temperature (°C) | • Averaged daily values to seasonal averages                                                                                                          |
| 1 km² gridded surface PM₂.₅ (1) | Average seasonal PM₂.₅ (µg/m³)                           | • Averaged daily values to seasonal averages                                                                                                          |
S1: Parcel imputation procedure

In order to estimate air exchange rate (AER) for all residential parcels in Massachusetts, we created a complete dataset of Massachusetts Assessor’s database variables by imputing missing parcel data by the Markov Chain Monte Carlo multivariate normal model (2). We assessed the missing at random assumption by examining missing data patterns. Data were log-transformed within the imputation procedure to fulfill the normality assumption. Missing parcel data were 11%, 19%, and 17% for building area, number of stories and year built, respectively. Using PROC MI, we created 15 imputed datasets as recommended by Graham et al (3). Details on variables included as predictors of missing observations and percent missingness can be found in Table S3.

| Variable                        | Percent Missing |
|---------------------------------|-----------------|
| Lot Area                        | 0               |
| Percent Impervious              | 0.07            |
| Square Meter Paved              | 0.08            |
| Percent Paved                   | 0.08            |
| Percent Building                | 0.16            |
| Total Value per Acre            | 0.32            |
| Building Value                  | 1.2             |
| Square Meters Impervious        | 1.8             |
| Number of Estimated Units*      | 1.9             |
| Land Value                      | 2.0             |
| Building/Land Ratio             | 2.2             |
| Floor/Area Ratio                | 2.5             |
| Square Meter Building           | 4.0             |
| Year Built*                     | 6.7             |
| Building Value per Square Foot  | 7.8             |
| Number of Stories*              | 8.6             |
| Building Area*                  | 10.2            |
| Number of Rooms                 | 16.8            |

*Variables used in AER equations
S2: Normalized Leakage (NL) Estimation

NL was calculated using literature-reported regression parameters from an empirically-derived leakage area model estimated from measurements of 70,000 homes across the United States (18,21,36). This particular area leakage model is suitable for the purposes of this work because it can be parameterized using publicly-available data sources. It has previously been found to perform equally well as alternative area leakage models (37).

In this model, median year built and floor area (m$^2$) were the main NL predictors (Equation S1a & S1b). Single family (SF) floor areas were taken directly from the assessor’s database parcel data, and we calculated floor area of a representative unit in a multi-family building by dividing building area by the number of units per floor. Chan et al. (2005) predicts NL separately for residents earning below 125% of the federal poverty line (NL$_a$) and all other homes (NL$_b$). Regardless of year built and floor area, homes below 125% of the poverty line from the database used in Chan et al (2005) were leakier than conventional homes, indicating residual neighborhood and individual-level characteristics beyond the age and size of the home that may influence home leakiness. We predicted NL$_a$ for parcels that fell within a BG categorized as a census-defined poverty area, where at least 20 percent of the households are below the poverty threshold (38). NL parameters for low-income and conventional homes are shown below:

\[
NL_a = e^{11.1+(-0.00537 \times \text{median year built}) + (-0.00418 \times \text{m}^2)} \quad \text{[Equation S1a]}
\]

\[
NL_b = e^{2.7+(-0.0107 \times \text{median year built}) + (-0.0022 \times \text{m}^2)} \quad \text{[Equation S1b]}
\]
S3: Stack Coefficient and Wind Coefficient Estimation

Parameters calculated for the infiltration parameter $S$ (Eq. 2) included:

$f_s$ is the stack coefficient, estimated as:

$$f_s = \sqrt{\frac{1 + \frac{R_{fac}}{2}}{3} x 1 - \left( \frac{x^{2}_{fac}}{2 - R^{2}_{fac}} \right)^{\frac{3}{2}} x g r a v x \frac{H}{T_{ref}}}$$  \[Equation S2\]

where $R_{fac}$ is the fraction of total leakage from the floors and ceilings (assumed to be 0.5 for homes built before 2011 and 0.25 for homes built on or after 2011 (4–6)); $X_{fac}$ is the difference between the leakage from a ceiling compared to that from a floor (assumed to be 0.25 (4)); $grav$ is the earth’s gravitational force ($9.8 \text{ m/s}^2$); and $T_{ref}$ is 298 K from the ideal gas law;

