Cardiovascular Effects of Air Pollution: Understanding Biological Risk, Progression of Heart Failure, and Exposure Measurement Error

Claire L Leiser

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Joel D. Kaufman, Chair

Alison Fohner

Adam Szpiro

Program Authorized to Offer Degree:

Department of Epidemiology

University of Washington

Abstract

Cardiovascular Effects of Air Pollution: Understanding Biological Risk, Progression of Heart Failure, and Exposure Measurement Error

Claire L Leiser

Chair of the Supervisory Committee:

Joel D Kaufman

Department of Environmental and Occupational Health Sciences, Department of Epidemiology, Department of Medicine

Research suggests causal associations between air pollution exposure and cardiovascular disease risk, however biologic pathways between these associations are largely unknown. Furthermore, it is possible that exposure measurement error results in biased estimates of health effects from analytic models, thus further evidence for a true causal effect is important to inform public health decision-making. We investigated associations of air pollution exposure between biologic pathways, progression of subclinical cardiovascular disease, and exposure measurement error in many large, epidemiological cohorts representing varying geographical regions.

Our first analysis aimed to estimate the association between outdoor air pollution concentrations and the prevalence of a cardiovascular disease associated acquired somatic mutation (clonal hematopoiesis of indeterminate potential, or CHIP) in two cohorts, the Multi-Ethnic Study of Atherosclerosis and the Women's Health Initiative. We did not detect an association between air pollutants and CHIP. We then performed an analysis of the association of outdoor air pollution concentrations on longitudinal progression of subclinical echocardiographic findings suggesting

pre-heart failure or heart failure in the Hispanic Latino Community Health Study/ Study of Latinos. We identified an association between air pollutants and subclinical changes consistent with development of early heart failure with reduced ejection fraction. Finally, we estimated differences in cardiovascular disease risk from varying geographic distances from the same predictive model in a pooled analysis of four harmonized epidemiological cohorts. We did not identify a strong impact of decreasing the resolution of the predictive model on inferences regarding air pollution – outcome relationships. Across our three studies, our results provide new insights into our understanding of the relationship between air pollution and cardiovascular disease. While we did not find that acquired somatic mutations related to CHIP were likely mediators of this relationship, nor that exposure misestimation due to reduced spatial precision of pollutant concentration predictions results in bias, we did confirm the relationship between air pollutants and cardiovascular mortality and the evolution of heart failure with reduced ejection fraction. Further research is needed to elucidate specific biological pathways of these exposure outcome relationships.

Dedication

For my parents, Pastor Steven John Leiser and Dr. Jennifer Paul Leiser.
You are my north stars.

