THESIS

EVALUATING STATISTICAL METHODS TO PREDICT INDOOR BLACK CARBON IN AN URBAN BIRTH COHORT

Submitted by

Sherry D. WeMott

Department of Environmental and Radiological Health Sciences

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer, 2022

Master's Committee:

Advisor: Sheryl Magzamen

Matthew Koslovsky
David Rojas-Rueda

ABSTRACT

EVALUATING STATISTICAL METHODS TO PREDICT INDOOR BLACK CARBON IN AN URBAN BIRTH COHORT

Though individuals in the United States spend a majority of their time indoors, epidemiologic studies often use ambient air pollution data for exposure assessment. We used several modeling approaches to predict indoor black carbon (BC) from outdoor BC and housing characteristics to support future efforts to estimate personal air pollution exposure given time spent indoors. Households from the Healthy Start cohort in Denver, CO were recruited to host two paired indoor/outdoor low-cost air samplers for one-week sampling periods during spring 2018, summer 2018, and winter 2019. Participants also completed questionnaires about housing characteristics like building type, flooring, and use of heating and cooling systems. Sampled filters were analyzed for BC using transmissometry. Ridge, Lasso and multiple regression techniques were used to build the best predictive model of indoor BC given the available set of covariates. Leave-one-out cross-validation was used to assess the predictive accuracy of each model. A total of 27 households participated in the study, and BC data were available for 39 filters. We had limited comparable data on seasonality as winter data were excluded from the analysis due to high variability and low confidence in outdoor measurements. Shortened runtimes and other performance issues suggest insufficient weatherproofing of our monitors for low temperatures. Of the three modeling approaches, Ridge LSE showed the best predictive performance (MPSE 0.50). The final inference model included the following covariates: outdoor $PM_{2.5}$, outdoor BC, hard floors, and pets in the home (adj. R^2=0.27). These factors accounted for

approximately 27% of the variability in indoor BC concentrations measured in Denver, CO homes. In the absence of personal monitoring, household characteristics and time-activity patterns may be used to calibrate ambient air pollution concentrations to the indoor environment for improved estimation of personal exposure.

ACKNOWLEDGEMENTS

I would like to extend my deepest gratitude to my advisor, Dr. Sheryl Magzamen, for providing the opportunity to work as a research associate in Environmental Health (EH), and eventually pursuing by Master's degree under her mentorship. Her enthusiasm, breadth of knowledge, writing skills and kindness cannot be underestimated. To my other committee members Dr. Matt Koslovsky and Dr. David Rojas-Rueda, the completion of analysis and thesis would not be possible without you. Dr. Koslovsky presented me with ideas to use for statistical analysis and guided me through the often-complicated process with patience and humor. Dr. Rojas-Rueda was an encouraging and positive instructor in two of my courses, as well as providing resources and ideas for writing up my thesis chapters. I cannot begin to express my thanks to my talented and brilliant lab team including Grace Kuiper, Dr. Sheena Martenies and Elizabeth Brooks. Grace has been my partner on this project since day one, and it never would have gotten off the ground without her impressive organizational and management skills. Ms. Kuiper and Dr. Martenies were both instrumental with cleaning and organizing an immense amount of complex data that was trimmed down and used for my analysis. Dr. Martenies also walked me through analysis in R and shared her script early in the project when I was still learning. Ms. Brooks was there at the beginning of the project to get me up to speed on logistical details. My sincere thanks to all the EH faculty that helped me build my epidemiology toolkit including Dr. Brooke Anderson, Dr. Jennifer Peel, Dr. Molly Gutilla and Dr. Maggie Clark. I'm deeply indebted to Dr. Bonnie Young for her writing and field work mentorship, and most importantly her friendship. Many thanks to our field team and collaborators on this project from University of Colorado, Anschutz including Stephen Brindley, Kacy Lorber, Cody Dawson and

Kevin Andresen. They worked tirelessly to deploy and our monitoring equipment, meet with participants and keep track of samples. We could not have completed the field campaigns without their on the ground work. I also had the great pleasure of going through the program with fellow graduate student, Matthew Hughes. His infectious enthusiasm and love of cheese helped keep my spirits up. Special thanks to Dan Miller-Lionberg and the rest of the staff at Access Sensor Technologies which manufactured and help us troubleshoot the personal sensors we used to collect air quality data across the Denver Metro area. I very much appreciate the work of Christian L'Orange and other team members of the Volckens Lab for their timely weighing, scanning and data processing of the multitude of filter samples we sent their way. And finally, my deepest appreciation to my all my friends and family whom provided me unwavering support, guidance, and motivation when I most needed it.

TABLE OF CONTENTS

CHAPTER ONE: LITERATURE REVIEW

This chapter provides an in-depth overview of the literature associated with this project. The largest sections will cover air pollution and human health, specifically pregnancy outcomes. There are also sections devoted to discussing the differences in indoor and outdoor pollution, as well as a detailed background on measurement and health effects of black carbon (BC).

Historical air pollution events

Air pollution is not just a 21st century problem; air quality issues have occurred throughout history in areas where people burn biomass to heat their homes or work in craft or industry requiring the use of such fuels. The Romans had a phrase for air pollution, *gravorioris caeli,* or heavy heaven.[1] Although it was often considered unpleasant, air pollution was also seen as a sign of wealth and progress in cities. This seeming conflict between health concerns and economic progress continues today. Several particularly deadly air pollution events led to the creation of acts and policies that would lead to improved air quality in many cities across the world.

Donora Smog 1948

This event in Donora, Pennsylvania, is considered the worst air pollution disaster in United States history. [22] The cities of Donora and the nearby village of Webster were covered in a yellow fog from October 26 through October 31 of 1948. The event led to 20 deaths and multiple hospitalizations due to respiratory or cardiac conditions. Following the event, investigators from local, state and national agencies such as the United States Public Health Service (USPHS) looked into the cause of the fog and resulting mortalities. Thus began the first large scale epidemiological study of an environmental health disaster in the U.S. The initial

conclusion by the USPHS was that the smog was caused by three primary factors.[3] First was pollution emitted by American Steel and Wire Plant and the Donora Zinc works. The Zinc Works emitted hydrogen fluoride, carbon monoxide, nitrogen dioxide, several sulfur compounds as well as heavy metals within a mix of fine particulate matter.[3] Also identified as important factors was an unusual weather system resulting in a temperature inversion, which worked to trap the fog in the valley. Donora's unique geography, surrounded by hills and cliffs, further contributed to the entrapment of the fog. Steel and zinc industries had a powerful presence in town and accounted for a large proportion of local jobs. At the conclusion of the investigation, American Steel & Wire settled without accepting blame for incident.[2] However, this event would have effects on public health and air pollution epidemiology for many years to come.

Great Smog of London 1952

Fog was a common occurrence in London, but not the toxic, smoke-laden fog that shrouded the British capitol for five days in December 1952. Although smoke and fog ("smog") events were becoming more frequent with the spread of the Industrial Revolution, none had reached the severity that was seen in 1952. In some parts of the city, people were unable to see their feet while walking.[4] Once again, weather phenomena in combination with pollution emitted from factories and coal burning home furnaces combined to form the deadly smog. As was the case with the Donora Smog, a temperature inversion had reduced air flow and trapped pollutants close the ground. The cold winter temperatures meant that people were burning large quantities of coal in their homes. It is estimated that 4,000 people died as a result of the fog, but some historians think it was closer to 12,000.[5] Although it took some time, the British government eventually passed the Clean Air Act in 1956, as a direct response to the 1952 smog event. [4]

New York City Smog 1966

Over Thanksgiving weekend in 1966, a stagnant air mass settled over the New York City area.[6] The resulting heavy, toxic smog lasted for three days and is estimated to have resulted immediately in 60 deaths, roughly 10.5 times the expected amount under ordinary circumstances for that time period.[7] Most of the deceased were those with chronic cardiorespiratory diseases or of advanced age. Autopsies showed inflammatory signatures of chemical irritants in the respiratory tract.

Air Pollution Legislation

The Clean Air Act (CAA) of 1963 was the first federal legislation in the US to describe air pollution control policy; the 1955 Air Pollution Control Act provided funds only for federal air pollution research. [8] The 1963 CAA Established a federal program within the U.S. Public Health Service and funded research in the areas of monitoring and control. In 1967, the Air Quality Act (ACA) expanded federal government authority on air pollution control and allowed the enforcement and monitoring of activities that resulted in interstate air pollution transport.

The federal government took an even larger role in air pollution control and research beginning with the CAA of 1970. In the same year, legislation was adopted that formed the U.S. Environmental Protection Agency (EPA). It's main function at the time was to implement federal and state regulations set forth by the 1970 CAA. The legislation set limits for emissions from both stationary (industrial) and mobile sources.

Amendments to the CAA in 1977 and 1990 further strengthened federal programs for air quality control, monitoring and research. The 1990 amendments, including regulatory programs for acid rain as well as an expanded program for controlling toxic air pollutants, are the basis for current federal legal authority regarding air pollution.

Current Federal Standards and Criteria Pollutants

The 1990 Clean Air Act requires the EPA to set National Ambient Air Quality Standards (NAAQS) for six principal pollutants ("criteria" air pollutants) that can harm public health and the environment. Primary standards aim to provide direct public health protection, including the health of sensitive groups such as children, elderly and people with asthma. Secondary standards protect the public welfare, inclusive of protection against damage to buildings, crops, animals and vegetation as well as decreased visibility.[9] Pollutants are measured in parts per million (ppm) by volume (ppm), parts per billion (ppb) by volume, and micrograms per cubic meter of air ($\mu g/m^3$). The current primary pollutants include Carbon Monoxide (CO), Lead (Pb), Nitrogen Dioxide (NO_2), Ozone (O_3), particulate matter ($PM_{2.5,}$ PM_{10}) and Sulfur Dioxide (SO_2). All but CO have standards as secondary pollutants as well. Below is a brief description of each pollutant as well as discuss its criteria level. Information on standards is derived from the US EPA NAAQS Table.

Carbon Monoxide (CO): a colorless, odorless, gas CO is released from combustion. Ambient sources of CO include cars, trucks and any other fossil fuel burning equipment of machinery. CO can also come from indoor sources such as unvented kerosene or gas space heaters, leaking chimneys or furnaces, gas stoves and generators. The standard for CO is maximum 36 ppm for a 1-hour period and 9 ppm for 8 hours.

Lead (Pb): There are various ways lead enters the air, including ore and metals processing, waste incinerators, utilities, and lead-acid battery manufacturing. According to the EPA, highest air concentrations are found near lead smelters. Regulation requiring

the removal of lead from vehicle fuel reduced the amount of lead in the air from 98 percent between 1980 and 2014. The maximum acceptable limit for lead in the air is 0.15 $\mu g/m^3$ over a rolling 3-month average.

Nitrogen Dioxide (NO_2): one of the several nitrogen oxides (NO_x), NO_2 typically arises via the oxidation of nitric oxide by oxygen in the air and is formed in most combustion processes. The most common anthropogenic sources are internal combustion engines and burning of fossil fuels. Maximum levels are 100 ppb hourly and 53 ppb annually.