$f_w$ is the wind coefficient, estimated as:

$$C_{fac} x (1 - R_{fac})^{\frac{1}{3}} x A_{fac} x \left( \frac{H}{T_{ref}} \right)^{B_{fac}}$$  \[Equation S3\]

where $C_{fac}$ is set to reported values developed by LBNL based on local wind shielding from surrounding obstructions (shelter class) and house height (4). $A_{fac}$ and $B_{fac}$ are also factors developed by LBNL related to the geophysical terrain around the residence, and chosen based on shelter class.

Shelter class is a surrogate of wind shielding from surrounding obstructions, which we determined based on the land use classification in which each parcel was located. Shelter class was assigned based on the following: very low density residential (class 2), low density residential (class 3), medium density residential (class 4) and multi-family and high density residential (class 5).

| Table S4. Shelter class $B_{fac}$ and $A_{fac}$ Parameters used in AER calculation (7) |
|---------------------------------------------|-------------------|-------------------|
| Shelter Class | $B_{fac}$ | $A_{fac}$ | Description |
|----------------|-----------|-----------|-------------|
| 1              | 0.10      | 1.30      | No obstructions or local sheltering |
| 2              | 0.15      | 1.00      | Flat terrain with some isolated obstacles |
| 3              | 0.20      | 0.85      | Rural areas with low buildings, trees, etc. |
| 4              | 0.25      | 0.67      | Urban, industrial or forest area |
| 5              | 0.35      | 0.47      | Center of large city |

For $C_{fac}$ inputs see Tables 2.1-2.4 in Sherman et al. 1980 (7)
S4. Applying LBNL Model to Multi-Family Units

To apply the LBNL model to a given residential unit in a multi-family building, we assigned all a parcels categorized as multi-family a shelter class of “5.” This approach decreases the influence of the wind effect by maximizing the density of surrounding obstructions to account for apartment units having fewer externally-facing walls as compared to single family homes and duplex/triplex units. Shelter class definitions based on Sherman et al. (1980) can be found in Table S4.
| NL Category (n (%)) | Total (n=1 659 098) | Single Family (n=1 383 249) | Duplex/Triplex (n=207 722) | 4-8 Apartment Buildings (n=22 387) | >8 Apartment Buildings (n=45 740) |
|---------------------|----------------------|-------------------------------|-----------------------------|-----------------------------------|----------------------------------|
| Low Income          | 168 346              | 10.2                          | 90 458                      | 6.5                               | 8 970                            | 40.1                            | 7 825                          | 17.1                           |
| Conventional        | 1490 752             | 89.9                          | 1 292                       | 93.5                              | 146 629                           | 70.6                            | 13 417                         | 59.9                            | 37 915                          | 82.9                           |

| Shelter Class (n (%)) | 1 | 2 | 3 | 4 | 5 |
|------------------------|---|---|---|---|---|
| Low Income             | 210 236 | 12.7 | 205 813 | 14.9 | 4 423 | 2.1 |
| Conventional           | 283 991 | 17.1 | 275 642 | 19.9 | 8 349 | 4.0 |
| Low Income             | 546 564 | 32.9 | 509 177 | 36.8 | 37 387 | 18.0 |
| Conventional           | 477 791 | 28.8 | 392 617 | 28.4 | 85 174 | 41.0 |
| Low Income             | 140 516 | 8.5 | 72 389 | 34.9 | 22 387 | 100 |
| Conventional           | 145 072 | 8.5 | 72 389 | 34.9 | 22 387 | 100 |