Table of Contents

List of Acronyms

CVD – Cardiovascular Disease

NAAQS – National Ambient Air Quality Standards

PM – Particulate Matter

PM2.5 – Fine particulate matter

NO2 – nitrogen dioxide

O3 – ozone

CHIP – Clonal hematopoiesis of indeterminate potential

MESA - Multi-Ethnic Study of Atherosclerosis

WHI - Women's Health Initiative

TOPMed - Trans-Omics for Precision Medicine

NIH - National Institutes of Health

NHLBI - National Heart, Lung, and Blood Institute

IQR – Interquartile Range

OR – Odds Ratio

VAF – Variant Allele Fraction

HPpEF - heart failure with preserved ejection fraction

HPrEF - heart failure with reduced ejection fraction

LV – left ventricular

ECHO-SOL - Echocardiographic Study of Latinos

HCHS/SOL - Hispanic Community Health Study/Study of Latinos

V1 – Visit 1

V2 – Visit 2

ICC – Intraclass correlation
LVMI – left ventricular mass index

RWT – relative wall thickness

LVEF – LV ejection fraction

EDV – end-diastolic volume

ESV – end-systolic volume

GLS – LV global longitudinal strain

LAVI – Left atrial volume index

NSES – neighborhood socioeconomic status

BMI – body mass index

CHS - Cardiovascular Health Study

FN - MESA Family New Recruits

REGARDS - Reasons for Geographic and Racial Differences in Stroke

Dbp – diastolic blood pressure

Sbp – systolic blood pressure

m - meter

List of Figures

List of Tables

Chapter 1: Introduction

Air Pollution and Cardiovascular Disease

Air pollution is a known risk factor for the development and exacerbation of cardiovascular disease (CVD). It is estimated that air pollution (both indoor and outdoor) is the most important environmental risk factor for human health and represents a larger burden of disease globally than soil contamination, water pollution, and environmental occupational exposures combined[1]. Air pollution is a complex mixture of particulates and gases from both anthropogenic and natural sources. Most studies of air pollution in the US on CVD focus on pollutants regulated under the National Ambient Air Quality Standards (NAAQS)[2]. These include particulate matter (PM) (including fine PM [$PM_{2.5}$]), nitrogen dioxide (NO_2), ozone (O_3), carbon monoxide, lead, and sulfur dioxide. As environmental health was emerging as a field, time series studies demonstrated the effects of short-term increases in air pollution exposure on acute CVD events and other health effects[3–6]. It is now estimated that more than half of all deaths related to air pollution are due to exposure of $PM_{2.5}$ on CVD mortality, and that CVD health effects vary according to the source and pollutant composition of $PM_{2.5}$ [1,7]. Further efforts have shown that long term exposure to air pollution is associated with incident CVD and progression of disease, between and within cities[8,9].

Biological Mechanisms to Air Pollution CVD Relationship

Despite the larger body of literature demonstrating associations between air pollution and CVD, pathophysiologic and biologic mechanisms underlying the air pollution CVD relationship remain unknown. It is therefore crucial to investigate novel possible mechanisms. Toxicological evidence has thus far provided evidence not only between air pollution and CVD, but also with

precursors to CVD including vascular endothelial dysfunction, hypercholesteremia, hyper glycaemia, arrythmias, atherosclerosis, hypercoagulability, and systemic inflammation[10–17]. There is growing scientific consensus that inhaled PM and other gaseous pollutants cause oxidative stress leading to respiratory and CVD responses in exposed populations over time[10]. Air pollution is associated with increases in serum C-reactive protein, fibrinogen, and hepatocyte growth factor, all biomarkers of systemic inflammation and tissue repair[18] . However, oxidative stress and inflammation are fundamental in most chronic diseases and thus it is crucial to identify air pollution and CVD specific pathways.

Subclinical Disease

Beyond biological pathways, the longitudinal study of subclinical pathophysiological measures could lead to robust insight into the role of risk factors in development of clinical CVD**.** In addition to investigating possible biological mechanisms, insight into air pollution and CVD etiology can be achieved through the study of subclinical measures of CVD. Air pollution has previously been shown to be associated with the presence and progression of coronary artery calcium and carotid intima-media thickness in adult populations, both preclinical measures of atherosclerosis[19–21]. Use of subclinical measurement can provide for further insight into the relationship between air pollution and CVD, particularly in vulnerable populations.

Exposure Measurement Error

Another component that plays a large role in the study of air pollution on CVD is exposure measurement error. In addition to spatial variability, air pollutant concentrations are highly temporally variable, thus making exposure estimation for epidemiological study complex.

For example, traffic related air pollutants peak during rush hours even within short distances to roadways and particulates like dust and wood smoke are variable seasonally, though may be less spatially variable[9,22] . The relationship between air pollutants and development and exacerbation of CVD has been consistently observed over a range of spatially and temporally heterogeneous air pollution concentrations, including those ranges below regulatory standards, and no "safe" exposure concentration has yet to be determined[2,22]. Due to small effect sizes in epidemiologic studies of air pollution, fine scale exposure assessment is needed for both long-term and short-term exposures. Modern approaches to exposure estimation rely on statistical models to predict concentrations over much larger geographic areas outside of monitored areas across the US and over many decades[23–32]. These models have high predictive performance, however questions remain as to the necessity of such precision. It is possible that a point based residential history in the predictive model does not affect the health effects estimates.

Summary of Chapters

The following analyses add to literature on the long-term effects of air pollution on CVD. In addition, we aimed to elucidate the differences in health effects between fine scale exposure assessment compared to more coarse measurement of air pollution.

In Chapter 2, the objective was to explore a novel biological mechanism for the development of CVD. In Chapter 3, we aimed to estimate progression of subclinical heart failure due to air pollution. In Chapter 4, we estimated differences health effects from varying spatial distances to estimate exposure measurement error. Chapter 5 summarizes our findings in the context of known literature.