Ozone (O_3): Also known as trioxygen, ozone is a pale blue gas and a powerful oxidant. It occurs at both ground level and in the earth's upper atmosphere. "Good" ozone occurs naturally in the stratosphere and shields the planet from the sun's harmful ultraviolet (UV) rays. "Bad" or tropospheric ozone is found at ground level and is a harmful air pollutant, often the main constituent of smog. Ground level ozone is created by chemical reactions between sunlight and air containing volatile organic compounds (VOCs) and NO_x. Urban areas with high density of motor vehicle use, which emit NOx and (VOCs), are often affected by the highest ozone levels. Ozone can also be transported by wind into rural areas. The maximum acceptable ozone concentration is 0.070 ppm over an 8-hour period. Nonattainment occurs when the annual fourth-highest daily maximum 8-hour concentration averaged over 3 years exceeds this value.

$PM_{2.5}$: Particulate pollution is a mixture of solid particles and liquid droplets found in the air.

This category of particulate matter (PM or particle pollution), includes the fine, inhalable particles with diameters generally 2.5 micrometers (μm) or less and can only be seen individually under a microscope. $PM_{2.5}$ is typically a product of combustion, including vehicle exhaust, industry and biomass and is often the biggest concern for epidemiologists as these particles can be inhaled deeply into the lung. The composition of $PM_{2.5}$ is thought mainly to be sulfate, nitrate, ammonium, elemental carbon, organic carbon, silicon and sodium ions. [10] The maximum limit for $PM_{2.5}$ as a primary pollutant is 12.0 $\mu g/m^3$ averaged over a 1-year period.

- **BC (black carbon):** also known as soot, BC particles are a constituent of $PM_{2.5}$ formed from incomplete combustion of biomass and fossil fuels. Although BC concentrations are not currently regulated by the EPA, BC exposure has been associated with cancer, respiratory and cardiovascular disease and is the second largest contributor to global temperature increases behind CO_2.

PM_{10}: Although still microscopic, PM_{10} particles are slightly larger with a diameter generally 10 μm or smaller. In addition to the combustion particles commonly found in $PM_{2.5}$, PM_{10} particles include dust from construction sites, landfills and agriculture, wildfires and waste burning, industrial sources, wind-blown dust from open lands, pollen and fragments of bacteria.[11]The primary and secondary standard for PM_{10} is 150 $\mu g/m^3$ – this value is not to be exceeded more than once per year on average over 3 years.

Sulfur Dioxide (SO_2): National air quality standards are designed to protect against exposure to all sulfur oxides (SO_x), among which SO_2 is the most concerning. The main source of SO_x pollutants is the combustion of fossil fuels by power plants or industrial facilities. The maximum limit for the primary standard of SO_2 is 75 ppb over one hour.

Evidence of Adverse Effects of Air Pollution on Birth Outcomes

This section explores recent (past 15 years) epidemiological literature looking at the association of air pollution with adverse health effects in humans, and specifically, birth outcomes.

Reviews and Meta-Analysis

In 2005, Sram *et al*. published a review of ambient air pollution and pregnancy outcomes.[12] A majority of studies reviewed in the paper were population based, with exposure to air pollution measured by fixed-site areawide or citywide monitors. Pollutants measured include total suspended particles (TSP), sulfur dioxide (SO_2), carbon monoxide (CO), $PM_{2.5}$, PM_{10}, nitrogen dioxide (NO_2), ozone (O_3) and oxides of nitrogen (NO_x). Some of the specific outcomes reviewed include child mortality (including pre- and post-natal outcomes), birthweight, premature births and intrauterine growth restriction (IUGR). Overall, the authors find the evidence shows an association between air pollution exposure and adverse birth outcomes. However, there is different strength of association between outcomes. The evidence linking particulate air pollution with decreased birth weight and respiratory deaths in the postneonatal period was sufficient to infer causality. The association between air pollution and preterm birth and IUGR was weaker, but still enough to justify further study.

For example, in a time-series study conducted by Loomis *et al*. in Mexico City (1993-1995) that investigated the association between PM_{10} and infant mortality, the authors found a 6.9% excess in infant mortality (95% CI, 2.5 – 11.3%) for each 10μg/m^3 increase in PM_{10}.[13] A study that took place in Sao Paulo, Brazil examined the association between intrauterine mortality (stillbirth) and PM_{10} found no effect with increasing exposure of PM_{10}.[14] A group of studies that investigated low birth weight (LBW; < 2,500 g) show more similar odds ratios, but mostly small effects sizes. Specifically, Wang *et al*. found an AOR of 1.21 (95% CI, 1.06 – 1.16)

when looking at total suspended particles (TSP) as the exposure,[15] while Ha *et al.* found a smaller AOR of 1.04 (95% CI 1.00 – 1.08) with increasing exposure to TSP in the first trimester. [16] The variance in these effect sizes highlights the challenge of proper exposure assessment in environmental epidemiology.

A 2013 multi-center meta-analysis looking at data from the International Collaboration on Air Pollution and Pregnancy Outcomes (ICAPPO) was published in 2013 by Dadvand *et al*. [17] The aim of this analysis was to evaluate the association of birth weight outcomes with maternal air pollution exposure. The authors combined effect estimates from each of the ICAPPO centers and investigated inter-center heterogeneity. A common analysis protocol was used at all of the ICAPPO centers, thus eliminating any effect differences resulting from divergent analysis approaches. The analysis focused primarily on the association between PM_{10} exposure during the duration of the pregnancy and term LBW outcome (birth weight < 2500 g at 37-42 completed weeks of gestation). Included in the analysis were 14 ICAPPO centers including 6 in North America, 5 in Europe and 1 each in South America, Asia and Oceania; more than 3 million births were analyzed in total. Random-effects models of combined ORs showed a positive association between term LBW and a $10\mu g/m^3$ increase in average maternal exposure to PM_{10} during entire pregnancy prior to adjustment (OR = 1.04; 95% CI: 1.01, 1.06), the OR decreased slightly to 1.03 after adjusted for SES (95% CI: 1.01, 1.05). Data on $PM_{2.5}$ was analyzed from seven centers and results indicated positive association with an OR of 1.10 after being adjusted for maternal SES (95% CI: 1.03, 1.18). Therefore, potential confounders for univariate results were not evaluated.

A comprehensive review on ambient air pollution (AAP) and pregnancy outcome studies was conducted by Klepac *et al*. at the National Institute of Public Health in Slovenia in 2019.[18]

For the first phase of the study, the authors extracted components of interest from each study including defined pregnancy outcomes, AAP exposure assessment methods, study design, sample size and statistical analysis methods. For the second phase, effect estimates of the most commonly studied ambient air pollutants on pregnancy outcomes were evaluated. Furthermore, the authors conducted a meta-analysis on pre-term birth (PTB) outcome, and used the Newcastle-Ottawa Scale (NOS) to assess study quality. The NOS is a tool used to assess the quality of nonrandomized studies such as case-control and cohort, and can be incorporated into the interpretation of meta-analytic results. The developers of the scale use a 'star system' in which a study is judged on three areas: selection of study groups, the comparability of groups, and the ascertainment of either the exposure or the outcome of interest for case-control or cohort studies respectively.[19] A total of 96 articles, a majority published in 2010 or later, met the quality inclusion criteria and were included in the meta-analysis. Of these studies, a large portion (n = 45) were conducted using a retrospective cohort design and assessed the association between routine ambient air monitoring and birth records data. The most commonly studied pollutants were PM_{10}, $PM_{2.5}$, NO_2, CO and O_3, while the most commonly investigated pregnancy outcomes were PTB, LBW, birth weight (BW) as a continuous variable (in grams), congenital anomalies and small for gestational age/intrauterine growth restriction (SGA/IUGR, BW < 10^{th} percentile for gestational age).

PTB (defined as < 37 weeks gestation) was the most widely studied outcome found in this review with a total of 28 studies included in the review and assessed for quality. Meta-analysis was used to calculate pooled effects estimates across the most common pollutants. The pooled effects estimates (OR) for exposure per $10\mu g/m^3$ increase in particulate matter over the entire pregnancy was significantly associated with risk of PTB: 1.09 (95% CI, 1.03 – 1.16) for

PM_{10} and 1.24 (1.08-1.41) for $PM_{2.5}$. Ozone (O_3) also showed association with PTB and the pooled effects estimates per 10 ppb increase were 1.03 (1.01-1.05) for 1st trimester, 1.12 (1.05-1.19) for second trimester and 1.03 (1.01-1.04) for the entire pregnancy. Evidence of an association between CO and PTB was also present as the pooled effect estimates per 1 ppb increase in CO concentration were 1.36 (1.15-1.62). Only one study in the review used personal air quality monitoring; this leaves open the possibility of exposure misclassification if changes of maternal residence address and time-activity patterns during pregnancy are not considered. Most studies also lacked information on indoor air pollution exposure and a recent study indicates that between 30-75% of indoor $PM_{2.5}$ may originate from the outdoor environment.[20]

Traffic Related Air Pollution

There is a growing body of evidence that suggests PM from combustion sources may be more harmful compared to PM from other sources.[21,22] Black carbon (BC) particles are formed from incomplete combustion of biomass and fossil fuels are often used to estimate exposure to traffic related air pollution (TRAP). Once inhaled, small BC particles can enter the circulation system and illicit inflammatory responses throughout the body.[23] Furthermore, there is evidence linking ambient air pollution exposure *in utero* to adverse health outcomes in infants in children.

A 2008 study by Brauer *et al*. used spatiotemporal exposure metrics to assess the impacts of air pollution exposure on pregnancy outcomes, specifically, low full-term birth weight (LBW) and preterm birth small for gestational age (SGA) birth weight. [24] Their findings showed that residence within 50 m of a highway was associated with a 11% (94% CI, 1.01 - 1.23) increase in LBW and a 26% increase in SGA (95% CI, 1.07 – 1.49). In addition, associations were observed with $PM_{2.5}$ and births < 37 weeks gestation; exposure to all pollutants except O_3 was associated with SGA. The study area of Vancouver, British Columbia is generally considered to have

relatively low ambient air pollution levels; nonetheless, intraurban findings did not include the null value. The null value, in an epidemiological context, is the value that corresponds to no effect or shows no association between exposure and health outcome. The authors note potential exposure misclassification due to the fact that subject mobility was not captured.

A prospective birth cohort study of healthy newborns conducted in Switzerland from 1999 – 2007 found that minute ventilation (tidal volume multiplied with respiratory rate, inverse measure of lung development and function) was higher in infants with higher pre-natal PM_{10} exposure (24.9 mL * min-1 per µg m-3 PM_{10}). Increased respiratory rate can be a sign of respiratory distress syndrome (RDS) if breaths reach more than 100 a minute. [25] Increased respiratory rate can be due to the lower work of breathing required with small rapid breaths. In addition, authors report that exhaled nitric oxide (eNO) was increased in infants with higher prental NO_2 exposure. Pollution measurements were captured at the monitoring station of Payerne (part of the Swiss National Air Pollution Monitoring Network), and regional data were used to calculate mean exposure of subjects to each pollutant during their pregnancy period. In addition, a stronger association was also found with minute ventilation and mothers that lived close to major roads. Although there were many strengths of this study, including standardized lung function measurement, exposure misclassification is still possible given that neither individual or indoor air pollution measurements were taken.

In 2012, Sapkota *et al*. published a meta-analysis of 20 peer-reviewed articles with quantitative estimates of exposure to particulate matter and adverse birth outcomes.[26] Random-effects meta-analysis results suggested a 9% increase in risk of LBW associated with a $10\mu g/m^3$ increase in $PM_{2.5}$ (combined OR, 1.09; 95% CI, 0.90 – 1.32), although it is worth noting the CI does include the null value, or number corresponding to no effect. Increase in risk of pre-term

birth (PTB) increased 15% for each 10μg/m^3 increase in $PM_{2.5}$ (combined OR, 1.15; CI, 1.14 – 1.16). The authors note that there is significant heterogeneity between study results; in particular with PM_{10}. This could be a contributing factor to the results of the analysis showing a smaller magnitude of risk associated with PM_{10} (2% per 10μg/m^3). Nearly all studies included in the meta-analysis obtained outcome data from public records such as birth certificates.