Table S6. AER (h⁻¹) estimates complete case data

| AER (h⁻¹), winter | AER (h⁻¹), summer |
|-------------------|-------------------|
| mean              | 0.85              | 0.45              |
| SD                | 0.52              | 0.31              |
| 25th              | 0.47              | 0.23              |
| 50th              | 0.74              | 0.37              |
| 75th              | 1.10              | 0.57              |
| SD                | 0.44              | 0.27              |
| 25th              | 0.42              | 0.21              |
| 50th              | 0.67              | 0.34              |
| 75th              | 0.94              | 0.50              |
| SD                | 0.55              | 0.41              |
| 25th              | 1.09              | 0.45              |
| 50th              | 1.39              | 0.67              |
| 75th              | 1.78              | 0.97              |
| SD                | 0.41              | 0.20              |
| 25th              | 1.22              | 0.43              |
| 50th              | 1.42              | 0.54              |
| 75th              | 1.74              | 0.68              |
| SD                | 0.41              | 0.20              |
| 25th              | 0.50              | 0.20              |
| 50th              | 0.87              | 0.33              |
| 75th              | 1.29              | 0.54              |
| SD                | 0.48              | 0.22              |
| 25th              | 0.50              | 0.20              |
| 50th              | 0.87              | 0.33              |
| 75th              | 1.29              | 0.54              |
S5. Indoor PM$_{2.5}$ of ambient origin estimation procedure

The equation for indoor PM$_{2.5}$ concentrations of ambient origin ($C_{in}$ in $\mu g/m^3$) is:

$$C_{in} = \frac{P \cdot a}{a + k} \cdot C_{out}$$  \hspace{1cm} \text{[Equation S4]}

where $P$=penetration efficiency (dimensionless), $a$=AER ($h^{-1}$) assigned at a parcel level in the winter and the summer, calculated from Eq. 1, $k$= PM$_{2.5}$ decay rate ($h^{-1}$), and $C_{out}$ = outdoor ambient PM$_{2.5}$ concentration ($\mu g/m^3$). Based on literature-reported parameters estimated previously in Breen et al. (2014) from Environmental Protection Agency Panel study data, we assumed that $P$=0.84 and $k$=0.21 $h^{-1}$ (8). The $P$ and $k$ values estimated in Breen et al. (2015) are consistent with previously reported estimates over a variety of housing stock and geographic regions (9–11). As a sensitivity analysis, we test the lower and upper confidence limits of $P$ and $k$ estimated in Breen et al. (2015) (8). These values are reported in Table S7, found in Supplement. $C_{out}$ was assigned to each residential parcel from the corresponding 1 km$^2$ PM$_{2.5}$ data described above.
Table S7. Sensitivity analysis estimating indoor PM$_{2.5}$ of ambient origin for different values of penetration efficiency ($P$) and decay rate ($k$) as reported in Breen et al. (2015)$^a$(8)

