DR. MARINA ZUSMAN (Orcid ID : 0000-0002-5513-8846)
DR. KIRSTEN KOEHLER (Orcid ID : 0000-0002-0516-6945)
PROF. JOEL DANIEL KAUFMAN (Orcid ID : 0000-0003-4174-9037)

Article type : Original Article

Title:

"Modeling Residential Indoor Concentrations of $PM_{2.5}$, NO_2, NO_x, and secondhand smoke in the Subpopulations and Intermediate Outcome Measures in COPD (SPIROMICS) Air Study"

A short running title: "SPIROMICS Indoor Exposure Modeling"

Authors[◊]:

Marina Zusman[1], Amanda J Gassett[1], Kipruto Kirwa[1], R. Graham Barr[2], Christopher B. Cooper[3], MeiLan K. Han[4], Richard E. Kanner[5], Kirsten Koehler[6], Victor E. Ortega[7], Robert Paine[8], Laura Paulin[9], Cheryl Pirozzi[10], Ana Rule[6], Nadia N. Hansel[11], Joel D. Kaufman[1,12,13] *

[◊] The authors have no conflicts of interest to disclose

[*] Corresponding author's e-mail address: joelk@uw.edu (J.D. Kaufman)

This is the author manuscript accepted for publication and has undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/INA.12760

[1]Department of Environmental and Occupational Health Sciences, University of Washington, Seattle, WA, United States
[2] Presbyterian Hospital, Columbia University Medical Center, New York, NY, United States
[3] Dept of Medicine, UCLA Sch of Medicine, Los Angeles, CA, United States
[4] Division of Pulmonary and Critical Care Medicine, University of Michigan, United States
[5] University of Utah Health Sciences Center, Department of Internal Medicine, Division of Respiratory, Critical Care & Occupational Medicine, Salt Lake City, Utah, United States
[6] Environmental Health and Engineering, Johns Hopkins University, Baltimore, MD, United States
[7] Department of Internal Medicine, Section on Pulmonary, Critical Care, Allergy and Immunologic Diseases Center for Precision Medicine. Wake Forest School of Medicine, Winston-Salem, NC, United States
[8] Division of Pulmonary Medicine, University Of Utah Hospital, Salt Lake City, UT, United States
[9] Pulmonary/Critical Care, Dartmouth Hitchcock Medical Center, Lebanon, NH, United States
[10] University Of Utah Hospital, Salt Lake City, UT, United States
[11] Department of Medicine, Johns Hopkins University, Baltimore, MD, United States
[12] Department of Medicine, University of Washington, Seattle, WA, United States
[13] Department of Epidemiology, University of Washington, Seattle, WA, United States

Acknowledgments

This publication was developed under a SPIROMICS AIR study that was supported by NIH NIEHS (R01ES023500). This project was also supported by NIH/NIEHS grants K23ES025781 and P30ES007033.

The SPIROMICS was supported by contracts from the NIH/NHLBI (HHSN268200900013C, HHSN268200900014C, HHSN268200900015C, HHSN268200900016C, HHSN268200900017C, HHSN268200900018C,

HHSN268200900019C, HHSN268200900020C), grants from the NIH/NHLBI (U01 HL137880 and U24 HL141762), and supplemented by contributions made through the Foundation for the NIH and the COPD Foundation from AstraZeneca/MedImmune; Bayer; Bellerophon Therapeutics; BoehringerIngelheim Pharmaceuticals, Inc..; Chiesi Farmaceutici S.p.A.; Forest Research Institute, Inc.; GlaxoSmithKline; Grifols Therapeutics, Inc.; Ikaria, Inc.; Novartis Pharmaceuticals Corporation; Nycomed GmbH; ProterixBio; Regeneron Pharmaceuticals, Inc.; Sanofi; Sunovion; Takeda Pharmaceutical Company; and Theravance Biopharma and Mylan.

The MESA Air study was developed under a STAR research assistance agreement, No. RD831697 (MESA Air) and RD83830001 (MESA Air Next Stage), awarded by the U.S Environmental Protection Agency. It has not been formally reviewed by the EPA. The views expressed in this document are solely those of the authors and the EPA does not endorse any products or commercial services mentioned in this publication.

The authors would like to thank the participants and investigators of the SPIROMICS and SPIROMICS Air studies for their critical contributions. Additional gratitude is for our fruitful cooperation with the Winston-Salem, Ann Arbor, San Francisco, Los Angeles, New York City, Salt Lake City and Baltimore clinical centers.

Data Availability

Author elects to not share data.

Research data used in the project include confidential and personal health information. Data availability are managed through the publication policies of the parent NHLBI SPIROMICS cohort study.

Conflict of Interest Statement

The authors have no conflicts of interest to disclose

Abstract

Increased outdoor concentrations of fine particulate matter ($PM_{2.5}$) and oxides of nitrogen (NO_2, NO_x) are associated with respiratory and cardiovascular morbidity in adults and children. However, people spend most of their time indoors and this is

particularly true for individuals with chronic obstructive pulmonary disease (COPD). Both outdoor and indoor air pollution may accelerate lung function loss in individuals with COPD, but it is not feasible to measure indoor pollutant concentrations in all participants in large cohort studies. We aimed to understand indoor exposures in a cohort of adults (SPIROMICS Air, the SubPopulations and Intermediate Outcome Measures in COPD Study, Air pollution). We developed models for the entire cohort based on monitoring in a subset of homes, to predict mean 2-week measured concentrations of $PM_{2.5}$, NO_2, NO_x, and nicotine, using home and behavioral questionnaire responses available in the full cohort. Models incorporating socioeconomic, meteorological, behavioral and residential information together explained about 60% of the variation in indoor concentration of each pollutant. Cross validated R^2 for best indoor prediction models ranged from 0.43 (NO_x) to 0.51 (NO_2). Models based on questionnaire responses and estimated outdoor concentrations successfully explained most variation in indoor $PM_{2.5}$, NO_x, NO_2, and nicotine concentrations.

Keywords: Indoor monitoring, Air pollutants, Prediction modeling, Residential behavior, Indoor exposure questionnaires, Exposure assessment

Practical implications:

- Questionnaire responses regarding home characteristics and residential behaviors explained a majority of the variation in indoor concentrations of key ambient air pollutants and secondhand smoke exposure.
- These model-based estimates can be used in epidemiological analyses in this cohort, taking into account remaining uncertainties.
- This approach and these models may be applicable to other populations with similar characteristics.

1. Introduction

Air pollution is a well-established risk factor for a variety of adverse health effects[1-2]. Epidemiological studies have found an association between air pollution levels and

increased risk of cardiovascular and respiratory disease. Increased concentrations of fine particulate matter ($PM_{2.5}$) and oxides of nitrogen (NO_2, NO_x) assessed in the ambient environment have been associated with adverse respiratory outcomes, including chronic obstructive pulmonary disease (COPD) [2-8].

While these relatively consistent associations have been seen with outdoor pollutant concentrations, the majority of adults, especially older adults, spend most of their time indoors. Individuals with COPD spend even more time at home than their age-matched counterparts. Both exposure to outdoor and to indoor air pollution may accelerate lung function loss in individuals with COPD and lead to exacerbations.

Spending most of residents' time at home and only a small part of time outside the house or in transit suggests that characterizing indoor exposures may improve our understanding of these relationships, since the severity of adverse respiratory outcomes linking air pollution depends on the concentration, frequency and duration of the personal exposure to each pollutant[1,9-11].

Researchers tend to classify residential indoor exposures as either the result of indoor-generated pollutants or the result of emissions from ambient origin. High outdoor concentrations can increase indoor concentrations of particulate pollution. Potential sources of indoor-generated air pollution include fuel-burning combustion processes, biologic agents, building and furnishing materials, tobacco smoke, and different heating/cooling devices. Indoor concentration can vary due to characteristics of the indoor environment[12]. Outdoor-derived pollutants are found in houses due to infiltration of these substances into the residential environment. The dynamics of outdoor-generated pollutants indoors, their concentration, and their reactivity are important factors for indoor pollution modeling that require detailed information on residence-specific characteristics, and resident behavior data that are typically unavailable, especially for a large multi-center cohort.

Since long-term individual indoor exposure measurement is a complex task which would be expensive for investigators and burdensome for participants, most studies directly measuring indoor exposure have small sample sizes[14-15] and the majority of

studies rely on outdoor exposure or modeled indoor concentrations[16-17], or examine personal exposure levels to specific air pollutants[18-22].

SPIROMICS Air, an ancillary study of NHLBI's Subpopulations and Intermediate Outcome Measures in COPD Study (SPIROMICS) multi-center prospective cohort study, was designed to examine the relationship between short and long term exposure to particulate matter with aerodynamic diameter less than 2.5 μm ($PM_{2.5}$), nitrogen dioxide (NO_2), nitrogen oxides (NO_x), sulfur dioxide (SO_2), ozone (O_3), black carbon (BC) and secondhand smoke (SHS) air pollutants, and disease progression in individuals with COPD (SO_2 and O_3 are not presented here). Participants were enrolled in twelve clinical centers across the United States (Winston-Salem, Ann Arbor, San Francisco, Los Angeles, New York City, Salt Lake City, Iowa City, Baltimore, Denver, Philadelphia, Birmingham, and Chicago) from 2012 to 2016 for SPIROMICS (see Figure 1). SPIROMICS Air was initiated in 2013[23].

Since it was not feasible to measure exposures for all 2,982 SPIROMICS participants we chose a modeling approach to assess each participant's long-term exposure to various air pollutants. In this study, we used indoor concentrations of $PM_{2.5}$, NO_2, NO_x, and nicotine measured in a subset of homes, estimates of ambient-origin infiltrated concentrations, and questionnaire-based behavioral and residence data (questionnaire responses are available for all SPIROMICS participants) to develop an individual-based model for residential indoor pollutant concentrations in SPIROMICS Air. We aimed to generalize each pollutant prediction model to the full SPIROMICS cohort in order to estimate each participant's indoor exposure to $PM_{2.5}$, NO_2, NO_x, and nicotine.