Wildfires

As climate change impacts accelerate, exposure to wildfire smoke will increase in the many parts of the world.[27] Holstius *et al*. investigated birth weight outcomes following pregnancy during the 2003 Southern California wildfires[28] and found a reduced average birthweight among infants exposed *in utero*. In this study, outcome data was collected from birth records in the South Coast Air Basin (SoCAB) during 2001-2005. Smoke exposure, which is typically measured using $PM_{2.5}$ concentrations, was primarily determined by reports of smoke from the California Department of Forestry and Fire Protection and inspection of Moderate Resolution Imaging Spectroradiometer (MODIS) satellite imagery. From this data, the window of potential wildfire smoke exposure was set as October 21 – November 10, 2003. Birth weights were compared between pregnancies that took place entirely before or after the wildfire event. Mean birthweight of babies born during the wildfire was 3.3 g lower (95% CI: -7.2, 0.6) when exposure occurred during first trimester, 9.7 g lower (95% CI: -14.5, -4.8) during second trimester and 7.0 g lower (95% CI: -11.8, -2.2) during third trimester. Potential exposure misclassification could be possible due to lack of information on spatial variability and a relatively small number of exposed mothers could have been affected more significantly than the estimates would demonstrate.

Asthma

A growing body of evidence suggests *in utero* exposure to air pollution can also increase the likelihood of development of childhood asthma. The objective of a study conducted in 2015 by Hsu *et al*. was to identify sensitive windows for effects of prenatal exposure to $PM_{2.5}$ on children's asthma development in an urban pregnancy cohort.[29] The final analysis was conducted on data from 736 full-term (≥ 37 weeks) children and their mothers. Mother child dyads were selected from a pool of patients receiving care at Brigham and Women's Hospital, Boston Medical Center and the affiliated community. A novel spatiotemporal MODIS aerosol optical depth (AOD) was used to estimate mothers' exposure to $PM_{2.5}$ over the course of the pregnancy. This model was layered with traditional land use regression (LUR) predictors to provide residence specific estimates of daily $PM_{2.5}$ exposure. Asthma diagnosis was maternal-reported and clinician diagnosed and was ascertained from birth to 6 years. There were 110 asthma cases among the final group of 736 children; a significant sensitive exposure window was observed between 12 and 26 weeks for boys but not girls when stratified by sex. In addition, a multivariable logistic regression model including interaction between $PM_{2.5}$ and sex showed a significant interaction ($p = 0.01$). The novel use of LUR/AOD to model daily $PM_{2.5}$ was one of the first studies to objectively estimate exposure windows during pregnancy.

Telomere Length

A new and interesting direction in air pollution research evaluates the telomere as a biomarker of biological aging. Oxidative stress and chronic inflammation are thought to be potential mechanisms for the adverse pregnancy outcomes associated with maternal exposure to air pollution.[30] One mechanism through which oxidation damages the human body on a cellular level is through oxidation of DNA, which may cause damage to the telomeres. The telomere

region in human DNA protects the end of the chromosomes from degradation and from sticking to other chromosomes (end-to-end fusion). Every time a cell carries out DNA replication, the chromosomes are shortened (by about 20 base pairs), but due to the telomere protected end, that is the only part of the chromosome that is lost. At a certain point, telomeres become too short and the chromosome reaches the "critical length"; it can no longer be replicated and the cell dies by apoptosis or programmed cell death. A 2020 pregnancy cohort study conducted in Copenhagen, Denmark as part of the Maternal Stress and Placental Function project,[31] aimed to assess the association between prenatal exposure to air pollution and telomere length (TL) in the maternal blood cells (leukocytes), placenta and umbilical cord blood cells. The participants were pregnant women who gave birth at the Copenhagen University Hospital, Department of Obstetrics; a total of 296 mother-child dyads participated in the study. Residential air pollution exposure was modeled at street resolution using the DEHM-UBM-AirGIS system (Danish Eulerian Hemispheric Model, Urban Background Model, traffic and street geometry model) and occupational address exposure modeled at 1 km^2 grid resolution using DEHM-UBM models only. Pollution exposure measurement started 8 weeks before the estimated start of pregnancy and was modeled in 7-day periods for a total of one year. Along with pollutant measurements, ambient temperatures were also modeled and were used to adjust in the distributed lag models. Indoor air pollution data was collected via questionnaire, along with other information on potential confounders including education (via an occupational questionnaire), pregnancy complications and maternal strain (or difficulty) of birth.

Associations were reported using percent interquartile range (IQR) increase in pollutant and mean relative telomere length across gestational weeks, and results were mixed. There were significant and *positive* associations between TL in umbilical cord cells and prenatal exposure to

BC (percent change (95% CI) = 22% (2, 46)), organic carbon (OC) 43% (12, 84), NO_2 20% (3, 39), NO_x 19% (3, 37), CO 70% (24, 132) and O_3 60% (24, 107) during the second trimester. TL in umbilical cord blood was significantly and *inversely* associated with prenatal exposure to $PM_{2.5}$ -23% (-35, -9), BC -19% (-29,6), OC -22% (-35, -7), SO_2 -33% (-47, -16), NH_4^+ -56% (-70,-34), CO -29% (-48, -5) and NO_2 -20% (-31, -6) in the third trimester. The barrier of the third trimester placenta is relatively thin compared to other trimesters and can allow more compounds such as PM, pro-inflammatory cytokines and pro-oxidant species to cross. The authors speculate that the positive association in the second trimester may be an indicator of the maternal inflammatory response to air pollution exposure and the inverse association seen in the third trimester a combined effect of inflammation and oxidative stress response caused by direct transfer of air pollution particles through the placenta. This study is novel in its inclusion of air pollution exposure levels at occupation locations as well as using such a large selection of air pollution markers. There are currently a limited amount of studies that have measured the effect of period-segregated prenatal air pollution exposure on TL in umbilical cord cells from pregnancies. [32–35]

Mechanisms

As discussed earlier, recent evidence suggests $PM_{2.5}$ from combustion sources may be more harmful than particulate matter from other sources.[36] Although collecting BC exposure data cannot help us differentiate between indoor and outdoor pollution sources, it can be used as indicator of exposure to combustion particles. BC is thought to cause damage in the human body through a series of pathways involving inflammation via direct and systemic toxicity.[23] Systemic toxicity occurs when BC particles trigger an inflammatory response in the lungs causing epithelial cells to secrete inflammatory mediators (chemokines and cytokines). This initial insult results in a continuing inflammatory cascade that can spread to other systems in the body. A recent study showed evidence that BC particles can cross the placental barrier, and have the potential to cause damage to the developing fetus. [37] Additional studies show associations between BC and cancer, respiratory diseases, and cardiovascular disfunction.[23] There are hypothesized mechanisms for the link between air pollution exposure and low birth weight, including intrauterine growth restriction (IUGR). This occurs when the baby does not develop as expected *in utero* due to problems with the placenta, mother's health or birth defects. The placenta is a crucial organ for fetal development and alterations on a molecular level, induced by constituents of air pollution, may be important as early origins of adverse health outcomes. Oxidative stress may be one of the key factors involved in air pollution-induced modifications of the placenta. Reactive oxygen species (ROS) generated by air pollution can directly induce lipid, protein, and DNA damage.[38] The data present compelling evidence that the association between exposure to air pollution and detrimental birth outcomes may act through mechanistic pathways

such as systemic inflammation on the maternal side as well as direct translocation to the placenta on the fetal side.

The literature documents multiple adverse childhood and pregnancy outcomes associated with exposure to air pollution. However, there is heterogeneity in the findings. The heterogeneity is attributable, in part, to the difference in exposure assessment methods, time scales of measurement, or confounding variables. To complicate matters, there is a lack of information on IAQ compared to AAP, even though people in developed countries spend a majority of their time indoors. Emerging research is working to understand how different sub-species of PM, such as BC, act as etiological agents in adverse health outcomes.[39–42]

Indoor Air Quality

Multiple studies have shown poor correlation between ambient $PM_{2.5}$ measurements and personal exposure data illustrating why ambient PM measures are often not suitable proxies for personal exposure.[43–45] Much of this can be attributed to the fact that personal exposure measurements encompass not only ambient PM, but also contributions from indoor and vehicle sources as well as exposure experienced during personal activity. These exposures vary from person to person and throughout time. Often, data regarding indoor air quality (IAQ) is missing in studies investigating associations between air quality and health outcomes. According the US EPA, indoor air quality (IAQ) refers to the air quality within and around buildings and structures, especially as it relates to the health and comfort of building occupants.[46] IAQ problems often originate from indoor pollution sources that release gases or particles into the air. High temperature and humidity as well as inadequate ventilation can increase concentration of some pollutants. Common sources of indoor air pollutants include fuel-burning combustion appliances, tobacco products building materials and furnishings, household cleaning and personal care products, central heating and cooling systems, excess moisture (mold and bacteria) and outdoor sources such as radon, pesticides and particulate matter. As with ambient air pollution, vulnerable populations including children, elderly, low-income and minority communities are often impacted more strongly by poor IAQ.

Indoor Sources of Particulate Matter

Indoor PM is composed of particles of outdoor origin as well as from indoor sources including cooking (especially from unvented gas stoves) and other combustion activities

including candle burning, fuel heaters, cigarette smoke and fireplaces. As discussed earlier, ventilation has often been used as a solution to exchange indoor pollutants with fresh outside air. However, there are situations including high ambient PM and wildfire events in which it may be undesirable to bring polluted or smoky air into a building.

$PM_{2.5}$ is made up of a number of different chemical constituents, with the mixture being source dependent. One of these constituents, BC, originates from the indoor sources mentioned above as well as outdoor sources such as TRAP and industrial combustion activities. The contribution of wildfire smoke to BC concentrations is also important to consider in wildfire prone areas. Due to the fact that PM is often measured by mass, it is difficult to distinguish between indoor and outdoor sources. Methods to distinguish between indoor and outdoor PM source include chemical tracers and positive matrix factorization modeling developed by the US EPA. The positive matrix factorization model takes speciated PM data and uses a mathematical approach to quantify the contribution of sources to samples based on the chemical "fingerprint" of the sources. [47] Although the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE) does set standards for air exchange indoors (minimum 0.35 air changes per hour but not less than 15 cubic feet of air per minute per person),[48] the EPA does not specifically monitor or regulate IAQ. This has led to dearth of information on IAQ compared to what is understood about ambient air pollution.

Infiltration of Ambient PM

Outdoor air generally enters and leaves a building through natural or mechanical ventilation or infiltration. Infiltration occurs when air moves between the indoor and outdoor environment by flowing through cracks, joint and openings in the building. Natural ventilation occurs when wind

and thermal buoyancy (denser cold air lifts warmer air up and out) move air through open windows and doors and between rooms.