Chapter 2: Associations between ambient air pollutants and clonal hematopoiesis of indeterminate potential (CHIP)

Introduction

Air pollution is well documented as a risk factor for cardiovascular disease (CVD), however the underlying biological mechanisms for this relationship are not well understood. Toxicological evidence has thus far provided evidence not only between air pollution and CVD, but also with precursors to CVD including vascular endothelial dysfunction, hypercholesteremia, hyper glycaemia, arrythmias, atherosclerosis, hypercoagulability, and systemic inflammation [10–15,17,33]. Air pollution is also associated with increases in serum C-reactive protein, fibrinogen, and hepatocyte growth factor, all biomarkers of systemic inflammation and tissue repair[18] . However, systemic inflammation and oxidative stress are central to most chronic diseases and it is therefore crucial to investigate novel associations that are specific to the air pollution/CVD pathway.

One such pathway could involve clonal hematopoiesis of indeterminate potential (CHIP). Hematopoiesis refers to polyclonal processes giving rise to erythroid, lymphoid, myeloid, or megakaryocytic cells[34]. Somatic mutations in stem cells during hematopoiesis may occur in genes that confer a selective advantage to non-mutated cells, resulting in a greater proportion of circulating clone populations in peripheral blood samples [35–37]. CHIP is defined as a mutation in a driver gene at a variant allele frequency (VAF) as >2% (which reflects the technical limits of sequencing), and an absence of both morphological variation in blood cells and hematologic malignancy[34]. CHIP is a common somatic mutation; approximately 10% of individuals over age 70 have detectable CHIP[35]. CHIP mutations are most commonly associated with loss-of-function mutations in *DNMT3A*, *TET2*, *ASXL1,* and *JAK2* with approximately 90% of CHIP carriers having a single mutation[35,37].

While CHIP is long-known risk factor for subsequent hematologic malignancies (HR ~12), studies have since demonstrated that CHIP mutations lead to a pro-inflammatory state resulting in greater atherosclerotic plaque burden and accelerated atherosclerosis[34,38–40]. CHIP thus confers increased risk in all-cause mortality and incident CVD, beyond traditional lifestyle risk factors [34,36,38–41]. Studies have demonstrated that CHIP carriers have an increased risk of incident coronary artery disease, ischemic stroke, and myocardial infarction, as well as worse prognosis for heart failure[36,40,42].

Age is the most predictive risk factor for the development of CHIP mutations[40] , however other risk factors have recently been described to induce CHIP including germline variation, lifestyle behaviors (e.g. tobacco smoking, obesity), disease comorbidities (e.g. inflammatory conditions, premature menopause, HIV), and the use of chemotherapeutic medication[34,37]. Environmental exposures, including air pollution, are largely unknown and understudied. Given previous evidence that exposure to air pollution is associated with systemic inflammation and that inflammation is a risk factor for CHIP [43,44], it is possible that ambient air pollution is a risk factor for CHIP acquisition. we explored exposure to ambient air pollution as a potential risk factor for CHIP. We estimated associations between long-term air pollution exposure and CHIP prevalence in two epidemiological cohorts, the Multi-Ethnic Study of Atherosclerosis (MESA) and the Women's Health Initiative (WHI).

Methods

Data for this study comes from the National Institutes of Health (NIH) National Heart, Lung, and Blood Institute's (NHLBI) Trans-Omics for Precision Medicine (TOPMed) consortium MESA and WHI participants selected to the US National Heart Lung and Blood Institute's Trans-Omics for Precision Medicine (TOPMed) program represent our sample[45,46].

The aim of TOPMed is to contribute to scientific understanding of biological processes underlying disease through the combination of existing parent studies containing large numbers of participants with well-characterized phenotypic and environmental data. TOPMed encompasses integrated cohort-dependent whole-genome sequencing (WGS) data, as well as data on RNA and protein measurements, metabolic profiles, and epigenetic data. The TOPMed database currently consists of over 155,000 participants from over 80 studies, including the Multi-Ethnic Study of Atherosclerosis (MESA) and the Women's Health Initiative (WHI) (previously described)[47,48].