2. Methods

2.1 Study design and monitoring

SPIROMICS enrolled 2,982 participants aged 40-80 years at baseline from 12 clinical centers including 202 non-smokers without airflow obstruction, 944 smokers without airflow obstruction, and 1836 current and former smokers with COPD.

Figure 1 shows jittered residential locations of the 27-30 participants (total of 216) from each of seven SPIROMICS Air clinical centers (Winston-Salem, Ann Arbor, San

Francisco, Los Angeles, New York City, Salt Lake City and Baltimore) who were selected to participate in detailed individual exposure assessment campaigns between 2014 and 2016[23]. Characteristics of the full SPIROMICS cohort and SPIROMICS Air participants who completed the two-week monitoring are provided in Supplementary Materials (SM) Table S1.

Measuring pollutant concentrations in each of the 2982 residences was not possible, so we measured pollutant concentrations in a sample of SPIROMICS participants and developed prediction models using the information from home characteristics and residential behaviors questionnaires administered to all SPIROMICS participants. Convenience samples of approximately 30 participants that were available during pre-determined sampling periods in each area were selected, with COPD Stratum 3 and 4 participants prioritized for inclusion. Since this work builds on previous work in the MESA cohort, locations in non-MESA cities were observed in two contrasting seasons. Other locations were observed in one season. Pollutant measurements were made inside and outside those participants' homes [20, 23].

a) *Indoor, outdoor, and personal exposure sampling*

Two-week integrated paired indoor and outdoor measurements of $PM_{2.5}$, NO_2, and NO_x were collected at the 216 homes. In Ann Arbor, San Francisco and Salt Lake City measurements were collected during two campaigns to account for seasonal differences (winter and summer in Ann Arbor and Salt Lake City and fall and spring in San Francisco). Ogawa passive samplers were used to measure NO_2, NO_x, and O_3. $PM_{2.5}$ mass was measured by collecting particles on a 37-mm Teflon filter within a Harvard Personal Environmental Monitor (HPEM) [24]. Indoors, these were connected by silicone tubing to a TSI SidePak SP530 pump, programmed to run on a 50% duty cycle (alternating 5 minutes on and off). These pumps were also used to collect outdoor $PM_{2.5}$ measurements in residences without an outdoor sampling space (some apartments or condo units), where samplers were attached to an arm extended from a window. Outdoors, HPEMs were connected to MEDO VP0125 pumps, with a similar 50% duty cycle. Both types of pumps were adjusted to achieve a target flow rate of 1.8 liters per

minute. Nicotine 2-week indoor integrated measurements were collected using a sodium bisulfate passive badge.

Indoor sampling units were preferably placed in a room where participants spent the majority of their waking hours. Outdoor units were placed away from particle sources such as grills or smoking areas.

$PM_{2.5}$ mass concentrations were gravimetrically determined from Teflon filters weighed in a temperature and humidity controlled environment[25] using standard filter weighing procedures[26]. Ogawa passive samplers were used to measure NO_2 and NO_x using ion chromatography and ultraviolet spectroscopy. Concentrations of each pollutant were calculated using equations provided by Ogawa & Co.[24]. Nicotine concentrations were determined from passive air samplers using a sodium bisulfate–treated filter with a polycarbonate filter diffusion screen[27]. Nicotine content was analyzed using gas chromatography with a nitrogen phosphate detector. The LOD for the passive air nicotine badges was 0.021 μg/mL[28].

Additional information about pollutant monitoring and analytical methods used have been previously described in Hansel et al, (2017)[23] and Cohen et al, (2009)[20].

b) *Temperature and relative humidity*

Indoor temperature and humidity were monitored using Onset HOBO data loggers[29]. Outdoor temperature and relative humidity (RH) data were obtained from government sources[30] at meteorological stations nearest to each study clinic. The data were averaged over the 2-week periods that matched the 2-week monitoring period (with about 5-10 days' variability in the start/end dates) to provide seasonal variation across sites.

c) *Questionnaires*

We integrated information from several instruments. All participants answered questions regarding smoking behaviors, including SHS exposure. These were administered annually for the 3-years of follow, and the questionnaire instruments are available online[31], whether or not they participated in the home monitoring. A home information questionnaire on residence characteristics, residential behaviors, and

approximate amounts of time spent indoors, outdoors, and in transit was completed by 2054 SPIROMICS participants once during 2014-2017 (see Table 1).

Field technicians completed a home inspection form for participants included in the air monitoring subset to verify the presence of appliances, window types, and other home characteristics. During each 2-week monitoring period, participants logged cooking activity, equipment use, amount of time spent indoors and outdoors, and any combustion. Research staff deploying home monitoring equipment also assessed the presence of specific appliances in the home and assessed relevant environmental characteristics of the residence.

d) *Neighborhood- scale socio-economic information*

Percent of owner-occupied housing units, education level of adults (age 25+), and median household income data were obtained at the block group level (an area that typically encompasses between 600 and 3,000 people), based on the U.S. Census and the American Community Survey sources [32].

e) *Estimation of $PM_{2.5}$ infiltration*

To estimate the level of ambient-derived $PM_{2.5}$ concentrations inside each home, we used a model for outdoor $PM_{2.5}$ that we had previously developed and validated in the MESA Air study[33]. In that model, the proportion of ambient-origin $PM_{2.5}$ that infiltrates into dwellings was estimated based on paired indoor-outdoor filters with elemental sulfur as a tracer. We employed infiltration coefficients from that model to calculate levels of indoor $PM_{2.5}$ in SPIROMICS Air households estimated to be of ambient origin.

2.2 Exposure modeling

The exposure assessment design for SPIROMICS Air has been described in Hansel et al, 2017 and has already been successfully used in MESA Air study[20, 33-34]. Briefly, cohort-specific air monitoring was conducted to support the development of air pollution prediction models that can be generalized to the study population. These models are based on spatio-temporal air pollution prediction methods that incorporate the study-specific outdoor monitoring data[23,34]. The SPIROMICS Air exposure prediction

modeling structure for $PM_{2.5}$, NO_2, NO_x, and SHS, with the available data sources, is shown in Figure 2. The left wing of the Figure 2 shows the outdoor predicted exposure modeling with the data that are incorporated into outdoor prediction models for $PM_{2.5}$, NO_2, NO_x, and ozone, although ozone is not further discussed in this paper. The right wing of this structure demonstrates indoor prediction model with the available data resources (the scope of this paper is highlighted by the content of the red rectangle). Indoor and outdoor measurements for indoor exposure prediction modeling were collected inside and outside participants' homes. This paper focuses on indoor exposure modeling for $PM_{2.5}$, NO_2, NO_x and SHS using data from questionnaires, predicted infiltration estimates, and neighborhood socioeconomic data.

2.2.1 Indoor exposures

Indoor pollutant concentration can be calculated as follows:

(1)

$$C^I = C^{IG} + C^A F_{inf}$$

where:

C^I - total indoor pollution concentration; C^{IG} - indoor generated pollution concentration; C^A - ambient (outdoor) pollution concentration; F_{inf} - pollutant specific infiltration rate.

SHS exposures were assessed based on responses from smoking related questions from all available questionnaires and then validated with air nicotine measurements.

Unlike $PM_{2.5}$, infiltration models for NO_2, NO_x, and SHS were not available from MESA Air, and to account for potential infiltration of oxides of nitrogen, we modeled the total measured indoor concentration (C^I) using the ambient concentration (C^A) as an input to the prediction model. We assumed that levels of SHS are dominated by indoor-generated sources.

2.3 Analytical methods and modeling decisions

Indoor prediction models were developed using the 2-week time-integrated indoor and outdoor (I/O) data (that were obtained during 2014-2016), and responses from available questionnaires that were mentioned above.

Multivariate linear regression models were developed for each pollutant.

2.3.1 Predictor selection

We built prediction models using a forward stepwise linear regression procedure. The response variable in all models was the measured pollutant-specific 2-week time-integrated indoor concentration (native-scale or transformed). As indoor concentrations vary across cities, seasons, and by neighborhood socioeconomic status, such variables as "city", "temperature" or/and "relative humidity" and several census-derived socioeconomic factors were always included in our models. Additionally, estimated infiltrated concentrations of $PM_{2.5}$ and measured outdoor concentrations of oxides of nitrogen were included in the respective models.

We examined 148 variables from both questionnaires (see SM, Table S2). We excluded questions with fewer than 10 responses and used only questions that had been posed to most of the SPIROMICS participants. We explored each variable and performed various transformations of continuous variables (e.g. quadratic, square root, logarithmic, and polynomial) to satisfy the assumptions of linear modeling.

Since home characteristics and behaviors were assessed in up to three ways (questionnaires completed by all SPIROMICS participants in the course of interviews by clinic staff, diaries completed by participants in the home monitoring study, and observations of the home by the study technician), we evaluated agreement between these sources (see SM, Table S3 and description of the analysis).

2.3.2 Model development and statistical methods

We built prediction model using two approaches. In a first "A" approach, starting with the mandatory variables outlined above, additional covariates were evaluated by assessing significant stepwise improvements in R^2 and leaving out predictors that contributed less than 0.01 to the R^2. In a second "B" approach, the residuals of the best model from the aforementioned stepwise method were first estimated. Then the variables that were excluded were explored individually against these residuals by forward stepwise regression. Any variables contributing improvements of more than 0.01 to the R^2 were selected, creating a second model with additional predictors. Post-

modeling diagnostics were performed for each model to assess collinearity check, outliers and high leverage and influential points. Outlier were identifies using the Bonferroni-adjusted outlier test finding the largest absolute studentized residual[35]. Interactions between variables of the model were also analyzed. Model performance was assessed using 10-fold cross-validation (CV).