The process of mechanical ventilation uses fans to pull air into and/or exhaust air out of a building. Buildings that are designed to minimize the infiltration of outdoor air can suffer from poor indoor air quality if they are lacking specialized mechanical ventilation systems that circulate and filter air. Ventilation and IAQ has become a particularly important topic since early 2020 and the beginning of the SARS-CoV2 pandemic. The spread of COVID-19 (Coronavirus disease 2019) occurs via airborne particles and droplets and being in indoor spaces with poor ventilation can increase risk of infection.[49]

The main objective of the Relationship of Indoor, Outdoor and Personal Air (RIOPA) study was to investigate the contribution of ambient outdoor sources to indoor and personal exposure concentrations.[50] Approximately 100 nonsmoking homes from three distinct geographical locations (Houston, TX; Los Angeles County, CA; and Elizabeth NJ) were sampled across four seasons from summer 1999 to spring 2001. The sampled homes included a mix of ages, styles ventilation methods. Measurements were collected for home volume, air exchange rate, time-activity information, temperature, relative humidity, VOC, aldehydes, PM mass and several PM species concentrations. Of the total of 212 houses sampled for $PM_{2.5}$, 162 were sampled twice. Personal $PM_{2.5}$ mass concentrations were measured on Teflon filters over a 48-hour period and duplicate samples were collected with collocated monitors inside or outside 35 study homes. Average air exchange rates over the 48-hour period were determined by measuring the house volume and the concentration of the inert nontoxic tracer perfluorinated methylcyclohexane.

Median indoor, outdoor and personal $PM_{2.5}$ mass concentrations were found to be 14.4, 15.5 and 31.4 $\mu g/m^3$, respectively. Personal concentrations were found to be more variable and

significantly greater than both indoor and outdoor concentrations. Combined indoor, outdoor and personal $PM_{2.5}$ mass concentrations were poorly to moderately correlated (R^2 = 1-19% for NJ and TX; R^2 = 21-44% for CA) while correlations between indoor and outdoor concentrations were stronger for homes where the ratio of indoor to outdoor concentrations was less than 1 (R^2 = 43-80%; I/O < 1 in 54-71% of homes by state). The higher correlations occur presumably due to less potent indoor sources and/or high exchange rates in these homes. Using a mass balance model approach,[51] the estimate for the mean contribution of ambient outdoor sources to indoor $PM_{2.5}$ concentrations was 8.7 $\mu g/m^3$ or 60% for all study homes. The ambient contribution to personal $PM_{2.5}$ exposure for subjects in the RIOPA study was estimated to be on average 26%. Results from the RIOPA study highlight the importance of collecting information on personal time-activity patterns and indoor PM exposure sources, in addition to data from fixed site ambient monitors.

A study conducted in a non-smoking household in northern Virginia over a period of 2 years (1998-2000) found that the main outdoor sources of BC in the house were from the general regional background (annual average 83-84% weekly) and sources of indoor BC originated from cooking and candle burning (annual average 16 and 31%, respectively).[52] Morning rush hour traffic was found to contribute 8-9% of the total BC and a seasonal evening peak in fall and winter, believed to be from wood burning fireplaces and stoves, contributed approximately 8% of the annual average BC.

CHAPTER TWO: MANUSCRIPT

Introduction

Although monitoring and regulation of air quality has improved in the last several decades, exposure to air pollution, from both ambient and indoor sources, is still a major public health problem worldwide.[1] Globally, exposure to ambient and indoor air pollution is estimated to be responsible for 7 million deaths each year. Almost all of the world's population (99%) lives in places where ambient air pollution exposure exceeds guideline limits established by the World Health Organization (WHO).[53] Exposure to ambient air pollution is considered to be a leading environmental risk factor worldwide. Although the United States showed a marked improvement in ambient air quality between 2000 to 2020 [54], the WHO estimated 1.7 million Disability Adjusted Life Years (DALY's) resulted from exposure to ambient air pollution in the US during that time;[55] the Institute for Health Metrics and Evaluation (IHME) estimated that 60,229 deaths in the US in 2019 could be attributed to ambient and indoor air pollution exposure. Specifically, among US children under the age of 5, air pollution exposure resulted in 4,100 DALY's in 2016.[56] Additional risk also exists for those in areas of lower socioeconomic status, as environmental hazards such as air pollution are often disproportionately distributed across populations. [57] Improved methods for indoor and ambient air pollution exposure assessment are critical to understand inequitable health burden and protect vulnerable populations.

Fine particulate matter ($PM_{2.5}$) and its chemical constituents are commonly measured to estimate exposure to ambient air pollution. The constituents of $PM_{2.5}$, which are source dependent, determine its chemical composition and influence its impact on health. For example, recent research has shown that $PM_{2.5}$ from combustion sources may be more harmful than

particulate matter from other sources.[58] One constituent of $PM_{2.5}$, black carbon (BC), is formed from the incomplete combustion of hydrocarbons and can be used as an indicator of exposure to traffic related air pollution (TRAP). Previous studies demonstrate associations between BC and cancer, respiratory diseases, and cardiovascular dysfunction.[22] BC is also known to have a more heterogeneous spatial distribution than $PM_{2.5}$, and therefore may be a more informative measurement at the intraurban scale.[59] Intraurban heterogeneity in BC concentrations related to TRAP can be useful for understanding differences in exposure across sociodemographic groups.

Air pollution exposure assessment for epidemiological research, particularly for mobile sources, can take various forms. Land Use Regression (LUR) is a popular technique for estimating medium and long-term TRAP exposure.[60] However, personal monitoring has been shown to be a better estimate of TRAP exposure [61] because air pollution exposure is measured in each microenvironment such as school, work, and transportation. Likewise, regional ambient air pollution data is often extrapolated to estimate individual exposure, but multiple studies have shown poor correlation between ambient $PM_{2.5}$ measurements and personal exposure data.[43,44] However, the indoor environment plays a significant role in personal exposure to air pollution that is not accounted for using these approaches. On average, Americans spend approximately 90% of their time indoors.[62] Particularly vulnerable populations such as the very young, elderly and those with cardiovascular and respiratory diseases may spend even more time inside. Unfortunately, representative samples sizes are difficult to achieve for indoor air quality monitoring in residential settings, which requires compliance from multiple households as well as reliable instruments to measure indoor air pollution. In lieu of personal or indoor monitoring, statistical methods to predict indoor air quality may be useful for estimating individual air pollution exposure more accurately than using ambient air quality data alone.

Indoor air quality is influenced by the infiltration of outdoor $PM_{2.5}$, including BC, as well as sources of indoor air pollution, such as cooking on unvented stoves, burning candles, and burning solid fuels for heating. Infiltration and exfiltration occur when air moves between indoor and outdoor environments through unintentional leaks in a building envelope, open doors and windows and mechanical ventilation. To better understand total air pollution exposure, we utilized indoor and outdoor air quality data along with household characteristic survey data from an urban birth cohort study in Colorado[63] to evaluate three predictive modeling techniques to predict indoor BC concentrations using outdoor BC concentrations and household characteristics. Our study objective is to inform epidemiologic studies on environmental influences on early childhood health by investigating how indoor and outdoor BC are related and what additional factors are important in predicting indoor BC concentrations by using ambient BC concentrations.

Methods

Study Population

Healthy Start (5UH3OD023248; PI: Dabalea) is a longitudinal pre-birth cohort study of ethnically diverse mother-infant dyads recruited at obstetrics clinics at the University of Colorado Hospital, which serves a nine-county region of Metropolitan Denver. Full descriptions of the study and recruitment strategy can be found in Harrod *et al.* 2014.[64] Recruitment of pregnant women who had not yet reached 24 weeks of gestation, took place between 2009 and 2014, with the last birth occurring in September 2014. Participants attended follow-up visits through pregnancy, delivery and into early childhood of the offspring. Eligibility criteria included age > 16 years old, singleton pregnancy and no history of chronic disease such as diabetes, cancer, asthma, treated with steroids, or medication-dependent psychiatric illness.

Those that experienced extremely preterm births (<25 weeks gestation) were also excluded. Participants reported their place of residence, age, race/ethnicity, education completed and the number of previous pregnancies at time of enrollment. A total of 1410 mother-child dyads were enrolled from whom data on pre-natal and perinatal outcomes were collected. Additional funding has allowed follow-up of the children and mothers over time to investigate the development of obesity and diabetes, in addition to other health endpoints including neurocognitive and respiratory outcomes. The Healthy Start study protocol was approved by the Colorado Multiple Institutional Review Board.

Study Area and Sampling Locations

At the time of the study, the majority of Healthy Start participants lived primarily in Denver, Arapahoe and Adams Counties. As such, the ambient air pollution monitoring campaigns were designed to estimate ambient air quality in these counties, primarily within the State Highway 470 loop around the metropolitan area. Monitoring sites were selected using a stratified sampling approach as described by Matte *et al*.[58] to determine optimal monitor number and placement to represent regional TRAP sources and sinks. Briefly, the study region was overlaid with a 300m x 300m grid to stratify the area based on common sources of ambient $PM_{2.5}$ (e.g., building density, road density, and traffic density) and to identify areas without coverage by central site monitors operated by the Colorado Department of Public Health and the Environment.[59] Residential monitoring locations were selected to optimize coverage of the study area and oversampled to reflect areas likely to be high sources of PM, while also considering the geographic location of the participants' residences. We conducted spring (Campaign 1), summer (Campaign 2) and winter (Campaign 4) campaigns (May 8 to July 3, 2018; July 10 to August 27, 2018; January 22 to March 12, 2019) to capture seasonal variability. (Due to budgetary

restrictions, households did not participate in Fall campaign (Campaign 3) data collection.) We deployed 52 Ultra Sonic Personal Air Samplers (UPAS; Access Sensor Technologies, Fort Collins, CO) across the study area that included Healthy Start participants' residences in addition to public sites. Paired residential samples were measured during weeks 1 or 2 of each campaign.

Participant Home Visits

A total of 27 households of Healthy Start II participants were selected to host two UPAS for indoor and outdoor measurements. Field team members visited each home at a pre-scheduled date and time to install air quality samplers inside and outside of the participants' homes. The filter-based UPAS allowed us to capture $PM_{2.5}$ and constituents, including BC, metals and reactive oxidative species (ROS). Each outdoor UPAS was outfitted with a custom protective case and an external battery to extend run time to meet a 5-day target sampling period. Indoor monitors were set up indoors in a shared space such as a living room or family room to more accurately represent the entire household's exposure. If the selected room had an outer wall, the outdoor monitor would be collocated on the outside of the house. The field team avoided placing UPAS in rooms that had little foot traffic, were near inflow/outflow vents or air conditioning units, or were to near other $PM_{2.5}$ sources such as near cooking sources in the kitchen, or bedrooms. Outdoor UPAS were installed at adult breathing zone height or slightly above if risk of theft or tampering was perceived to be a concern. Lamps, fence posts, porch beams, and gutters were some of the items on which UPAS were attached via zip tie. Indoor samplers remained in the participants' homes for 5 days, while outdoor samplers stayed for the duration of the outdoor sampling campaign. The field staff also collected urine samples, which were banked for future biomarker analysis. Participants were compensated $25 for hosting a monitor and an

additional $25 for providing a urine sample; participants could receive a maximum of $150 in compensation provided for full participation in all three campaigns (spring, summer and winter).

Survey methods

During study visits, the adult Healthy Start participant in the house responded to an English language questionnaire relating to household characteristics and occupation. The survey, adapted from the EPA's BASE study[65] was designed to gather information about household characteristics that may influence indoor air quality, such as housing type, flooring, heating and cooling methods. Participants were also asked about their occupation and smoking status. Survey data were collected and managed using REDCap[66,67] electronic data capture tools hosted by the Colorado Clinical and Translational Sciences Institute.