MESA is a longitudinal cohort study of subclinical cardiovascular disease in 6,814 adults aged 45-84 years at baseline. Subjects were recruited from six US cities (Los Angeles, California; Winston-Salem, North Carolina; New York City, New York; St. Paul, Minnesota; Baltimore, Maryland; and Chicago, Illinois) and were required to be free of clinical cardiovascular disease at enrollment. MESA subjects are self-reported 38% White, 28% African American, 22% Hispanic, and 12% Asian (predominantly of Chinese descent)[47]. The WHI study includes 161,808 postmenopausal women aged 50-79 years who were enrolled at 40 clinical centers across the US between 1993-1998. WHI participants are self-reported 82% White, 9% Black or African-American, 4% Hispanic, 3% Asian or Pacific Islander, 0.4% American Indian or Alaskan Native, and 1% Other race/ethnicity [48]. Participants in WHI were all dbGaP-eligible stroke or venous thromboembolism cases 1:1 matched to an age-, race-, and hormone therapy-stratified, random sample of controls.

CHIP mutations were identified in TOPMed. Information on CHIP mutation identification is available[37]. We applied the GATK MuTECT2 somatic variant caller to whole genome sequencing of blood-derived DNA and identified CHIP based on a list of pre-specified

leukemogenic driver mutations[37]. CHIP was defined as variant allele fraction (VAF) > 2%. Individuals with history of hematologic malignancy or prior use of antineoplastic medication were excluded. We were only able to assess CHIP at one time point (i.e. time of sequencing).

Universal kriging models predicted individual-level concentrations for fine particulate matter ($PM_{2.5}$) and nitrogen dioxide (NO_2) weighted by time in residential address[32,49] Primary exposures were average annual $PM_{2.5}$ and NO_2 concentrations one year prior to blood sampling (MESA 2000-2006 , WHI 1994-2003). Our sample included N_{MESA} = 4445 and N_{WHI} = 7701. Cross-sectional associations between air pollutant concentrations and CHIP prevalence at blood draw were estimated by logistic regression, separately for MESA and WHI and per interquartile range (IQR). We selected potential confounders *a priori*. Our modeling approach was as follows:

MESA model 1a: $PM_{2.5}$, age, gender, smoking status, race/ethnicity, income, year of blood draw
MESA model 1b: Model 1 + contextual covariates (census tract median household income + MESA site)
MESA model 2a: NO_2, age, gender, smoking status, race/ethnicity, income, year of blood draw
MESA model 2b: Model 2 + contextual covariates (census tract median household income + MESA site)

WHI model 1: $PM_{2.5}$, age, smoking status, race/ethnicity, income, year of blood draw
WHI model 2: NO_2, age, smoking status, race/ethnicity, income, year of blood draw

Subanalysis: We restrict CHIP mutations to those with DNMT3A mutations and VAF > 10%

Our specific analytic strategy was as follows:

$$log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots \beta_n x_n$$

Where
p *= probability of CHIP mutation*
β_1 *= estimated difference in log odds CHIP for one unit air pollution exposure*
β_0 *= intercept term*
β_2: β_n *= regression coefficients for predictor variables*

Sensitivity analysis

In both cohorts, we tested for associations between air pollution and CHIP, *DNMT3A*-specific mutations, and VAF>10%. In the WHI cohort, we used multiple imputation using chained equations to impute missing covariate information based on non-missing values among participants. We used ten imputation datasets in our analyses account for differences between imputations. To correct for the complex sampling design for inclusion of participants into TOPMed, we used inverse probability of the sampling weights to correct our analyses. Specifically, individuals selected into TOPMed from WHI were based on cases (stroke or venous thromboembolism cases) and controls. Controls were DBGaP eligible, with DNA already extracted and needed to have at least 5 biomarkers (LDL, HDL, Triglycerides, Glucose, Insulin, CRP and/or Creatinine). Controls were matched based on age at baseline and HRT participation. These results can be inferred to the larger WHI population from which the sample was drawn. Because MESA participants were not selected based on clinical diagnosis of stroke and were free of CVD at the time of sampling, we did not perform multiple imputation on that cohort.

Results

Table 1 shows descriptive statistics. CHIP was identified in 196(4.4%) MESA participants and 668 (8.7%) WHI participants. Most commonly mutated genes were *DNMT3A, TET2,* and *ASXL1*. Participants with CHIP were on average older than participants without CHIP (66.9 years vs 61.0 years in MESA; 71.6 years vs 69.2 years in WHI). Participants with CHIP were more likely to have history of smoking. Air pollution exposure was similar between participants with and without CHIP ($PM_{2.5-CHIP}$ = 16.9 μg/m^3 vs $PM_{2.5-NO\ CHIP}$ = 16.8 μg/m^3; NO_{2-CHIP} = 22.6 ppb vs $NO_{2-NO\ CHIP}$ = 22.0 ppb in MESA) ($PM_{2.5-CHIP}$ = 13.4 μg/m^3 vs $PM_{2.5-}$

$_{NO\ CHIP}$ = 13.5 μg/m^3; $NO_{2\ -\ CHIP}$ = 14.1 ppb vs $NO_{2\ -\ NO\ CHIP}$ = 14.1 ppb in WHI) (Figure 1 and Figure 2).