To prevent inappropriate extrapolation, predictions for the full cohort were generated to eliminate models that produced predictions well outside the range of the observations (see SM, Figure S1).

All statistical analyses were conducted in R version 3.6.0.

3. Results

The 2-week time-integrated indoor and outdoor (I/O) measurements were collected from 216 homes for NO_2 and NO_x, from 201 homes for $PM_{2.5}$ and from 205 homes for nicotine (Table 1). After inclusion of two monitoring campaigns at homes in Ann Arbor, San Francisco, and Salt Lake City, and removal of missing data, there were 287 paired I/O observations for NO_2 and NO_x, 263 for indoor and 264 for outdoor $PM_{2.5}$, and 265 for indoor nicotine available for further analysis. Some samples were invalidated for duration, flow rate, or physical damage.

The indoor and outdoor concentrations by pollutant and city are shown in Figure 3a and additionally by season in SM, Table S4. On average, indoor measurements were higher and more variable than outdoor measurements. Between paired measurements, 54% of $PM_{2.5}$ measurements, 53% of NO_2 and 72% of NO_x measurements were higher indoors than outdoors. However, average indoor measurements for NO_2 in Los Angeles (LA) and Salt Lake City (SLC) were lower than outdoors. The highest indoor $PM_{2.5}$ values were found in Winston-Salem in spring, while the highest outdoor concentrations were found in SLC in summer. These values were unexpectedly higher than the winter $PM_{2.5}$ outdoor measurements in SLC area, due to an uncommon pollution event during our particular sampling period.

The highest average indoor and outdoor NO_x and NO_2 concentrations were measured in LA in fall (see SM, Table S4). Of 265 nicotine samples collected, only 61

measurements were above the level of detection (LOD) of 0.013 μg/m^3 of which participants in 29 of these 61 homes were current smokers. Out of the 61 homes with nicotine concentrations above LOD, 21 residents were living in the same household with another smoker and 21 were not current smokers nor living with smoker. Out of the residences with detected nicotine levels, 25 were living in "single family" type of house, 3 in row house, townhouse, duplex or triplex types of houses and 5 in apartments.

Indoor/outdoor temperature and RH are shown in Figure 3b. Outdoor average temperature (15.5±7.8°C) was more variable across sites than indoor temperature (22.6±2.1°C); and indoor temperature was largely consistent between cities. Since temperature was highly correlated with RH, the one that contributed larger improvements in R^2 was chosen for each model.

Table 2 demonstrates the results for the best indoor prediction models of $PM_{2.5}$, NO_2, NO_x and nicotine using the two model-building approaches "A" and "B". Linear and square root models provided the best fit for indoor $PM_{2.5}$ (Model A1 - best linear model; Models A2-A3 - two best square root models with interactions with/without outliers using first "A" approach; and Models B4-B5 - two additional best square root models with/without outliers using second "B" approach) are described in Table 2.

We observed a very low correlation between indoor and outdoor measurements of $PM_{2.5}$ (Pearson's r=0.01, p-value=0.91, Table 2). There were higher correlations between indoor and outdoor NO_2 and NO_x measurements (Pearson's r=0.34, p-value<0.001 and r=0.41, p-value<0.001, respectively). The ranges of indoor NO_2 and NO_x concentrations are wide, especially for NO_x (1.3 - 477.7 ppb), suggesting a few extreme outliers (there were two indoor NO_x values larger than 200 ppb).

Models based on log-transformed measured indoor concentrations performed better than linear for NO_2, NO_x and nicotine (Table 2). The best logarithmic prediction model fit for NO_2 using the first approach explained 63% of variation in concentrations, and the second approach, in which two additional predictors were added, yielded a model with slightly improved performance (for NO_2 Model B4 R^2=0.63, and Model B5 R^2=0.65). For NO_x the best prediction logarithmic model fits explained 58% of the variation in concentrations (Model A3 based on the first approach) and for nicotine 61%.

The second approach did not result in selection of any additional predictors both for NO_x and for nicotine.

Analyzing outlying observations provided some important insights for the $PM_{2.5}$. Model A1, suggesting that these outliers were explained by two residents using wood fireplaces during periods with low outdoor temperatures. Hence, applying a model with an interaction between "the use of wood fireplace" and temperature improved substantially Model A1's fit (cross validation (CV) R^2 rose from 0.30 to 0.44 see $PM_{2.5}$ Models A1-A2, Table 2). Finally, two variables associated with window-opening residential behavior factors, and the interaction between opening windows and smoking factors during second approach improved the model's performance to an R^2 of =0.66 (CV R^2=0.45, RMSE=1.06 μg/m^3; see Model B5, Table 2).

Table 3 shows the goodness of fit from the best prediction model based on the first approach (Model A3 per pollutant), and the coefficients of the predictors selected via forward stepwise regression. The full list of variables explored during model building is shown in the SM, Table S2.

The reliability and correlation of data for the subset of variables with multiple information sources available is shown in the SM, Tables S4, S5a and S5b.

4. Discussion

We developed residential indoor exposure prediction models for measured $PM_{2.5}$, NO_2, NO_x, and nicotine based on home characteristics and residential behavior information in a well-characterized cohort. Using socioeconomic, meteorological, behavioral, residential and ambient-pollutant concentration data obtained from questionnaires, direct observations and measurements, we built models that explained about 60% of the variability in measured indoor pollutant concentrations.

$PM_{2.5}$

It is well-established that both indoor and outdoor sources contribute to indoor $PM_{2.5}$ concentrations. Coal and wood burning for cooking and heating, the use of candles, and tobacco smoke increase indoor $PM_{2.5}$ concentrations[36-40], and outdoor particulates infiltrate indoors variably based on the tightness of the home environment and natural

and/or mechanical ventilation and air cleaning systems[41]. While the mean average $PM_{2.5}$ I/O ratio 1.74±2.14 μg/m^3 in the current study was higher than typically reported (Wichmann et al. (2010)[8], Geller et al. (2002)[42] and Jones et al. (2000)[43] report from 1.00 to 1.02 mean $PM_{2.5}$ I/O ratio in their studies), our median $PM_{2.5}$ I/O ratio (1.08) was close to these values. We found little correlation between $PM_{2.5}$ indoor and outdoor measurements. This suggests that in our sampled homes, indoor sources of fine particles are the major source of variation in indoor concentrations, rather than infiltrated ambient particles. As expected, we observed significant reduction in indoor $PM_{2.5}$ with the use of an air cleaner/filter. We also observed that smoking and the use of wood fireplace were significantly correlated with the concentration of indoor fine particles. Our results are consistent with Meng et al (2010)[39] who observed an increase in $PM_{2.5}$ mass during wood burning, woodworking and tobacco smoke. We also observed that parking more than two cars in the attached garage significantly increased the indoor concentrations of ambient particles compared to the residences that had no garage. Additionally, we observed significant reduction in indoor ambient particles for residents living in second floor or higher compared to residents living most of their time in basement and ground floor. Unlike several other studies[44-46], we didn't find that cooking-related variables were significantly predictive of indoor $PM_{2.5}$.

NO_2 and NO_x

Consistent with prior research, the main indoor sources of NO_2 and NO_x were cooking-related factors such as gas stove usage and frequency of cooking[47-52]. Hansel et al. (2008)[23] report positive association between indoor NO_2 and the use of gas heaters, gas stoves and space heaters. Our results are consistent with these findings, suggesting that both the use of gas oven and the frequency of stove cooking were significantly predictive of higher NO_2 and NO_x indoor concentrations. However, neither the use of gas heating nor gas space heater appliances were associated with increased concentrations of oxides of nitrogen. The use of forced air ventilation significantly reduced indoor concentrations both in NO_2 and NO_x models.

We found that the presence of pilot lights on clothes dryer significantly increased oxides of nitrogen levels, consistent with prior report[49]. The presence of pilot lights on

water heater were associated with a decrease of NO_x but not NO_2 indoor concentrations. We observed positive associations between oxides of nitrogen and the use of an attached garage for parking, which has been previously described[53].

Among a variety of variables examined, we found that indoor NO_x was associated only with one smoking related factor - the most intensive category of smoking activity. Previous investigators had found several tobacco variables to predict indoor NO_2 levels[47, 52].

As in other studies investigating exposure from various air pollutants, we observed significant reduction in oxides of nitrogen indoor concentration with the use of air cleaner appliances and increased window-opening behavior. Opening windows in the summer for a few days a month (compared to not opening at all) was associated with significantly reduced indoor NO_2 and NO_x levels. The presence of double pane windows was associated with increased indoor NO_2 but not NO_x levels.

Somewhat surprisingly, the relationships between outdoor pollutant concentrations and indoor concentrations of oxides of nitrogen were stronger than the associations for indoor concentrations of $PM_{2.5}$. Not only were outdoor NO_x and NO_2 predictions associated with higher indoor concentrations of these pollutants, but higher outdoor concentrations of $PM_{2.5}$ were also significantly associated with increased indoor NO_2 and indoor NO_x concentrations. We can only speculate about this relationship; it is possible that some outdoor sources of variation in $PM_{2.5}$ concentration, such as proximity to traffic sources, are also predictors of indoor concentrations of oxides of nitrogen.

Nicotine

We modeled indoor nicotine concentrations in order to understand secondhand smoke exposures. As expected, indoor nicotine concentrations were strongly related to smoking habits. Active cigarette smoking of any amount (as of one month prior to sampling), more intense smoking habits (more than 20 cigarettes per day), and permitting smoking in all rooms each increased the levels of indoor nicotine.

We found that natural ventilation (e.g. reporting opening windows in the summer), significantly reduced the level of indoor nicotine concentrations. In this study, the use of a radiator for heating was significantly and positively associated with indoor nicotine concentrations, unlike other sources of heating. This might be related to the fact that people tend to smoke indoors when temperature decreases but our data can't clearly confirm that.