Exposure Assessment

Quantification of PM and BC

The UPAS employs a size-selective cyclone inlet to filter out particulate matter of > 2.5 μm in diameter. For this analysis we used 5-day sampling period measurements from participants' homes that hosted paired indoor and outdoor UPAS monitoring to collect BC and $PM_{2.5}$ measurements collected during the spring, summer and/or winter campaigns. We collected particulate samples on polytetrafluoroethylene filters (MTL Corporation, Minneapolis, MN). UPAS flow rate was set to 1L/min and units were run at an 80% duty cycle. Filters were pre-and post-weighed using an Automated Air Analysis Facility (AIRLIFT).[68]

BC analysis methods for this project have been previously described in Martenies *et al*. 2020. [59] Briefly, filters used to collect $PM_{2.5}$ samples were analyzed for BC using SootScan Model OT21 transmissometer (Magee Scientific, Berkeley, CA); we used a previously established protocol described by Ahmed *et al.*[69] to calculate the mass of BC on each filter.

Particle light absorption is calculated by measuring the amount of light attenuated when passed through a sampled filter; BC absorbs light strongly at the 880 nm wavelength. Attenuation at 880 nm was converted to BC density (μg cm^{-2}) by dividing the attenuation by the mass absorbance cross section σ_{880} ($m^2\ g^{-1}$). The final mass of BC on the filter was obtained by multiplying the BC density by the filter sampling area (α = 7.065 cm^2). We used a value of 4.2 for the mass-absorbance cross section based on analysis by Presler-Jur *et al.*[70]

We deployed field blanks during each sampling campaign in order to assess potential contamination of filters. Measurements of BC from Campaign 1 and 2 showed a negative bias (mean BC measurements were -4.0 ug and -2.6 ug, respectively, with coefficients of variation (CV) < 25%); therefore, BC measurements from these campaigns were blank corrected.[59] The limit of detection (LOD) value for BC was 1.41 μg based on the lower limit for the SootScan (0.2 μg/cm^2) and the standard area of our filters (7.065 cm^2). Filters below LOD, as well as filters with evidence of contamination (i.e., difference in pre- and post-sampling filter $PM_{2.5}$ mass exceeded 1000 μg) were removed from the analysis.

Variability of BC measurements was evaluated at each site by campaign. As described in Martenies *et al*. [59], we expected to see higher variability in winter outdoor BC measurements due to the effect of cold on the sampling equipment (i.e., battery life sometimes decreased in the presence of lower ambient temperatures). The coefficient of variation was used as our criterion to evaluate variability in UPAS measurements collected at each site by campaign. Any observation that had values that exceeded the coefficient of variation for BC (0.30) during the sampling campaign was dropped.

Statistical Analysis

The overall aims of this analysis were to assess the seasonality of BC levels indoors and outdoors and build the best predictive model given the available set of covariates. Within the context of the models, we also hoped to better understand which factors were predictive of indoor BC levels.

One of our main goals was to gain a better understanding of factors that are associated with BC levels inside homes, as people spend a majority their time indoors. We were also interested in evaluating seasonal differences in BC concentrations, which are expected to change seasonally both indoor and outdoor due to fireplace/woodstove use, opening of windows, wildfires and seasonal weather pattern variations. To test for significance differences in BC concentrations across seasons, we used the Kruskal-Wallis rank sum test. We used Wilcoxon rank sum tests to evaluate the relationship between indoor and outdoor BC concentrations across the entire study period. Wilcoxon signed rank tests were also applied to evaluate the relationship between mean indoor and outdoor BC concentrations for the same season. Significance was determined using an alpha of 0.05.

To understand the relationship between indoor and outdoor BC concentrations and build the best predictive models, we employed three modeling techniques and evaluated their performance. Two of these modeling approaches were machine learning techniques: Lasso (least absolute shrinkage and selection operator) and Ridge regression, to reduce model complexity and prevent over-fitting.[71] Lasso and Ridge modeling techniques (also known as L1 and L2 regularization, respectively) are similar as they aim to reduce over-fitting by using a penalty parameter that increases bias, thus introducing a reduction in variance. In machine learning techniques bias is defined as the inability of the modeling method to capture the true relationship

between the variables, while variance represents the difference in fits between data sets.[71] The ideal algorithm will have low bias and low variability and produce consistent predictions across different data sets.

Although the two modeling techniques are useful in helping to find balance between bias and variance, simplicity and complexity, there are important differences between the two. Both Ridge and Lasso regression use a penalty parameter lambda (λ) to penalize regression coefficients, however they differ in their penalty function. In linear regression, the cost function is the average error of n-samples in the data, represented by the root mean squared error (RMSE) or mean squared error (MSE). The cost function for Ridge regression is the sum of squared residuals plus the penalty term:

$$L_{hridge}(\hat{\beta}) = \sum_{i-1}^{n}(y_i - x_i'\hat{\beta})^2 + \lambda \sum_{j-1}^{m} w_j \hat{\beta}_j^2. \qquad \text{(Equation 1)}$$

The severity of the penalty is determined by λ, which is selected with cross validation of multiple values to determine which one results in the lowest variance. If λ = 0, then the equation is the same as the ordinary least squares (OLS) linear regression model; if λ > 0, then it will add a constraint to the coefficient. As the constraint increases, the value of the coefficient will *tend toward* zero, shrinking coefficients of correlated predictors towards each other. Ridge regression never reduces a coefficient to zero, only to near zero; therefore, it will decrease the complexity of the model but will not reduce the number of variables. For this reason, Ridge therefore does better when most variables in the model are "useful" or provide information on the relationship prediction variables and outcome.[72] Ridge regression is also a good tool for improving predictions when working with small samples sizes as it makes predictions less sensitive to training data due to increasing bias on the initial fit.

The cost function for Lasso regression is the sum of squared residuals plus the penalty term, which is the sum of the absolute value of the estimated regression coefficients.

$$L_{lasso}(\hat{\beta}) = \sum_{i-1}^{n}(y_i - x_i'\hat{\beta})^2 + \lambda \sum_{j-1}^{m} |\hat{\beta}_j|.$$

(Equation 2)

As with ridge regression, the severity of the penalty for Lasso is determined by λ. However, when is $\lambda > 0$, as the constraint increases the coefficient value will decrease until it *reaches* zero. The Lasso penalty is known to pick one of a group of coefficients of correlated predictors and discard the others. Thus, Lasso will reduce variables in the model, making it a viable option variable selection. This approach may be helpful when working with a dataset with a lot of noise variables or when there are too many predictor variables to have prior knowledge of which might be associated with the outcome.

For both Ridge and Lasso models, we used two functions from the R package `glmnet`, `lambda.min`( ) and `lambda.lse`( ), to choose the λ parameter.[73] The `lambda.min`( ) function calculates λ of the minimum mean cross-validated error, while `lamda.lse`( ) uses the largest value of λ such that the error is within 1 standard deviation of the cross-validated errors for `lambda.min`( ). With these two approaches for selecting λ, two models each were fit for both Ridge and Lasso.

The third predictive modeling approach we used was to fit a "clinical" model with multiple linear regression, based on factors associated with indoor BC concentrations that have been previously described in the literature including infiltration of outdoor particulate matter, wood burning, seasonal change, and residential building characteristics.[74] Factors examined in the modeling process are listed in Table 1 and include indoor air quality ($PM_{2.5}$, BC) and housing

characteristics (number of rooms, home ownership, presence of hard flooring, gas appliances, window AC use, presence of pets in home, and supplemental heat use).

Cross-validation is often used to evaluate different modeling techniques and compare their predictive performance. A data set is split into a subset used to "train" the algorithm and a subset used to "test" the modeling parameters developed in the training phase. Using only one training set can be problematic as the MSE obtained from the testing phase can vary greatly depending on which observations were used in the testing and training sets. K-fold cross-validation is a method of cross-validation which splits the observed data into *k* folds or groups.[75] Then, one group is selected as the testing data and the remaining k-1 groups are used for training each model. Predictive performance of the trained model is then evaluated on the testing data. This process is repeated using each fold for testing data and the predictive performance scores are averaged into a comprehensive validation score.

For our analysis, we implemented leave-one-out cross-validation (LOOCV), which can be more computationally expensive but is a good fit for small datasets as it allows for the smallest amount of data removed from the training set in each iteration (*n-1*). In this method of cross-validation, the dataset is split into a training set and a testing set using all but one observation as part of the training set. The training data is used to build the model and predict the response value of the "left out" observation. The most common method for evaluating the performance of a model on the dataset is calculating the mean squared prediction error (MSPE) to measure how well the predictions match the observed data. The closer model predictions are to observations, the lower the MSPE. The MSPE is calculated with each testing set, giving *n* values of MSPE. The LOOCV estimate for the test MSPE is the average of all *n* MSPE values.

The final model was chosen based on the lowest MSPE. A final multiple linear regression model was then fit for inference.

Prior to modeling, indoor BC concentrations were log transformed. All data cleaning and statistical analyses were conducted in R v.3.6.2 (R Core Team, 2019), primarily using the `tidyverse`[76] and `glmnet`[77] packages.

Results

Measurements and Housing Characteristics

A total of 27 homes were sampled over the three campaigns. Additional households were recruited to replace two that were lost to follow-up during the spring or summer campaigns, for a total of 25 homes for each sampling period.

We collected a total of 73 indoor/outdoor pairs (spring = 25, summer = 24, winter = 24) of residential filters. Nine filter pairs were removed from analysis due to outdoor $PM_{2.5}$ concentrations below LOD (n = 8) or potential contamination (PM mass > 1000ug, n =1). Additional pairs (n = 10) were removed that had indoor $PM_{2.5}$ concentrations below LOD or a BC coefficient of variation > 0.30 (n = 8, all from winter campaign). Due to the high variability and low confidence in the winter samples all winter data (n = 24) were excluded from analysis.

Following data cleaning, 39 measurements from 27 homes remained for the final analysis. Of these, 21 (54%) were collected in spring and 18 (46%) in summer. Table 1 presents summary statistics of the homes used in the study. Survey data from the 27 homes indicate that 69.2% of participants were home owners, 43.6% had gas appliances and 51.3% had pets in the home. All but one home had carpet, (97.4%) so this variable was removed from the analysis. In addition to carpet, 53.8 % of homes had more than two types of hard flooring. Table 2 presents sample measurement data stratified by season. The median indoor BC concentration was 1.00

$\mu g/m^3$ (0.76-1.42) during the spring and 0.87 $\mu g/m^3$ (0.67-1.12) during the summer, with an overall median of 0.98 $\mu g/m^3$. Homes monitored during the spring had a ratio of median indoor to median outdoor BC concentrations equal to 0.99, while homes monitored during summer had a ratio of median indoor to median outdoor BC concentrations equal to 0.76.

Using a Wilcoxon rank-sum test, neither median indoor ($p = 0.31$) or outdoor ($p = 0.86$) BC concentrations showed a significant difference by season (Table 2). We also found no significant difference between paired indoor and outdoor BC concentrations by season using a Wilcoxon signed-rank test (spring = 0.56, summer = 0.18; Table 3).