|                  | Winter Ambient PM$_{2.5}$ (µg/m$^3$) | Summer Ambient PM$_{2.5}$ (µg/m$^3$) |
|------------------|-------------------------------------|-------------------------------------|
|                  | mean S D 25$^{1}$ h 75$^{1}$ h CV | mean S D 25$^{1}$ h 75$^{1}$ h CV |
|                  | n=1 659 098                         | n=1 659 098                         |
| Total            | Indoor (lower CI)                  | Indoor (lower CI)                  |
|                  | 5.7 0.9 5.1 6.3 0.1 6              | 3.9 0.9 3.3 4.5 0.22               |
|                  | Indoor (upper CI)                  | Indoor (upper CI)                  |
|                  | 6.0 1.3 5.1 7.0 0.2 2              | 3.7 1.2 2.9 4.6 0.31               |
|                  | Indoor (lower CI P, upper CI $k$)  | Indoor (lower CI P, upper CI $k$)  |
|                  | 4.8 1.1 4.1 5.5 0.2 2              | 3.0 0.9 2.3 3.6 0.31               |
|                  | Indoor (upper CI P, lower CI $k$)  | Indoor (upper CI P, lower CI $k$)  |
|                  | 7.1 1.1 6.4 7.9 0.1 6              | 4.9 1.1 4.2 5.7 0.22               |
| Single Family    | Indoor (lower CI)                  | Indoor (lower CI)                  |
| (n=1 383 249)    | 5.5 0.9 5.0 6.1 0.1 6              | 3.7 0.8 3.2 4.3 0.22               |
|                  | Indoor (upper CI)                  | Indoor (upper CI)                  |
|                  | 5.7 1.3 4.9 6.7 0.2 2              | 3.5 1.1 2.8 4.3 0.30               |
|                  | Indoor (lower CI P, upper CI $k$)  | Indoor (lower CI P, upper CI $k$)  |
|                  | 4.6 1.0 3.9 5.3 0.2 2              | 2.8 0.9 2.2 3.4 0.30               |
|                  | Indoor (upper CI P, lower CI $k$)  | Indoor (upper CI P, lower CI $k$)  |
|                  | 6.9 1.1 6.2 7.7 0.1 6              | 4.7 1.0 4.0 5.5 0.22               |
| Duplex/Triplex   | Indoor (lower CI)                  | Indoor (lower CI)                  |
| (n=207 722)      | 6.5 0.5 6.3 6.9 0.0 8              | 4.7 0.6 4.4 5.2 0.13               |
|                  | Indoor (upper CI)                  | Indoor (upper CI)                  |
|                  | 7.4 0.8 7.0 7.9 0.1 0              | 4.9 1.0 4.3 5.7 0.19               |
|                  | Indoor (lower CI P, upper CI $k$)  | Indoor (lower CI P, upper CI $k$)  |
|                  | 5.9 0.6 5.6 6.3 0.1 0              | 3.9 0.8 3.4 4.5 0.19               |
|                  | Indoor (upper CI P, lower CI $k$)  | Indoor (upper CI P, lower CI $k$)  |
|                  | 8.2 0.6 7.9 8.6 0.0 8              | 6.0 0.8 5.5 6.5 0.13               |
| 4-8 Apartment    | Indoor (lower CI)                  | Indoor (lower CI)                  |
| Buildings (n=22 387) | 6.6 0.5 6.4 6.9 0.0 8 | 4.7 0.6 4.3 5.0 0.12 |
|                  | Indoor (upper CI)                  | Indoor (upper CI)                  |
|                  | 7.5 0.7 7.2 8.0 0.1 0              | 4.7 0.8 4.3 5.2 0.17               |
|                  | Indoor (lower CI P, upper CI $k$)  | Indoor (lower CI P, upper CI $k$)  |
|                  | 6.0 0.6 5.7 6.4 0.1 0              | 3.8 0.6 3.4 4.2 0.17               |
|                  | Indoor (upper CI P, lower CI $k$)  | Indoor (upper CI P, lower CI $k$)  |
|                  | 8.3 0.7 8.0 8.7 0.0 8              | 5.8 0.7 5.5 6.3 0.12               |
| >8 Apartment     | Indoor (lower CI)                  | Indoor (lower CI)                  |
| Buildings (n=45 740) | 6.1 0.7 5.6 6.7 0.1 2 | 4.0 0.9 3.4 4.7 0.21 |
|                  | Indoor (upper CI)                  | Indoor (upper CI)                  |
|                  | 6.6 1.1 5.7 7.6 0.1 7              | 3.9 1.1 2.9 4.8 0.29               |
|                  | Indoor (lower CI P, upper CI $k$)  | Indoor (lower CI P, upper CI $k$)  |
|                  | 5.3 0.9 4.6 6.0 0.1 7              | 3.1 0.9 2.3 3.8 0.29               |
|                  | Indoor (upper CI P, lower CI $k$)  | Indoor (upper CI P, lower CI $k$)  |
|                  | 7.7 0.9 7.0 8.4 0.1 2              | 5.1 1.1 4.2 5.9 0.21               |