Our ability to predict indoor nicotine levels based on the questionnaire responses for smoking was not as high as we might anticipate. We found some inconsistencies between indoor measured nicotine concentrations, which we expect to only be measurable when smoking occurs in the home, and participants' answers about smoking behavior at home (see SM, Table S5a and S5b). For example, 33 participants responded that no one smoked in their house in the past year while indoor nicotine concentrations were above the detection limit. There are several possible reasons for this, including the potential temporal mismatch between questionnaire completion and sampling (for example, the average difference between the start date of I/O sampling and the date of RDS questionnaire response is 956 days, and between I/O sampling and HIQ response is 185 days, see SM, Figure S2), a persistence of indoor nicotine after smoking cessation, as well as a bias to report what is believed to be the desired response.

General Issues and Limitations

Our indoor exposure prediction models are meant to predict exposure for every participant by incorporating high level resolution data (indoor and outdoor measurements of $PM_{2.5}$, NO_2, and NO_x collected in a subset plus socioeconomic, meteorological, behavioral and residential information from all participants) from seven different regions to improve spatial and seasonal variation. Similar prediction models have not previously been available. These models can facilitate exposure characterization of research cohorts with much less effort and expense than monitoring in all participants. Future studies may find this indoor exposure assessment method applicable for use in other populations while assessing indoor exposure with home characteristic and residential behavior questionnaire data.

We found reasonably high R^2 estimates of best model fit, though cross validation analysis showed lower R^2 values. Lower R^2 estimates in cross-validation approaches are not surprising since these approaches can create substantial areas of missing data and hence lowered effective sample size in these approaches. The lower cross-validated performance metrics also suggest some degree of over-fitting of our models.

Linear and square-root transformation models produced similar performance results for the $PM_{2.5}$ model, but the square root transformation model was preferred because it does not produce negative predictions and thus showed better generalizability to the full SPIROMICS population.

Several predictors that we found to be important in model fit contained substantial missing information that could not be estimated or imputed easily; these include age of building and outdoor $PM_{2.5}$ measurements. Additionally, we created models that included all cities and seasons under study but since each city had a relatively small number of indoor samples and seasonal representation was incomplete, we are limited in our ability to compare regional and seasonal contrasts in this dataset. Some individual 2-week periods may have included non-representative weather or pollution events which were beyond our control. For example, high concentrations of $PM_{2.5}$ observed in Salt Lake City may be explained by dust storms which are occasionally observed from the spring to fall, originating from the Great Salt Lake, and wildfires which increase ambient $PM_{2.5}$ concentration in this area. This could impact the generalizability of the model in the affected city.

The SPIROMICS full cohort is somewhat younger than the subgroup with indoor air monitoring studied here. This may reduce the generalizability of our models.

Conclusions

In a subset of participants representative of a multi-city cohort, we developed models that explained most variation in indoor $PM_{2.5}$, NO_x, NO_2, and estimated secondhand smoke concentrations using a set of variables available in most of the members of this major longitudinal cohort of chronic pulmonary disease. These models can be used in the full SPIROMICS cohort, may be of use in other epidemiological projects, and can be leveraged to study the lung health effects of several important indoor air contaminants.

5. References

1. U.S. Environmental Protection Agency. Office of Chemical Safety and Pollution Prevention. Sustainable Futures: P2 Framework Manual. EPA-748-B12-001, 2012.
2. Brunekreef B, Holgate ST. Air pollution and health. *The Lancet*. 2002 Oct 19;360(9341):1233-42.
3. Andersen ZJ, Hvidberg M, Jensen SS, Ketzel M, Loft S, Sørensen M, Tjønneland A, Overvad K, Raaschou-Nielsen O. Chronic obstructive pulmonary disease and long-term exposure to traffic-related air pollution: a cohort study. *American Journal of Respiratory and Critical Care Medicine*. 2011 Feb 15;183(4):455-61.
4. Hooper LG, Young MT, Keller JP, Szpiro AA, O'Brien KM, Sandler DP, Vedal S, Kaufman JD, London SJ. Ambient air pollution and chronic bronchitis in a cohort of US women. *Environmental Health Perspectives*. 2018 Feb 6;126(2):027005.
5. Medina-Ramon M, Zanobetti A, Schwartz J. The effect of ozone and PM10 on hospital admissions for pneumonia and chronic obstructive pulmonary disease: a national multicity study. *American Journal of Epidemiology*. 2006 Jan 27;163(6):579-88.
6. Pope III CA, Burnett RT, Thun MJ, Calle EE, Krewski D, Ito K, Thurston GD. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA*. 2002 Mar 6;287(9):1132-41.
7. Wang M, Aaron CP, Madrigano J, Hoffman EA, Angelini E, Yang J, Laine A, Vetterli TM, Kinney PL, Sampson PD, Sheppard LE, Szpiro AA, Kipruto K, Smith B, Lederer DJ, Diez-Roux AV, Vedal S, Kaufman JD, Barr RG. Association between long-term exposure to ambient air pollution and change in quantitatively assessed emphysema and lung function. *JAMA*. 2019 Aug 13;322(6):546-56.
8. Wichmann J, Lind T, Nilsson MM, Bellander T. $PM_{2.5}$, soot and NO_2 indoor–outdoor relationships at homes, pre-schools and schools in Stockholm, Sweden. *Atmospheric Environment*. 2010 Nov 1;44 (36):4536-44.

9. Jenkins PL, Phillips TJ, Mulberg EJ, Hui SP. Activity patterns of Californians: use of and proximity to indoor pollutant sources. *Atmospheric Environment*. Part A. General Topics. 1992; 1;26(12):2141-8.
10. Klepeis NE, Nelson WC, Ott WR, Robinson JP, Tsang AM, Switzer P, Behar JV, Hern SC, Engelmann WH. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. *Journal of Exposure Science and Environmental Epidemiology*. 2001 Jul;11(3):231.
11. Spalt EW, Curl CL, Allen RW, Cohen M, Adar SD, Stukovsky KH, Avol E, Castro-Diehl C, Nunn C, Mancera-Cuevas K, Kaufman JD. Time–location patterns of a diverse population of older adults: the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Journal of Exposure Science and Environmental Epidemiology*. 2016 Jun; 26(4):349.
12. U.S. Environmental Protection Agency. Indoor Air Quality (IAQ). Available at https://www.epa.gov/indoor-air-quality-iaq/introduction-indoor-air-quality. Accessed March 6th, 2018.
13. Liu L, Ruddy T, Dalipaj M, Poon R, Szyszkowicz M, You H, Dales RE, Wheeler AJ. Effects of indoor, outdoor, and personal exposure to particulate air pollution on cardiovascular physiology and systemic mediators in seniors. *Journal of Occupational and Environmental Medicine*. 2009 Sep 1;51(9):1088-98.
14. Norris C, Goldberg MS, Marshall JD, Valois MF, Pradeep T, Narayanswamy M, Jain G, Sethuraman K, Baumgartner J. A panel study of the acute effects of personal exposure to household air pollution on ambulatory blood pressure in rural Indian women. *Environmental Research*. 2016 May 1;147:331-42.
15. Diapouli E, Chaloulakou A, Koutrakis P. Estimating the concentration of indoor particles of outdoor origin: A review. *Journal of the Air & Waste Management Association*. 2013 Oct 1;63(10):1113-29.
16. Hodas N, Turpin BJ, Lunden MM, Baxter LK, Özkaynak H, Burke J, Ohman-Strickland P, Thevenet-Morrison K, Kostis JB, Rich DQ. Refined ambient PM 2.5 exposure surrogates and the risk of myocardial infarction. *Journal of Exposure Science and Environmental Epidemiology*. 2013 Nov;23(6):573.
17. Magalhaes S, Baumgartner J, Weichenthal S. Impacts of exposure to black carbon, elemental carbon, and ultrafine particles from indoor and outdoor

sources on blood pressure in adults: A review of epidemiological evidence. *Environmental Research*. 2018 Feb 28;161:345-53.

18. Ando M, Katagiri K, Tamura K, Yamamoto S, Matsumoto M, Li YF, Cao SR, Ji RD, Liang CK. Indoor and outdoor air pollution in Tokyo and Beijing supercities. *Atmospheric Environment*. 1996 Mar 1;30(5):695-702.
19. Belli AJ, Bose S, Aggarwal N, DaSilva C, Thapa S, Grammer L, Paulin LM, Hansel NN. Indoor particulate matter exposure is associated with increased black carbon content in airway macrophages of former smokers with COPD. *Environmental Research*. 2016 Oct 1;150:398-402.
20. Cohen MA, Adar SD, Allen RW, Avol E, Curl CL, Gould T, Hardie D, Ho A, Kinney P, Larson TV, Sampson P. Approach to estimating participant pollutant exposures in the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Environmental Science & Technology*. 2009 Apr 2;43(13):4687-93.
21. Clayton CA, Perritt RL, Pellizzari ED, Thomas KW, Whitmore RW, Wallace LA, Ozkaynak H, Spengler JD. Particle Total Exposure Assessment Methodology (PTEAM) study: distributions of aerosol and elemental concentrations in personal, indoor, and outdoor air samples in a southern California community. *Journal of Exposure Analysis and Environmental Epidemiology*. 1993;3(2):227-50.
22. Dockery DW, Pope CA, Xu X, Spengler JD, Ware JH, Fay ME, Ferris Jr BG, Speizer FE. An association between air pollution and mortality in six US cities. *New England Journal of Medicine*. 1993 Dec 9;329(24):1753-9.
23. Hansel NN, Paulin LM, Gassett AJ, Peng RD, Alexis N, Fan VS, Bleecker E, Bowler R, Comellas AP, Dransfield M, Han MK...and Kaufman JD. Design of the Subpopulations and Intermediate Outcome Measures in COPD (SPIROMICS) AIR Study. *BMJ open respiratory research*. 2017 Aug 1;4(1):e000186.
24. Ogawa and Company, U.S.A., Inc. NO, NO_2, NO_x, and SO_2 sampling protocol using the Ogawa sampler. Pompano Beach, FL: 1998.
25. Allen RW, Box M, Liu LJ, Larson TV. A cost-effective weighing chamber for particulate matter filters. J. *Air Waste Manag. Assoc.* 2001;51(12):1650–1653.