Model Performance

In terms of predictive performance, the clinical model had an MPSE of 0.054 and estimates for the four covariates selected *a priori* are as follows: outdoor $PM_{2.5}$ (β=0.027), outdoor BC (β=0.052), dirty supplemental heat (wood stove, fireplace or kerosene heater) (β=-0.038) and spring (β=0.027). The Lasso model obtained an MPSE of 0.064 and selected three predictors: outdoor $PM_{2.5}$ (β=0.013), presence of pets in home (β=0.064), and presence of greater than two types of hard flooring (β=0.033). The Ridge LSE model obtained the best predictive performance with a MPSE of 0.50 (Table 6). Since Ridge does not eliminate covariates and cannot be used specifically for feature selection, we chose the four predictors with the largest coefficients to go in our final model: outdoor $PM_{2.5}$ concentration (β=0.014), outdoor BC concentration (β=0.041), presence of pets in home (β=0.094) and presence of greater than two types of hard flooring (β=0.092). A linear regression model with these four covariates was fit for inference (Table 4). The R^2 for the final model was 0.27. Households that had more than two types of hard flooring had a 0.09 (95% CI: -0.48, 0.23) unit increase in log-transformed indoor BC concentrations compared to homes that had one or fewer types of hard flooring. Households

with pets in the home had a 0.12 (95% CI: -0.02, 0.25) unit increase in log-transformed indoor BC compared to homes with no pets. Figure 2 shows correlation plots regressing model predicted BC concentrations on observed BC values. R represents the coefficient of correlation, or the degree of the relationship between the x and y variables, and can vary between -1 and 1. The p-value shown is from the overall analysis of variance report. The report is the overall test to determine the significance of the entire model. The Ridge LSE model plot has the highest R at 0.36 and the lowest p-value of 0.026.

Discussion

Human exposure to BC is likely to be associated with air quality of indoor environments, due to the fact that people spend most of their time inside. Most studies assign individual exposure using data from regional ambient monitors as it is not often feasible to measure residential indoor BC in a large-scale study. However, if outdoor air pollution is associated with indoor air pollution, we can estimate total air pollution exposure based on models as knowledge about household infiltration and indoor sources of air pollution. Focusing on black carbon holds great utility for this exercise, as BC is considered to be a leading etiological agent for many adverse health endpoints associated with air pollution. Black carbon also demonstrates strong spatial gradients within an urban setting to help differentiate places of high and low exposure. Further, traffic is considered to be the dominant source of BC in the study area; understanding indoor sources of BC as well as housing characteristics can help estimate total exposure.

The results of our predictive modeling strategies demonstrate outdoor air pollution measurements along with housing characteristic data can explain approximately 27% of the variability in indoor BC concentration. Baxter *et.al* used a linear modeling strategy to predict indoor elemental carbon (EC) and found an $R^2 = 0.32$.[78] They used publicly available central site

monitor and GIS data, property assessment records and questionnaire responses from lower socioeconomic status (SES) households. This study was conducted with a subset of participants of a prospective birth cohort in Boston, MA. Sampling was conducted in two seasonal periods – heating (December – March) and non-heating (May-October). Although indoor sources were identified for $PM_{2.5}$ (cooking time) and NO_2 (stove usage), no indoor source was identified for EC. The truck traffic indicator proved to be an important factor in the modeling approach but was strongly modified by the variable "windowsopen" which was collected from the survey question asking if participants did or did not open their windows. In a similar study by Isiugo *et.al,* researchers developed a predictive model for BC that explained 78% of the variability in indoor BC concentrations. The presence or absence of electrostatic/HEPA HVAC filters was one of the most important factors influencing indoor BC concentration – homes with filters had an indoor/outdoor BC ratio of 0.81 while homes without had a ratio of 0.62. Both of these studies point to the importance of correctly capturing the heterogeneity in outdoor BC concentrations as BC tends to be dominated by outdoor sources.

While our three different modeling strategies didn't perform exactly the same, they were quite similar with MPSE's ranging from 0.050 with Ridge LSE to 0.064 with Lasso. Ridge is a useful modeling tool when working with multiple covariates that offer some explanation regarding the variability of the outcome while Lasso is better at eliminating noise variables. In this study, many of the covariates can at least partially explain the variability of the outcome of indoor BC concentrations. Ridge is better at handling multi-collinearity, or correlation between predictor variables. When predictor variables in the same regression model are correlated, they cannot independently predict the value of the dependent variable. Lasso will usually remove some of the collinear variables in the model while Ridge will keep all collinear variables while

shrinking some of their coefficients. Although Ridge does not eliminate predictor variables and reduce model complexity, it may be the better statistical tool for predicting indoor BC concentrations due to the information on covariates that is retained.

Applications of a Predictive Model

A predictive model for indoor BC would be a helpful tool to support public health risk assessment objectives and to reduce exposure misclassification in studies that assign individual BC exposure levels. Cumulative exposure is a combination of indoor concentration, outdoor concentration and time-activity patterns related to those environments. Collecting indoor BC measurements is a time and cost burden that could potentially be avoided by collecting household survey data and information on time-activity patterns. However, results from this study, as well of those from Baxter *et al.* and Isiugo *et. al.*, demonstrate that building characteristic data and other information such as distance to a highway, may not explain a sufficient amount of variability in indoor BC measurements. In addition, estimating only residential exposure does not capture exposure from other microenvironments including school, work and commuting. With the increasing availability of reliable, low-cost monitors, personal monitoring may be a more robust and feasible method to estimate individual exposure to BC. Additional questions regarding indoor BC sources could be addressed by placing a sampler inside the home during the personal monitoring period. Improvements in monitor technology, such as GPS functionality and real-time PM sensors will help us better understand exposure by microenvironment.

Limitations

We had limited comparable data on seasonality as winter data were excluded from the analysis. Shortened runtimes and other performance issues suggest that the poor data quality of

winter BC measurements may be explained in part by impaired function and reliability and insufficient weatherproofing of our monitors in cold weather. Thus, the only two seasons we had for comparison were summer and spring, which overall did not show a significant difference in indoor or outdoor BC concentrations. Other studies have shown outdoor BC to be negatively correlated with temperature and wind speed with higher average concentrations found during the winter. [59,79] Temperature inversions that trap pollution and wood burning to heat homes are thought to be related to the winter increases.

Information on housing characteristics was limited and did not include variables that were shown to be informative in an indoor BC predictive modeling study by Isiugo *et.al* such as kitchen stove ventilation status, candle use during the study period and type of HVAC filter used in the home. For future studies, we would also like to include more specific information on type of ventilation used (open windows, fans, HVAC, etc.) as well as percent of flooring area covered in either carpet or hard floors such as tile or concrete. Although rate of infiltration and exfiltration through the building envelope are important variables when considering predictors of indoor air quality, gathering that information directly would not be feasible for large-scale studies since it would require individual assessment at each participant residence. However, estimates of infiltration can be discerned from existing models including factors such as building age, size and climate zone.[80] Distance of a homes from nearby roads with high vehicular, particularly diesel traffic, may also modify indoor BC concentrations and will be incorporated into future modeling strategies.

While there is sufficient evidence in the literature to suggest that personal BC exposure estimates are often poorly represented by ambient air quality data, there are limited published works evaluating modeling strategies to predict indoor BC concentrations. We conducted this

study with the intent to investigate the relationship between indoor and outdoor black carbon concentrations, evaluate seasonal differences in BC concentrations and explore housing characteristics that may play a role in indoor BC concentrations. What we found is the ratio of indoor to outdoor BC concentrations was just below one in both spring and summer, with indoor slightly lower than outdoor concentrations. While we did not find a significant difference in concentrations between seasons, we acknowledge that having only viable data from spring and summer seasons (both considered "non-heating" seasons) limited us in detecting differences that would be seen between colder and warmer months. Using three predictive modeling approaches, including Ridge and Lasso regression, we developed a final model from our data to predict indoor BC concentrations using housing characteristics and outdoor BC and $PM_{2.5}$ concentrations. The R^2 for the final model was 0.27 and the four covariates were outdoor $PM_{2.5}$, outdoor BC, presences of more than two types of hard flooring and presence of pets in the home. We hypothesize that the presence of more than two types of hard flooring may be indicative of less carpeted surfaces, and perhaps carpeted surfaces capture more PM that is filtered out when vacuumed. Asking more specific questions about vacuuming habits and percentage of flooring that is carpeted would be helpful information for future projects. Pets the home as a factor predicting indoor BC concentrations may be related to ventilation – pet owners may be more likely to leave windows or doors open so their pets can easily go between the house and the yard, or they may be opening doors more frequently to let their pets outside. We could further improve our survey by asking more specific questions regarding how often door is opened for pets per day and if windows or door are left open.

We could further develop our modeling by using differing variable selection methods that allow for more than 2-level categorical covariates, investigate interaction terms or potential non-

linear relations. A specific tool for variable selection, Bayesian modeling, uses conditional probability, or the probability of an event A given event B, and can be calculated using the Bayes rule. The Bayes rule is used to compute and update probabilities after obtaining new data. Bayesian analysis can be useful for small samples, as it provides inferences that are conditional on the data and are exact, with no reliance on asymptomatic approximation.[81]

Regarding what we chose as our "final" model – this was designed specifically so we could make inference regarding factors associated with indoor BC. The Ridge LSE had the strongest predictive performance and we would suggest using that model if the main objective is to predict indoor BC with similar factors as we used in our study.

CHAPTER 3: CONCLUSION

Study Findings

There is limited literature available evaluating statistical modeling to predict indoor BC concentrations. We conducted this study with the intent to investigate the relationship between indoor and outdoor BC concentrations, evaluate seasonal differences in BC concentrations and explore housing characteristics that may play a role in indoor BC concentrations. What we found is the ratio of indoor to outdoor BC concentrations was just below one in both spring and summer, with indoor slightly lower than outdoor concentrations. While we did not find a significant difference in concentrations between seasons, we acknowledge that having only viable data from spring and summer seasons (both considered "non-heating" seasons) limited us in detecting differences that would be seen between colder and warmer months. Using three predictive modeling approaches, including Ridge and Lasso regression, we developed a final model from our data to predict indoor BC concentrations using housing characteristics and outdoor BC and $PM_{2.5}$ concentrations. The R^2 for the final model was 0.27 and the four covariates were outdoor $PM_{2.5}$, outdoor BC, presences of more than two types of hard flooring and presence of pets in the home.

Future Research

Accurately estimating personal exposure to air pollution is not a simple measurement. However, as noted in our research, simply using ambient estimates to capture a person's daily exposure does not account for pollutants they may be exposed to during a commute, at home or at work. Exposure misclassification errors are introduced when the spatiotemporal mobility and indoor exposure of study subjects are ignored. I would like to further develop our predictive model,

using more detailed information such as distance to nearest highway, infiltration rates, HVAC usage and filter type, as well as cooking activities, and ventilation status (windows open or closed, stove ventilation). A further area or interest is evaluating the differences in IAQ between different socioeconomic groups and the resulting health impacts. People of color and those struggling with poverty are often disproportionately affected by poor ambient and indoor air quality. Their housing options are more likely to be closer to highways, agriculture or industrial areas and their indoor spaces often lack proper ventilation. For example, in farming communities in Fresno, CA, many of the agriculture workers and their families live within 200 meters of a field or orchard. Farming activities such as aerial pesticide application and harvesting of crops can lead to drifting chemicals and clouds of dust. In addition to already poor air quality in this area, residents are subjected to drifting mix of pesticides and dust and a majority to do not have HVAC systems that would filter out such contaminants. When temperatures climb in the spring and summer residents are left with the option of keeping their windows closed and enduring the sweltering heat, or opening up windows and doors through which the mixture of dust and chemicals blow in with the breeze. Research that further examines the factors that contribute to personal air pollution exposure provides tools to public health professionals regarding steps that can taken to protect vulnerable populations.