$^a$Jackknife estimates and 95% confidence limits estimated in Breen et al (2015) are as follows: $P$: 0.84 (0.74, 0.93); $k$: 0.21 h$^{-1}$ (0.13, 0.29 h$^{-1}$)
Figure S1. Sociodemographic characteristics of block groups containing the residential parcels with the lowest air exchange rates (AER) (=<10th %tile) in areas with the highest ambient PM$_{2.5}$ (=>90th %tile) versus block groups containing parcels with the highest AER and lowest PM$_{2.5}$. Source: US Census 2010.

Figure S2. Sociodemographic characteristics of block groups containing the residential parcels with the lowest air exchange rates (AER) in areas with the lowest ambient PM$_{2.5}$ (low-exposure, =<10th %tile) versus block groups containing parcels with the highest AER and PM$_{2.5}$ (high-exposure, =>90th %tile), stratified by urban/rural class. Source: US Census 2010.
Figure S3. Map of Eastern Massachusetts in 2010 showing distribution of a) summer air exchange rates at parcel level, b) summer PM$_{2.5}$ concentrations at parcel level, c) % Hispanic at block-group, and d) % median annual household income below $20,000 at block group.
References

1. Kloog I, Chudnovsky AA, Just AC, Nordio F, Koutrakis P, Coull BA, et al. A new hybrid spatio-temporal model for estimating daily multi-year PM2.5 concentrations across northeastern USA using high resolution aerosol optical depth data. Atmos Environ [Internet]. Elsevier Ltd; 2014;95:581–90. Available from: http://dx.doi.org/10.1016/j.atmosenv.2014.07.014

2. Schafer JL. Analysis of Incomplete Multivariate Data. New York, NY: Chapman and Hall; 1997.

3. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. Prev Sci. 2007;8(3):206–13.

4. Sarnat JA, Sarnat SE, Flanders WD, Chang HH, Mulholland J, Baxter L, et al. Spatiotemporally resolved air exchange rate as a modifier of acute air pollution-related morbidity in Atlanta. J Expo Sci Environ Epidemiol [Internet]. 2013;23(6):606–15. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23778234

5. US Department of Energy. Air Leakage Guide [Internet]. Building Technologies Program: Air Leakage Guide. 2011. Available from: https://www.energycodes.gov/sites/default/files/documents/BECP_Buidling Energy Code Resource Guide Air Leakage Guide_Sept2011_v00_lores.pdf

6. ASHRAE. American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE). Handbook of Fundamentals. Vol. 30329. 2009. 926 p.

7. Sherman MH, Grimsrud DT. Infiltration-Pressurization Correlation: Simplified Physical Modeling. ASHRAE Trans. 1980;86:778–807.

8. Breen MS, Long TC, Schultz BD, Williams RW, Richmond-Bryant J, Breen M, et al. Air pollution exposure model for individuals (EMI) in health studies: evaluation for ambient PM2.5 in Central North Carolina. Environ Sci Technol [Internet]. 2015;acs.est.5b02765. Available from: http://pubs.acs.org/doi/10.1021/acs.est.5b02765

9. Burke JM, Zufall MJ, Ozkaynak H. A population exposure model for particulate matter: case study results for PM2.5 in Philadelphia, PA. J Expo Anal Environ Epidemiol. 2001;11(6):470–89.

10. Meng QY, Turpin BJ, Korn L, Weisel CP, Morandi M, Colome S, et al. Influence of ambient (outdoor) sources on residential indoor and personal PM2.5 concentrations: analyses of RIOPA data. J Expo Anal Environ Epidemiol. 2005;15(1):17–28.

11. Thatcher TL, Lunden MM, Revzan KL, Sextro RG, Brown NJ. A Concentration Rebound Method for Measuring Particle Penetration and Deposition in the Indoor Environment. Aerosol Sci Technol. 2003;37(11):847–64.