26. U.S. Environmental Protection Agency. Quality assurance guidance document 2.12 monitoring $PM_{2.5}$ in ambient air using designated reference or class I equivalent methods. In: N.E.R.L., editor. U.S. Environmental Protection Agency. 1998.
27. Hammond SK, Leaderer BP. A diffusion monitor to measure exposure to passive smoking. *Environmental Science & Technology*. 1987 May;21(5):494-7.
28. Torrey CM, Moon KA, D'Ann LW, Green T, Cohen JE, Navas-Acien A, Breysse PN. Waterpipe cafes in Baltimore, Maryland: Carbon monoxide, particulate matter, and nicotine exposure. *Journal of Exposure Science and Environmental Epidemiology*. 2015 Jul;25(4):405.
29. Onset Computer Corporation, 2014-2016. Available at: https://www.onsetcomp.com. Accessed on May 6th, 2020.
30. National Oceanic and Atmospheric Administration (NOAA), National Climatic Data Center (NCDC) 2014-2016. Available at: http://www7.ncdc.noaa.gov/CDO/dataproduct. Accessed on March 6th, 2018.
31. SubPopulations and InteRmediate Outcome Measures in COPD Study (SPIROMICS), 2010 -2019. Available at: https://www2.cscc.unc.edu/spiromics. Accessed on March 6th, 2020.
32. U.S. Census Bureau, 2010. Available at: https://www.census.gov. 2010 data was retrieved through the National Historical Geographic Information System5 (http://www.nhghis.org/) and from Social Explorer (http://www.socialexplorer.com/).
33. Allen RW, Adar SD, Avol MC, Curl CL, Larson T, Liu LJ, Sheppard L, Kaufman JD. Modeling the residential infiltration of outdoor $PM_{2.5}$ in the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Environmental Health Perspectives*. 2012 Jun;120(6):824.
34. Keller JP, Olives C, Kim SY, Sheppard L, Sampson PD, Szpiro AA, Oron AP, Lindström J, Vedal S, Kaufman JD. A unified spatiotemporal modeling approach for predicting concentrations of multiple air pollutants in the multi-ethnic study of atherosclerosis and air pollution. *Environmental Health Perspectives*. 2015 Apr;123(4):301.

35. Fox J, Friendly GG, Graves S, Heiberger R, Monette G, Nilsson H, Ripley B, Weisberg S, Fox MJ, Suggests MA. *The car package*. R Foundation for Statistical Computing. 2007 Oct 27.
36. Canha N, Almeida SM, do Carmo Freitas M, Wolterbeek HT, Cardoso J, Pio C, Caseiro A. Impact of wood burning on indoor $PM_{2.5}$ in a primary school in rural Portugal. *Atmospheric Environment*. 2014 Sep 1;94:663-70.
37. Höllbacher E, Ters T, Rieder-Gradinger C, Srebotnik E. Emissions of indoor air pollutants from six user scenarios in a model room. *Atmospheric Environment*. 2017 Feb 1;150:389-94.
38. McCormack MC, Breysse PN, Hansel NN, Matsui EC, Tonorezos ES, Curtin-Brosnan J, D'Ann LW, Buckley TJ, Eggleston PA, Diette GB. Common household activities are associated with elevated particulate matter concentrations in bedrooms of inner-city Baltimore pre-school children. *Environmental Research*. 2008 Feb 1;106(2):148-55.
39. Meng QY, Spector D, Colome S, Turpin B. Determinants of indoor and personal exposure to $PM_{2.5}$ of indoor and outdoor origin during the RIOPA study. *Atmospheric Environment*. 2009 Nov 1;43(36):5750-8.
40. Northcross AL, Hammond SK, Canuz E, Smith KR. Dioxin inhalation doses from wood combustion in indoor cookfires. *Atmospheric Environment*. 2012 Mar 1;49:415-8.
41. Martins NR, da Graça GC. Impact of $PM_{2.5}$ in indoor urban environments: A review. *Sustainable Cities and Society*. 2018 Oct 1;42:259-75.
42. Geller MD, Chang M, Sioutas C, Ostro BD, Lipsett MJ. Indoor/outdoor relationship and chemical composition of fine and coarse particles in the southern California deserts. *Atmospheric Environment*. 2002 Feb 1;36(6):1099-110.
43. Jones NC, Thornton CA, Mark D, Harrison RM. Indoor/outdoor relationships of particulate matter in domestic homes with roadside, urban and rural locations. *Atmospheric Environment*. 2000 Jan 1;34(16):2603-12.
44. Ji W, Zhao B. Contribution of outdoor-originating particles, indoor-emitted particles and indoor secondary organic aerosol (SOA) to residential indoor $PM_{2.5}$

concentration: A model-based estimation. *Building and Environment*. 2015 Aug 1;90:196-205.

45. Li T, Cao S, Fan D, Zhang Y, Wang B, Zhao X, Leaderer BP, Shen G, Zhang Y, Duan X. Household concentrations and personal exposure of $PM_{2.5}$ among urban residents using different cooking fuels. *Science of the Total Environment*. 2016 Apr 1;548:6-12.
46. Olson DA, Burke JM. Distributions of $PM_{2.5}$ source strengths for cooking from the Research Triangle Park particulate matter panel study. *Environmental Science & Technology*. 2006 Jan 1;40(1):163-9.
47. Baxter LK, Clougherty JE, Laden F, Levy JI. Predictors of concentrations of nitrogen dioxide, fine particulate matter, and particle constituents inside of lower socioeconomic status urban homes. *Journal of Exposure Science and Environmental Epidemiology*. 2007 Aug;17(5):433.
48. Ielpo P, Mangia C, Marra GP, Comite V, Rizza U, Uricchio VF, Fermo P. Outdoor spatial distribution and indoor levels of NO_2 and SO_2 in a high environmental risk site of the South Italy. *Science of the Total Environment*. 2019 Jan 15;648:787-97.
49. Lee K, Yanagisawa Y, Spengler JD, Fukumura Y, Billick IH. Classification of house characteristics in a Boston residential nitrogen dioxide characterization study. *Indoor Air*. 1996 Sep;6(3):211-6.
50. Levy JI. Impact of residential nitrogen dioxide exposure on personal exposure: an international study. *Journal of the Air & Waste Management Association*. 1998 Jun 1;48(6):553-60.
51. Paulin LM, Williams DA, Peng R, Diette GB, McCormack MC, Breysse P, Hansel NN. 24-h Nitrogen dioxide concentration is associated with cooking behaviors and an increase in rescue medication use in children with asthma. *Environmental Research*. 2017 Nov 1;159:118-23.
52. Tunno BJ, Dalton R, Cambal L, Holguin F, Lioy P, Clougherty JE. Indoor source apportionment in urban communities near industrial sites. *Atmospheric Environment*. 2016 Aug 1;139:30-6.

53. Spengler JD, Samet JM, McCarthy JF. *Indoor Air Quality Handbook* (McGRAW-HILL: New York, Chicago, San Francisco, Athens, London, Madrid, Mexico City, Milan, New Delhi, Singapore, Sydney, Toronto, 2001).

Tables:

Table 1: Available questionnaire and measurements data for SPIROMICS Air participants

Type of data	Participants	Monitoring Sample N. of participants (N. of measurements) †‡	Sample with home monitoring and questionnaire data
Questionnaires			
Home Information Questionnaire (HIQ)	2054		287
Respiratory Disease and Smoke Exposure Questionnaire (RDSE)	2912		283
Questionnaires for validation			
Daily Activity Questionnaire		217	209
Home Inspection Form		216	209
Measurements			
Indoor $PM_{2.5}$ (µg/m^3)		201 (270)	194 (263)
Outdoor $PM_{2.5}$ (µg/m^3)		197 (271)	190 (264)
Indoor NO_2 (ppb)		216 (294)	209 (287)
Outdoor NO_2 (ppb)		216 (294)	209 (287)
Indoor NO_x (ppb)		216 (294)	209 (287)
Outdoor NO_x (ppb)		216 (294)	209 (287)
Indoor nicotine (µg/m^3)		205 (274)	198 (265)
Outdoor RH			209 (287)
Outdoor temperature			209 (287)

Note: †- not including participants with missing measurements; ‡-Participants from Ann Arbor, San Francisco, and Salt Lake City centers had up to two sets of home monitoring measurements.
Abbreviations: SPIROMICS - Subpopulations and Intermediate Outcome Measures in COPD Study; PM2.5 – particulate matters with diameter less than 2.5 μm; NO_2 - nitrogen dioxide; NO_x - oxides of nitrogen; RH- relative humidity.