Post Thesis Reflections

I have always had a varied interest in environmental and human health issues and I found the Environmental Health program with focus in epidemiology to be a good fit with my wide-ranging interests. The way I see it, environmental and human health are one. As humans damage or destroy our environment, we often harm ourselves by creating unhealthy living conditions. Advances in scientific research have certainly helped the world community understand this

relationship more clearly over the last 50 years. Although some progress has been made, the human community still struggles to live harmoniously with the natural world. It is now clear that climate change is no longer a future scientific prediction; it is here now. In addition to working to mitigate further temperature increases, we need to learn to adapt to the extreme heat waves, fires floods and droughts that are occurring. All of these events can lead to a drastic decrease in ambient and indoor air quality, especially in poorer and underserved communities. As I continue my work as a research associate and lab manager in Dr. Magzamen's group, I want to further my knowledge and expertise translating research findings into useful measures to protect and improve community and environmental health. I also want to improve the ability of the public to access and understand relevant research findings as well as empower people to participate in citizen science research. I hope to accomplish this by working in the field to collect data while simultaneously speaking with communities to educate them on the environmental health hazards they face, and what they can do about it. I also plan to volunteer with local Fort Collins groups such as the Quarter Project that promotes females in STEM (science, technology, engineering, and math) and offers resources and experience to primarily minority and low-income girls. I will continue to develop my skillset in study design, data analysis, field methods writing and communication in an effort to have a positive impact as an environmental epidemiologist and community member.

TABLES & FIGURES

Table 1. Characteristics of households from which particulate matter was measured.

	Overall (N = 39)
Number of Rooms	
Mean (SD)	9.82 (3.36)
House	
no	8 (20.5%)
yes	31 (79.5%)
Own home	
no	12 (30.8%)
yes	27 (69.2%)
Carpet	
no	1 (2.6%)
yes	38 (97.4%)
Hard flooring (2 or more types)	
no	18 (46.2%)
yes	21 (53.8%)
Gas appliances	
no	22 (56.4%)
yes	17 (43.6%)
Window AC	
no	35 (89.7%)
yes	4 (10.3%)
Smoker in home	
no	34 (87.2%)
yes	5 (12.8%)
Pets in home	
no	19 (48.7%)
yes	20 (51.3%)
Dirty supplemental heat (wood stove, wood fireplace, kerosene gas)	
no	35 (89.7%)
yes	4 (10.3%)

Table 2. Median BC and $PM_{2.5}$ measurements ($\mu g/m^3$) by season

	Spring (N=21)	**Summer (N=18)**	**Overall (N=39)**
Indoor BC			
Median [Min, Max]	1.00 [0.25, 2.25]	0.871 [0.45, 4.65]	0.978 [0.25, 4.65]
Outdoor BC			
Median [Min, Max]	1.01 [0.36, 2.12]	1.15 [0.052, 2.57]	1.10 [0.052, 2.57]
Indoor PM2.5			
Median [Min, Max]	13.8 [6.13, 32.4]	11.9 [2.83, 75.1]	12.3 [2.83, 75.1]
Outdoor PM2.5			
Median [Min, Max]	9.15 [7.19, 17.1]	10.3 [3.48, 18.6]	9.62 [3.48, 18.6]
Indoor/outdoor BC ratio	0.990	0.757	0.843

Table 3. Model Summary – comparison of coefficients from lasso, ridge and clinical models. For ridge and lasso, minimum of mean cross-validated errors is used to calculate λ, while LSE models use the largest value of λ such that the error is within 1 standard error of the cross-validated minimum.

Covariates/Predictors					
	Lasso	**Lasso LSE**	**Ridge**	**Ridge LSE**	**Clinical**
Outdoor PM	0.013	2.15×10^{-17}	0.016	0.014	0.027
Indoor PM			0.004	0.003	
Outdoor BC			0.045	0.041	0.052
Rooms			-0.003	-0.002	
House			0.084	0.066	
Own home			-0.054	-0.046	
Hard flooring (>2 types)	0.033		0.127	0.092	
Gas appliance			-0.058	-0.022	
Window AC unit			0.078	0.052	
Smokes			-0.044	-0.013	
Pets in home	0.064		0.119	0.094	
Dirty supplemental heat			-0.038	-0.039	-0.038
Summer			-0.006	-0.009	
Spring			0.006	0.009	0.027
MPSE	**0.064**	**0.053**	**0.056**	**0.050**	**0.054**

Figure 1. UPAS with weather resistant enclosure mounted to participant home.

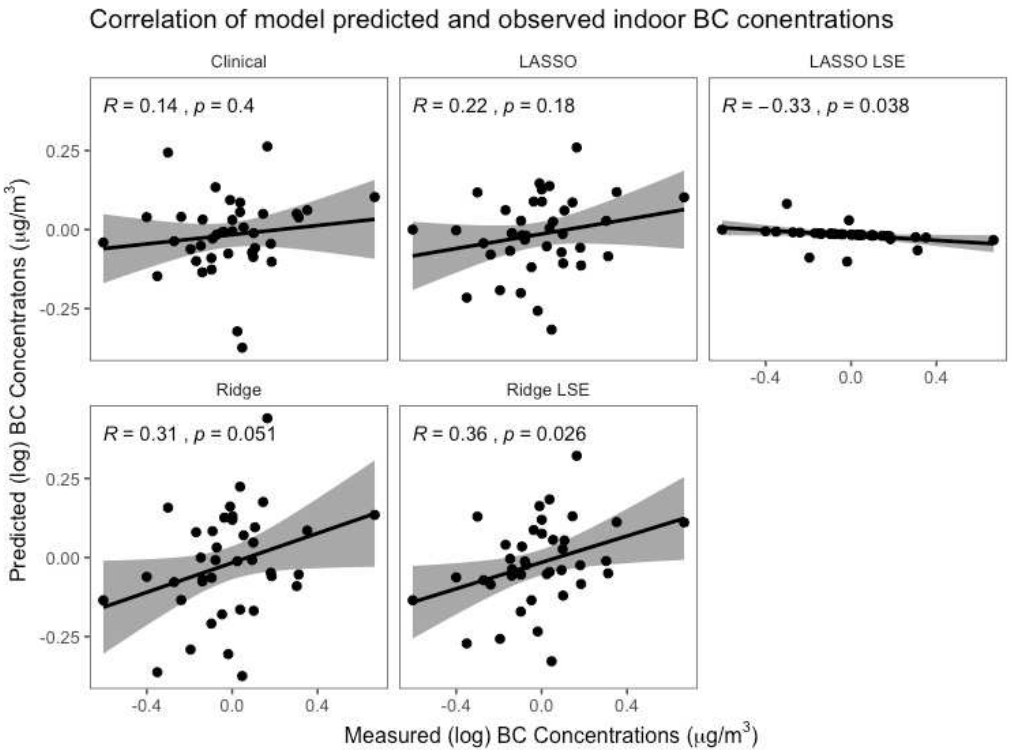

Figure 2. One-to-one correlation plots of Clinical, LASSO, LASSO LSE, Ridge Regression and Ridge Regression LSE predictions of log indoor BC concentrations for all measurements sites, spring and summer campaigns.

Table 4. Summary of linear regression model fit for inference

	Estimate	**Std. Error**	**p-value**	**95% Confidence Interval**
Intercept	-0.403	0.113	0.001	(-0.634, -0.174)
Outdoor $PM_{2.5}$	0.023	0.013	0.093	(-0.004, 0.051)
Outdoor BC	0.036	0.070	0.613	(-0.107, 0.178)
Hard flooring (> than two types)	0.090	0.068	0.192	(-0.475, 0.228)
Pets in home	0.117	0.067	0.091	(-0.020, 0.253)

REFERENCES

1. Air pollution. *Environmental history* https://environmentalhistory.org/about/airpollution/ (2017).
2. The Deadly Donora Smog of 1948 Spurred Environmental Protection—But Have We Forgotten the Lesson? | History | Smithsonian Magazine. https://www.smithsonianmag.com/history/deadly-donora-smog-1948-spurred-environmental-protection-have-we-forgotten-lesson-180970533/.
3. PubMed Central Full Text PDF.
4. Great Smog of London | Facts, Pollution, Solution, & History. *Encyclopedia Britannica* https://www.britannica.com/event/Great-Smog-of-London.
5. The Great Smog of 1952. *Met Office* https://www.metoffice.gov.uk/weather/learn-about/weather/case-studies/great-smog.
6. Document Display | NEPIS | US EPA. https://nepis.epa.gov.
7. Air pollution and health. *Bull N Y Acad Med* **42**, 588–619 (1966).
8. US EPA, O. Evolution of the Clean Air Act. *US EPA* https://www.epa.gov/clean-air-act-overview/evolution-clean-air-act (2015).
9. US EPA, O. NAAQS Table. https://www.epa.gov/criteria-air-pollutants/naaqs-table (2014).
10. Dominici, F. *et al.* Chemical Composition of Fine Particulate Matter and Life Expectancy. *Epidemiology* **26**, 556–564 (2015).
11. Inhalable Particulate Matter and Health (PM2.5 and PM10) | California Air Resources Board. https://ww2.arb.ca.gov/resources/inhalable-particulate-matter-and-health.

12. Sram, R. J., Binkova, B. B., Dejmek, J. & Bobak, M. Ambient air pollution and pregnancy outcomes: A review of the literature. *Environ. Health Perspect.* **113**, 375–382 (2005).
13. Loomis, D., Castillejos, M., Gold, D. R., McDonnell, W. & Borja-Aburto, V. H. Air pollution and infant mortality in Mexico City. *Epidemiology* **10**, 118–123 (1999).
14. Pereira, L. A. *et al.* Association between air pollution and intrauterine mortality in São Paulo, Brazil. *Environ Health Perspect* **106**, 325–329 (1998).
15. Wang, X., Ding, H., Ryan, L. & Xu, X. Association between air pollution and low birth weight: a community-based study. *Environ Health Perspect* **105**, 514–520 (1997).
16. Ha, E. H. *et al.* Is air pollution a risk factor for low birth weight in Seoul? *Epidemiology* **12**, 643–648 (2001).
17. Dadvand, P. *et al.* Maternal Exposure to Particulate Air Pollution and Term Birth Weight: A Multi-Country Evaluation of Effect and Heterogeneity. *Environ. Health Perspect.* **121**, 367–373 (2013).
18. Klepac, P., Locatelli, I., Korošec, S., Künzli, N. & Kukec, A. Ambient air pollution and pregnancy outcomes: A comprehensive review and identification of environmental public health challenges. *Environmental Research* **167**, 144–159 (2018).
19. Wells, G. *et al.* The Newcastle-Ottawa Scale (NOS) for Assessing the Quality of Nonrandomised Studies in Meta-Analyses. *undefined* (2014).
20. Lv, Y., Wang, H., Wei, S., Zhang, L. & Zhao, Q. The Correlation between Indoor and Outdoor Particulate Matter of Different Building Types in Daqing, China. *Procedia Engineering* **205**, 360–367 (2017).

21. Janssen, N. A. H. *et al.* Black Carbon as an Additional Indicator of the Adverse Health Effects of Airborne Particles Compared with PM10 and PM2.5. *Environ. Health Perspect.* **119**, 1691–1699 (2011).
22. Thurston George D. *et al.* Ischemic Heart Disease Mortality and Long-Term Exposure to Source-Related Components of U.S. Fine Particle Air Pollution. *Environmental Health Perspectives* **124**, 785–794 (2016).
23. Niranjan, R. & Thakur, A. K. The Toxicological Mechanisms of Environmental Soot (Black Carbon) and Carbon Black: Focus on Oxidative Stress and Inflammatory Pathways. *Front Immunol* **8**, 763 (2017).
24. Brauer Michael *et al.* A Cohort Study of Traffic-Related Air Pollution Impacts on Birth Outcomes. *Environmental Health Perspectives* **116**, 680–686 (2008).
25. Respiratory Minute Volume - an overview | ScienceDirect Topics.
26. Sapkota, A., Chelikowsky, A. P., Nachman, K. E., Cohen, A. J. & Ritz, B. Exposure to particulate matter and adverse birth outcomes: a comprehensive review and meta-analysis. *Air Qual. Atmos. Health* **5**, 369–381 (2012).
27. Reid, C. E. & Maestas, M. M. Wildfire smoke exposure under climate change: impact on respiratory health of affected communities. *Curr Opin Pulm Med* **25**, 179–187 (2019).
28. Holstius, D. M., Reid, C. E., Jesdale, B. M. & Morello-Frosch, R. Birth Weight following Pregnancy during the 2003 Southern California Wildfires. *Environ. Health Perspect.* **120**, 1340–1345 (2012).
29. Chiu, Y.-H. M. *et al.* Prenatal particulate air pollution exposure and body composition in urban preschool children: Examining sensitive windows and sex-specific associations. *Environ. Res.* **158**, 798–805 (2017).

30. Kannan, S., Misra, D. P., Dvonch, J. T. & Krishnakumar, A. Exposures to Airborne Particulate Matter and Adverse Perinatal Outcomes: A Biologically Plausible Mechanistic Framework for Exploring Potential Effect Modification by Nutrition. *Environmental Health Perspectives* **114**, 1636–1642 (2006).
31. Scholten, R. H. *et al.* Telomere length in newborns is associated with exposure to low levels of air pollution during pregnancy. *Environ. Int.* **146**, 106202 (2021).
32. Lee, A. G. *et al.* Prenatal particulate air pollution and newborn telomere length: Effect modification by maternal antioxidant intakes and infant sex. *Environmental Research* **187**, 109707 (2020).
33. Martens, D. S. *et al.* Prenatal Air Pollution and Newborns' Predisposition to Accelerated Biological Aging. *JAMA Pediatrics* **171**, 1160–1167 (2017).
34. Rosa, M. J. *et al.* Association between prenatal particulate air pollution exposure and telomere length in cord blood: Effect modification by fetal sex. *Environmental Research* **172**, 495–501 (2019).
35. Song, L. *et al.* Effects of maternal exposure to ambient air pollution on newborn telomere length. *Environment International* **128**, 254–260 (2019).
36. Hime, N. J., Marks, G. B. & Cowie, C. T. A Comparison of the Health Effects of Ambient Particulate Matter Air Pollution from Five Emission Sources. *Int. J. Environ. Res. Public Health* **15**, 1206 (2018).
37. Bove, H. *et al.* Ambient black carbon particles reach the fetal side of human placenta. *Nat. Commun.* **10**, 3866 (2019).
38. Saenen, N. D. *et al.* Air pollution-induced placental alterations: an interplay of oxidative stress, epigenetics, and the aging phenotype? *Clinical Epigenetics* **11**, 124 (2019).

39. MohanKumar, S. M. J., Campbell, A., Block, M. & Veronesi, B. Particulate matter, oxidative stress and neurotoxicity. *Neurotoxicology* **29**, 479–488 (2008).
40. Morales-Rubio, R. A. *et al.* In utero exposure to ultrafine particles promotes placental stress-induced programming of renin-angiotensin system-related elements in the offspring results in altered blood pressure in adult mice. *Part. Fibre Toxicol.* **16**, 7 (2019).
41. Salinas, P., Veuthey, C., Rubio, I., Bongiorno, A. & Romero, I. Exposure to Wood Smoke Pollution During Pre-Gestational Period of Rat has Effects on Placenta Volume and Fetus Size. *Int. J. Morphol.* **38**, 1356–1364 (2020).
42. Bolton, J. L. *et al.* Prenatal air pollution exposure induces neuroinflammation and predisposes offspring to weight gain in adulthood in a sex-specific manner (vol 26, pg 4743, 2012). *Faseb J.* **28**, 2737–2738 (2014).
43. Lachenmyer, C. Urban Measurements of Outdoor-Indoor PM2.5 Concentrations and Personal Exposure in the Deep South. Part I. Pilot Study of Mass Concentrations for Nonsmoking Subjects. *Aerosol Science and Technology* **32**, 34–51 (2000).
44. Oglesby, L. *et al.* Validity of ambient levels of fine particles as surrogate for personal exposure to outdoor air pollution--results of the European EXPOLIS-EAS Study (Swiss Center Basel). *J Air Waste Manag Assoc* **50**, 1251–1261 (2000).
45. Sexton, K., Spengler, J. D. & Treitman, R. D. Personal exposure to respirable particles: A case study in Waterbury, Vermont. *Atmospheric Environment (1967)* **18**, 1385–1398 (1984).
46. US EPA, O. Introduction to Indoor Air Quality. https://www.epa.gov/indoor-air-quality-iaq/introduction-indoor-air-quality (2014).
47. Norris, G., Duvall, R., Brown, S. & Bai, S. Positive Matrix Factorization (PMF) 5.0 Fundamentals and User Guide. 136.

48. Standards 62.1 & 62.2. https://www.ashrae.org/technical-resources/bookstore/standards-62-1-62-2.
49. How Indoor Ventilation Systems Can Help Prevent or Permit the Spread of COVID-19 | Johns Hopkins Bloomberg School of Public Health. https://publichealth.jhu.edu/2020/how-indoor-ventilation-systems-can-help-prevent-or-permit-the-spread-of-covid-19.
50. Meng, Q. Y. *et al.* Influence of ambient (outdoor) sources on residential indoor and personal PM2.5 concentrations: Analyses of RIOPA data. *J Expo Sci Environ Epidemiol* **15**, 17–28 (2005).
51. Diapouli, E., Chaloulakou, A. & Koutrakis, P. Estimating the concentration of indoor particles of outdoor origin: A review. *Journal of the Air & Waste Management Association* **63**, 1113–1129 (2013).
52. LaRosa, L. E., Buckley, T. J. & Wallace, L. A. Real-time indoor and outdoor measurements of black carbon in an occupied house: an examination of sources. *J Air Waste Manag Assoc* **52**, 41–49 (2002).
53. Air pollution. https://www.who.int/westernpacific/health-topics/air-pollution.
54. US EPA, O. Particulate Matter (PM2.5) Trends. https://www.epa.gov/air-trends/particulate-matter-pm25-trends (2016).
55. Ambient air pollution attributable DALYs. https://www.who.int/data/gho/data/indicators/indicator-details/GHO/mbd-aap-ambient-air-pollution-attributable-dalys.
56. Data Review: How many people die from air pollution? *Our World in Data* https://ourworldindata.org/data-review-air-pollution-deaths.

57. Unequal exposure and unequal impacts — European Environment Agency. https://www.eea.europa.eu/publications/unequal-exposure-and-unequal-impacts.

58. Matte, T. D. *et al.* Monitoring intraurban spatial patterns of multiple combustion air pollutants in New York City: Design and implementation. *J Expo Sci Environ Epidemiol* **23**, 223–231 (2013).

59. Martenies, S. E. *et al.* A Spatiotemporal Prediction Model for Black Carbon in the Denver Metropolitan Area, 2009-2020. *Environmental science & technology* **55**, 3112–3123 (2021).

60. Hoek, G. *et al.* A review of land-use regression models to assess spatial variation of outdoor air pollution. *Atmospheric Environment* **42**, 7561–7578 (2008).

61. Buonanno, G., Stabile, L. & Morawska, L. Personal exposure to ultrafine particles: the influence of time-activity patterns. *Sci Total Environ* **468–469**, 903–907 (2014).

62. Klepeis, N. E. *et al.* The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. *J Expo Sci Environ Epidemiol* **11**, 231–252 (2001).

63. Harrod, C. S. *et al.* Physical Activity in Pregnancy and Neonatal Body Composition: The Healthy Start Study. *Obstetrics & Gynecology* **124**, 257–264 (2014).

64. Harrod, C. S. *et al.* Exposure to prenatal smoking and early-life body composition: the healthy start study. *Obesity (Silver Spring)* **23**, 234–241 (2015).

65. US EPA, O. Building Assessment Survey and Evaluation Study. https://www.epa.gov/indoor-air-quality-iaq/building-assessment-survey-and-evaluation-study (2014).

66. Harris, P. A. *et al.* The REDCap consortium: Building an international community of software platform partners. *Journal of Biomedical Informatics* **95**, 103208 (2019).

67. Harris, P. A. *et al.* Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics* **42**, 377–381 (2009).
68. L'Orange, C., Neymark, G., Carter, E. & Volckens, J. A High-throughput, Robotic System for Analysis of Aerosol Sampling Filters. *Aerosol Air Qual. Res.* **21**, 210037 (2021).
69. Ahmed, T. *et al.* Measurement of black carbon (BC) by an optical method and a thermal-optical method: Intercomparison for four sites. *Atmospheric Environment* **43**, 6305–6311 (2009).
70. Presler-Jur, P., Doraiswamy, P., Hammond, O. & Rice, J. An Evaluation of Mass Absorption Cross-Section for Optical Carbon Analysis on Teflon Filter Media. *Journal of the Air & Waste Management Association (1995)* **67**, (2017).
71. Regularization Tutorial: Ridge, Lasso & Elastic Net Regression. https://www.datacamp.com/tutorial/tutorial-ridge-lasso-elastic-net.
72. StatQuest with Josh Starmer. *Regularization Part 1: Ridge (L2) Regression*. (2018).
73. Hastie, T., Qian, J. & Tay, K. An Introduction to glmnet. 38.
74. Isiugo, K. *et al.* Predicting indoor concentrations of black carbon in residential environments. *Atmospheric Environment* **201**, 223–230 (2019).
75. Cross-Validation Essentials in R - Articles - STHDA. http://www.sthda.com/english/articles/38-regression-model-validation/157-cross-validation-essentials-in-r/.
76. tidyverse package - RDocumentation. https://www.rdocumentation.org/packages/tidyverse/versions/1.3.1.

77. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Soft.* **33**, (2010).

78. Baxter, L. K., Clougherty, J. E., Paciorek, C. J., Wright, R. J. & Levy, J. I. Predicting residential indoor concentrations of nitrogen dioxide, fine particulate matter, and elemental carbon using questionnaire and geographic information system based data. *Atmos Environ* **41**, 6561–6571 (2007).

79. Kucbel, M. *et al.* Temporal and seasonal variations of black carbon in a highly polluted European city: Apportionment of potential sources and the effect of meteorological conditions. *J Environ Manage* **203**, 1178–1189 (2017).

80. Chan, W., Joh, I. & Sherman, M. *Analysis of Air Leakage Measurements from Residential Diagnostics Database*. LBNL-5967E, 1163524 http://www.osti.gov/servlets/purl/1163524/ (2012) doi:10.2172/1163524.

81. van de Schoot, R. *et al.* Bayesian statistics and modelling. *Nat Rev Methods Primers* **1**, 1–26 (2021).