Table 2: Prediction of indoor exposure for NO_2, NO_x, SHS and $PM_{2.5}$

		NO_2 (ppb)	**NO_x (ppb)**	**Nicotine (SHS) (μg/m³)**	**$PM_{2.5}$ (μg/m³)†**
Range of concentration (min, max of measured)	Indoor	(1.01, 115.91)	(1.27, 477.68)	(0.01, 19.92)	(1.19, 58.80)
	Outdoor	(0.99, 29.79)	(0.15, 143.29)	-	(1.66, 29.18)
Correlation of indoor to outdoor	Pearson's r (p-value)	*r*=0.34 (p<0.001)	*r*=0.41 (p<0.001)	-	*r*=0.01 *(p=0.92)*
Model A1‡: best linear	R^2/adj. R^2	0.40/ 0.35	0.38/ 0.33	0.40/ 0.30	0.52/0.45
	CV R2 (RMSE)	0.21 (8.88)	0.14 (38.36)	-0.12 (1.93)	0.30 (10.67)
Model A2‡: best logarithmic	R^2/adj. R^2	0.60/ 0.55	0.57/ 0.51	0.59/ 0.54	-
	CV R2 (RMSE)	0.46 (0.51)	0.39 (0.72)	0.45 (1.41)	-
Model A2‡: squared root with interactions	R^2/adj. R^2	-	-	-	0.58/ 0.52
	CV R2 (RMSE)	-	-	-	0.44 (1.09)
Model A3‡: removing outliers from Model A2	R^2/adj. R^2	0.63/ 0.57	0.58/ 0.53	0.61/ 0.56	0.60/ 0.54
	CV R2 (RMSE)	0.48 (0.47)	0.43 (0.69)	0.45 (1.37)	0.45 (1.06)
Model B4‡: correlation with residuals	R^2/adj. R^2	0.63/ 0.58	-	-	0.65/ 0.56
	CV R2 (RMSE)	0.49 (0.491)	-	-	0.42 (1.11)
Model B5‡: removing outliers from Model B4‡	R^2/adj. R^2	0.65/ 0.59	-	-	0.66/ 0.58
	CV R2 (RMSE)	0.51 (0.46)	-	-	0.45 (1.06)

Note:

† - For $PM_{2.5}$, all models (except the first one) are in square root function: Model A1 - is a model with linear function without interactions; Model A2 - square root function with interactions; Model A3 - removing outliers from Model A2; Model B4 - testing residuals; Model B5- removing outliers from Model B4;

‡ Letter "A" represents models that were built using the first "A" approach, while letter "B" represents models that were built using the second "B" approach (see subsection 2.3.2)

Table 3: Comparison of Model A3 prediction fit per pollutant

		Variables/groups	**Model A3: $PM_{2.5}$†** *Est.*	**Model A3: NO_2‡** *Est.*	**Model A3: NO_x‡** *Est.*	**Model A3: nicotine‡** *Est.*
		(Intercept)	2.52***	1.34**	2.28***	-5.49***
City§	*(compared to "Ann Arbor" city)*	Baltimore	0.06	0.03	0.59**	1.4***
		San Francisco	-0.01	0.21**	-0.22	0.67**
		Los Angeles	-0.48	-1.6E-03	-0.13	0.71
		New York	0.75**	0.03	-0.15	0.13
		Salt Lake City	-0.10	-0.17	0.60**	0.74
		Winston -Salem	-0.02	0.32*	0.18	2.53***
Meteorological measurements§		Outdoor temperature in C° (2-week av.)	-3.10E-04		-0.01***	
		Outdoor RH (2-week av.)		-0.01*		0.02
Proxy for infiltration estimation§		$PM_{2.5}$ µg/m^3 predicted infiltration	0.32			
		NO_2 outdoor (ppb)		0.49***		
		NO_x outdoor (ppb)			0.32***	
Socio-economic factors (based on Census		% adults with education less than High School	0.02**			
		Median value ($) for specified owner-occupied	6.7E-07			

data)§		housing units				
		Median family income		-3.14E-06 ***		
		% occupied housing units that are owner-occupied			-0.01***	-3.4E-03
		Median household income			3.9E-06 **	
Other pollutants measurements		$PM_{2.5}$ outdoor (µg/m^3)		0.03**	0.05**	
Smoking habits questions	*N. cigarettes per day were smoked in the past year by any smoker in the house? (compared to "none" group)*	Up to 20 cig.	1.07***			
		20 and more cig.	0.53**			
	Your approach to tobacco smoking in your home? (compared to "Never allow smoking in home" group)¶	Allow smoking only in certain rooms	-0.36			0.51
		Allow smoking In all rooms	1.28***			1.16**
	Do you smoke cigarettes (as of one month ago)? (Y/N)¶	Yes	0.99***			1.06***
	N. of cigarettes per day by each smoker. Is it more than 20 cigarettes? (Y/N)	Yes			0.45**	1.82***
	For how many years the allowing smoking at house approach?¶	(in years)				4.9E-03
	Traveled by car with	Yes				0.95**

	someone else who was smoking (during last week)? (Y/N)¶					
Building related questions	*Age of building*	(in years)	-2.64E-03	-8.5E-05	-9.7E-04	-0.01
	What floor do you live on? (compared to basement and ground floor)	2nd floor and higher	-1.19***			
	What type of building do you live in? (compared to "single family" type)	Rowhouse/townhouse/duplex/ triplex Apartment/condo Manufact./mobile				-0.33 -0.57* -0.63
	What the garage is used for? (compared to "no garage")	Parking 1 car Parking 2 cars Parking more than 2 cars Storage	-0.21 -0.17 1.32*** -0.32	0.28** 0.03 0.10 -0.07	0.23 0.15 0.39** -0.15	
Cleaner appliances questions	*Is an air cleaner/filter used? (Y/N)*	Yes	-0.57**			
	Type of air cleaner/filter (Y/N)	Electrostatic precipitator (Yes)		-0.41**		
	How often is the air cleaner/filter used?	More than half of the days and less Almost daily or daily		-0.17 0.08		
Windows use questions	*How often did you open the window (in summer)?*	A few days a month More than half of the days and less than daily Almost daily or daily		-0.19** 0.11 0.08	-0.26* -0.10 -0.11	-1.10*** -0.29 -0.42
	Double pane windows (Y/N)	Yes		0.13		
Heating	*What are the heating*	Fireplace wood (Yes)	3.68***			

sources questions	*sources used in your residence? (Y/N)*	temperature: Fireplace gas	-0.19***			
		Fireplace gas (Yes)	-0.64*			
		Forced Air (vents)(Yes)		-0.15**	-0.24**	
		Radiators (Yes)				1.21**
Cooking habits questions	*What type of oven is used? (compared to "electric" group)*	Gas oven		0.24**	0.37**	
	How often does someone cook (on stove) in residence? (compared to "not cooking" group)	Less than daily	-0.41	0.40**	0.41**	
		Almost daily or daily	0.26	0.40***	0.42**	
Pilots lights questions	*The presence of a pilot lights on: (Y/N)*	Oven (Yes)			0.15	
		Water Heater (Yes)			-0.39***	
		Clothes Dryer (Yes)		0.32***	0.52***	
Observations			195	223	230	192
R^2/ Adj. R^2			0.60/ 0.54	0.63/ 0.57	0.58/ 0.53	0.61/ 0.56
AIC			569.64	282.47	468.72	647.55
Cross Validation R^2 (RMSE)			0.45 (1.06)	0.48 (0.47)	0.43 (0.69)	0.45 (1.37)

Notes:

† – Model A3 is based on $PM_{2.5}$ squared root function

‡ - Model A3 is based on logarithmic function

§ - These variables are mandatory and were included in each model regardless of the choice of stepwise regression.

¶ – The questions that were obtained from Respiratory Diseases and Smoke Exposure questionnaire. Other questions related to house characteristics were obtained from Home Information Questionnaire.

*** - p-value <=0.001; **- p-value <0.05; * - p-value <0.1

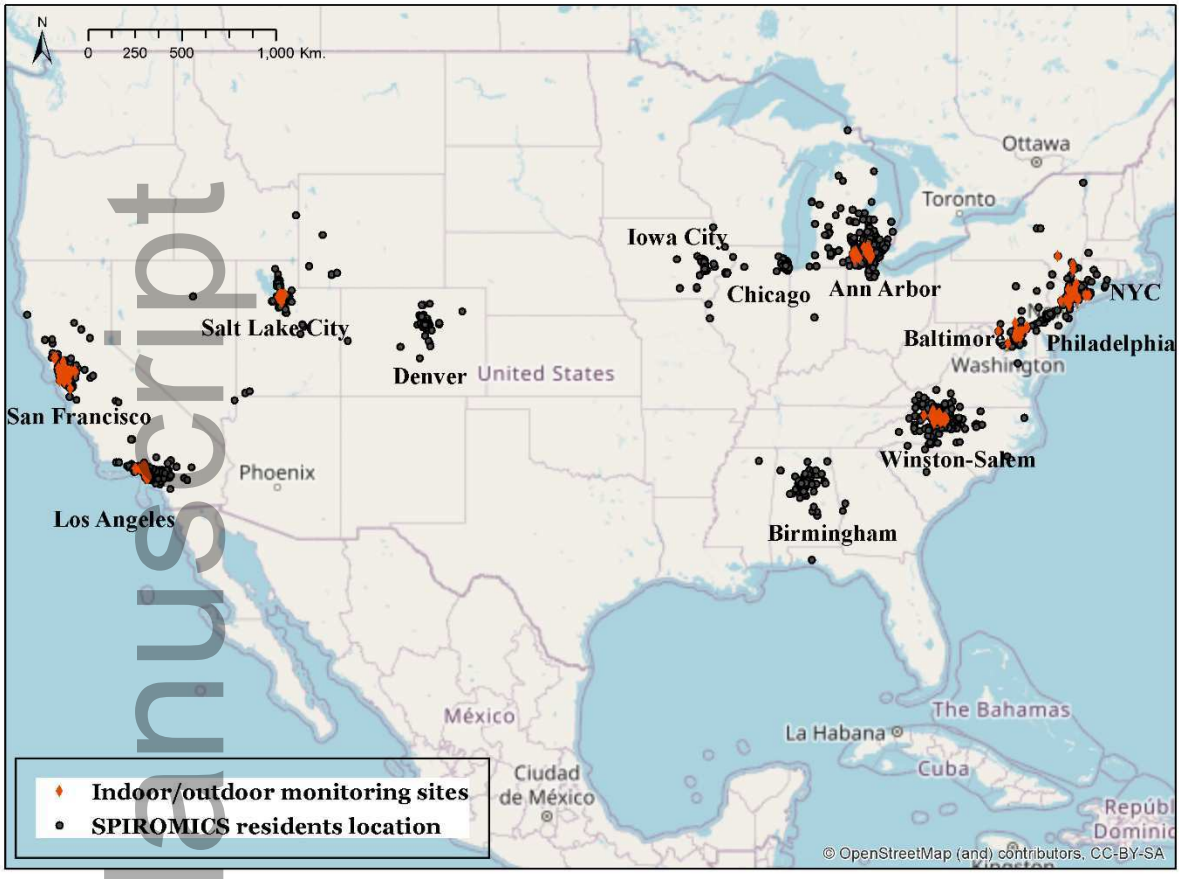
Figure legends:

Figure 1: Map of regions covered by SPIROMICS and SPIROMICS Air with the location of participants, indoor/outdoor monitoring sites

Figure 2: The structure of the SPIROMICS AIR total exposure prediction modeling

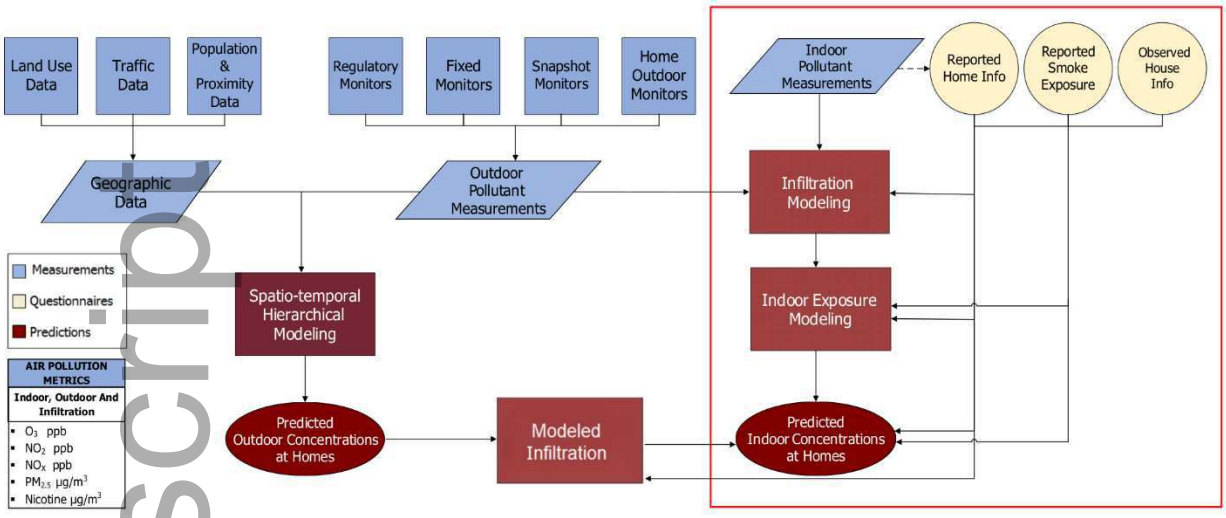
Figure 3a: Indoor and outdoor concentrations grouped by city

Figure 3b: Indoor and outdoor temperature °C and relative humidity grouped by city and season



Note: All participants' locations have been jittered; Black dots represent the locations that were reported by participants at the time of enrollment. Participants that were not recruited from one of the seven SPIROMICS Air sites were excluded from the analysis.

Figure 1: Map of regions covered by SPIROMICS and SPIROMICS Air with the location of participants, indoor/outdoor monitoring sites



Note: The scope of the current paper is marked by red rectangle. Indoor and outdoor measurements for indoor exposure prediction modeling were collected inside and outside participants' homes.

Figure 2: The structure of the SPIROMICS AIR total exposure prediction modeling

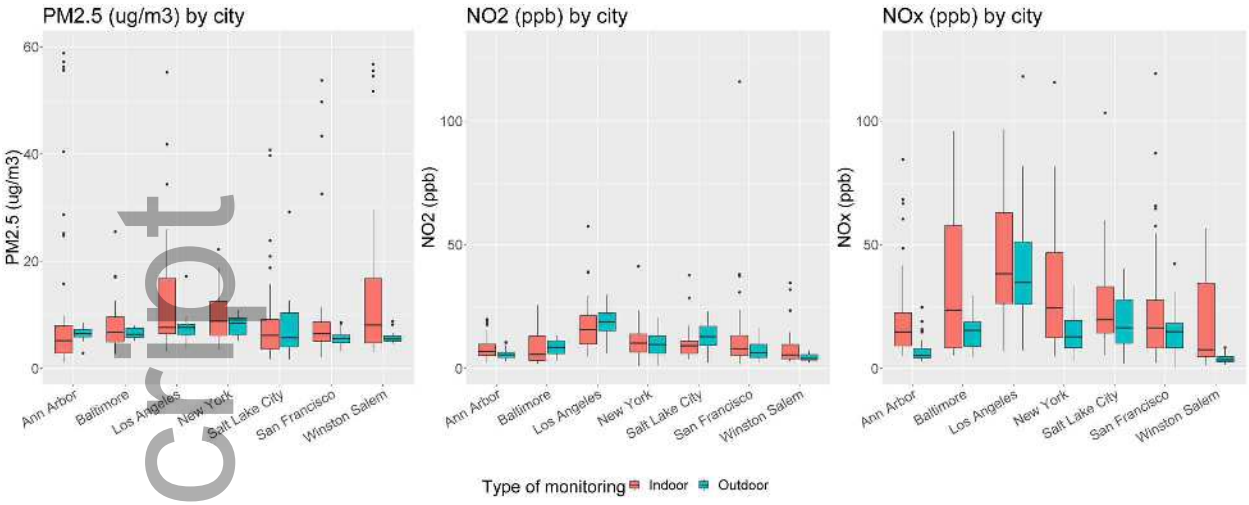


Figure 3a: Indoor and outdoor concentrations grouped by city

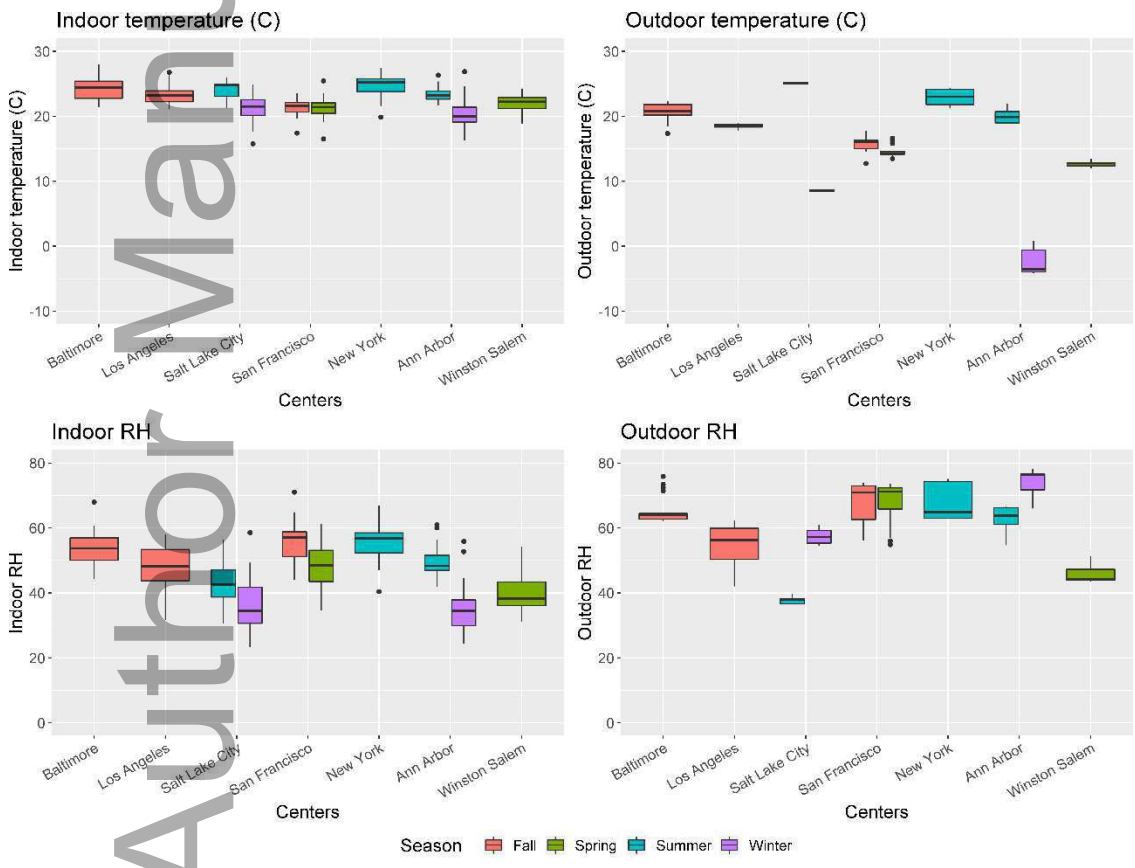
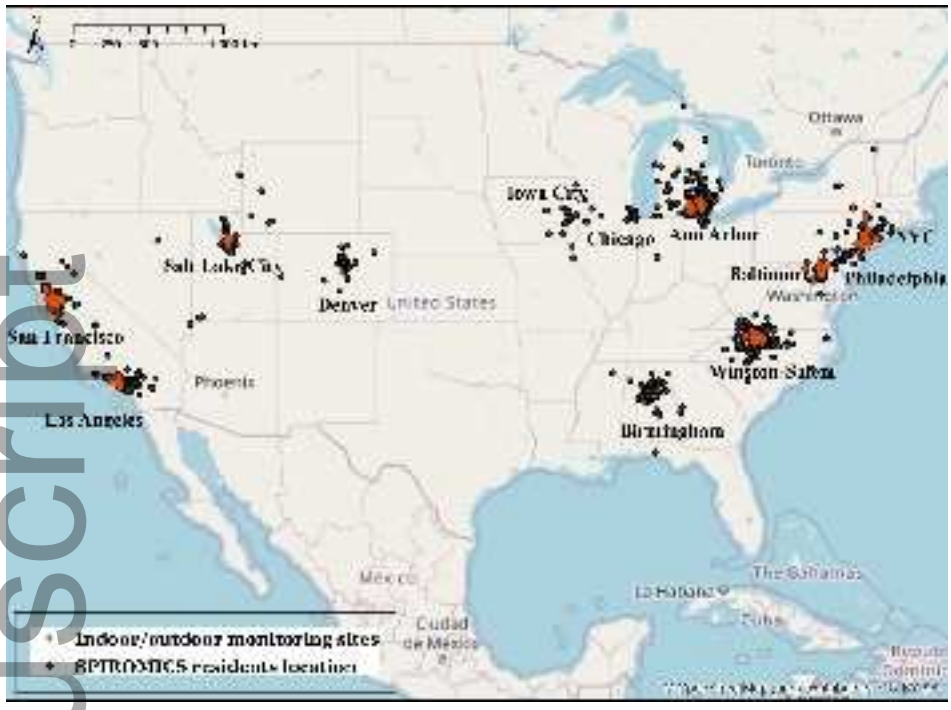
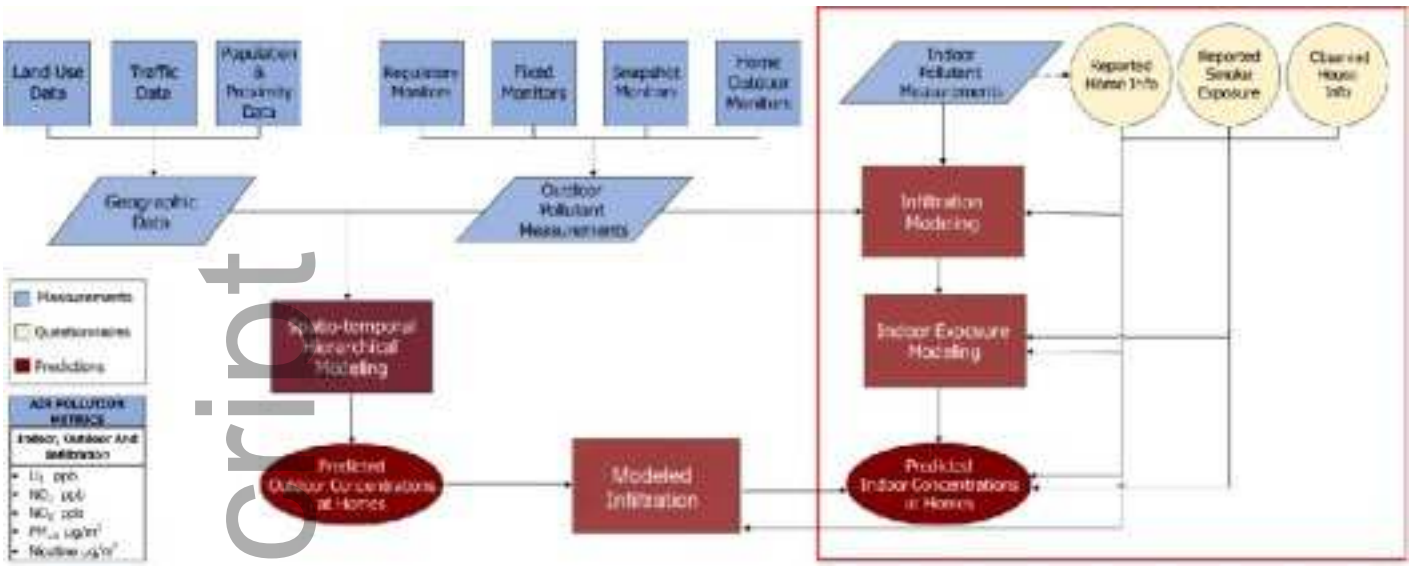


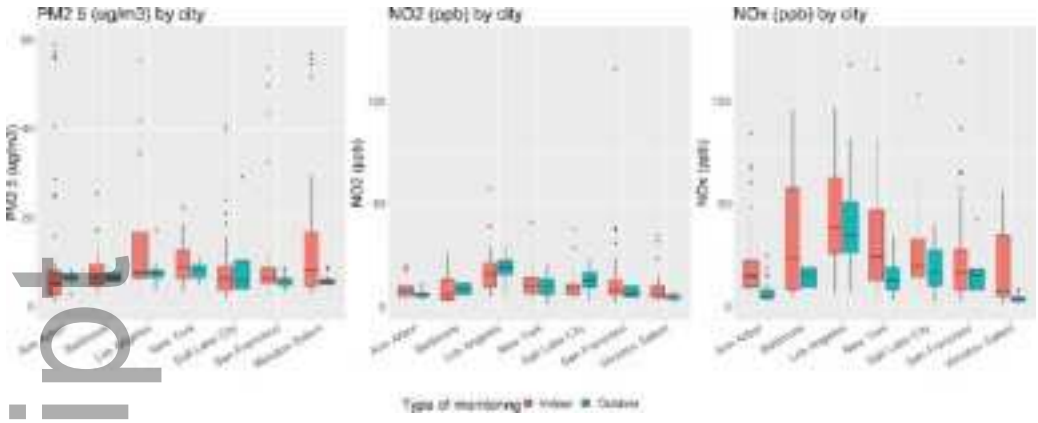
Figure 3b: Indoor and outdoor temperature °C and relative humidity grouped by city and season



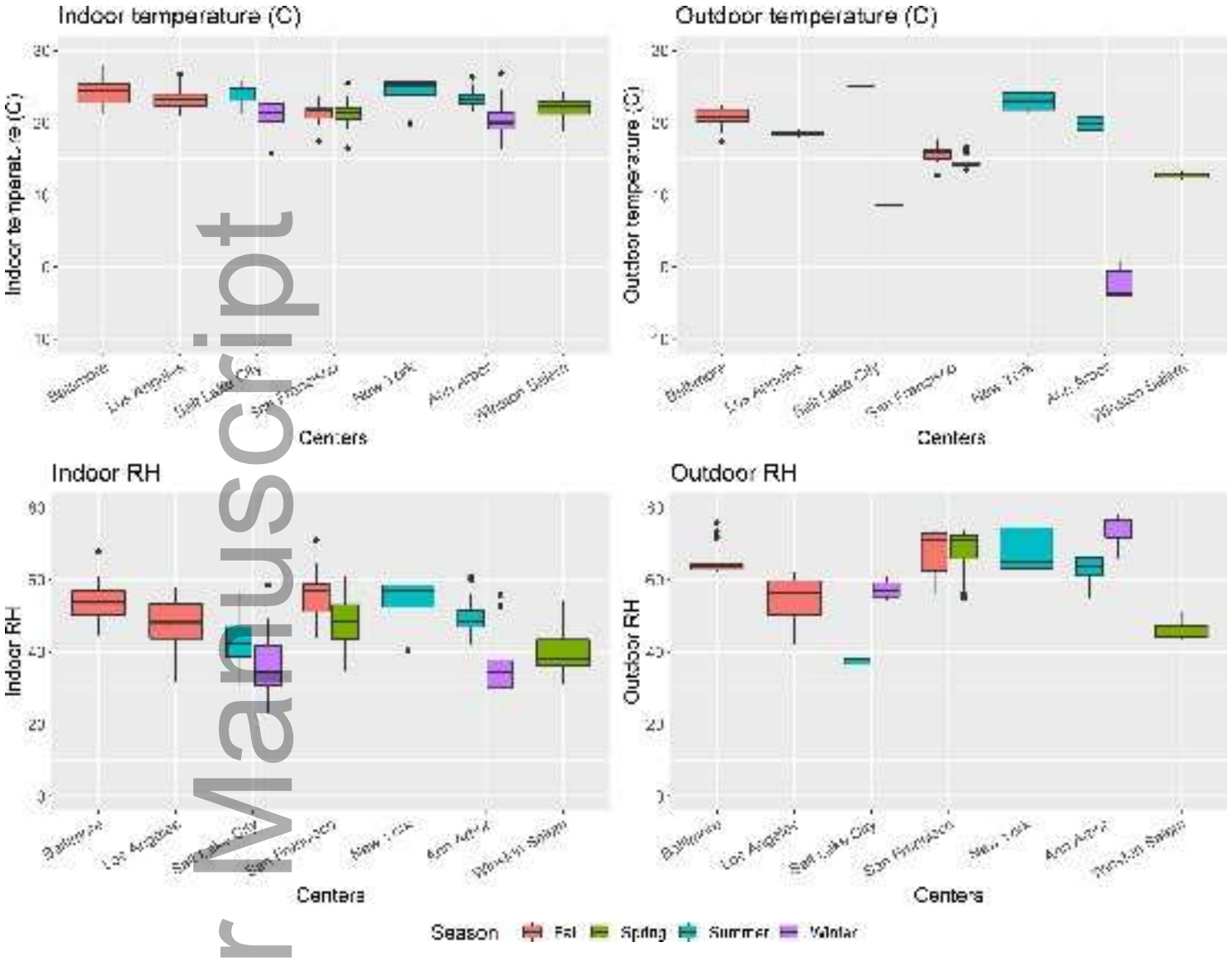
ina_12760_f1.png



ina_12760_f2.png



ina_12760_f3a.png



ina_12760_f3b.png