Table 2 shows the characteristics of the CHIP variants for both cohorts. For both MESA and WHI, a majority of individuals had only one variant. Most mutations were on chromosome 2 and most mutations were observed in *DNMT3A*, followed by *TET2* and *ASXL1*. The most common mutation was a non-synonymous single nucleotide variant.

Table 3 shows the results of logistic regression for air pollutants on CHIP mutation prevalence in MESA and WHI. IQR increases in $PM_{2.5}$ were not associated with odds of CHIP in MESA in the main model (OR_{MESA1a} = 1.06; 95% CI 0.91 - 1.23) or models adjusted for contextual covariates (OR_{MESA1b} = 1.03; 95% CI 0.70- 1.50). Table 4 shows the results off logistic regression for NO_2 on CHIP mutation prevalence in MESA. IQR increases in NO_2 were not associated with odds of CHIP in MESA in the main model (OR_{MESA2a} = 1.12; 95% CI 0.87 - 1.59) or models adjusted for contextual covariates (OR_{MESA2b} = 1.02; 95% CI 0.48- 2.15). Table 5 shows the effect of air pollutants on CHIP subgroup in MESA. IQR increases in $PM_{2.5}$ were not associated with odds of *DMNT3A*-specific CHIP in MESA in the main model (OR_{MESA1a} = 1.13; 95% CI 0.94 - 1.37) or models adjusted for contextual covariates (OR_{MESA1b} = 1.08; 95% CI 0.69 - 1.69). IQR increases in NO_2 were not associated with odds of *DMNT3A*-specific CHIP in MESA in the main model (OR_{MESA2a} = 1.27; 95% CI 0.88- 1.82) or models adjusted for contextual covariates (OR_{MESA2b} = 1.27; 95% CI 0.52 – 3.10). IQR increases in $PM_{2.5}$ were not associated with odds of CHIP with VAF > 10% in MESA in the main model (OR_{MESA1a} = 1.02; 95% CI 0.87 - 1.20) or models adjusted for contextual covariates (OR_{MESA1b} = 0.95; 95% CI 0.63 - 1.43). IQR increases in NO_2 were not associated with odds of CHIP with VAF > 10% in MESA in the main model (OR_{MESA2a} = 1.18; 95% CI 0.86 - 1.63) or models adjusted for contextual

covariates (OR_{MESA2b} = 1.06; 95% CI 0.49 – 2.32). IQR increases in $PM_{2.5}$ were not associated with odds of CHIP in WHI (OR_{WHI1} = 0.97; 95% CI 0.86 - 1.03). IQR increases in NO_2were not associated with odds of CHIP in WHI (OR_{WHI2} = 0.98; 95% CI 0.87 - 1.10). Table 7 shows the effect of air pollutants on CHIP subgroup in WHI. IQR increases in $PM_{2.5}$ were not associated with odds of *DMNT3A*-specific CHIP in WHI (OR_{WHI1} = 1.08; 95% CI 0.94 - 1.25). IQR increases in NO_2were not associated with odds of *DMNT3A*-specific CHIP in WHI (OR_{WHI2} = 1.11; 95% CI 0.97 - 1.28). IQR increases in $PM_{2.5}$ were not associated with odds of CHIP with VAF > 10% in WHI (OR_{WHI1} = 0.99; 95% CI 0.87 - 1.11). IQR increases in NO_2were not associated with odds of CHIP with VAF > 10% in WHI (OR_{WHI2} = 1.01; 95% CI 0.90 - 1.14).

Table 8 shows the effect of pollutants on CHIP mutation prevalence among non-smokers or never smokers. Again, we did not observe a significant effect of air pollution on CHIP mutation in MESA for $PM_{2.5}$ ($OR_{MESAPM2.5}$= 1.01; 95% CI 0.68 – 1.50) or NO_2 ($OR_{MESANO2}$= 1.08; 95% CI 0.49 – 2.35). We also did not observe a significant effect of air pollution on CHIP mutation in WHI for $PM_{2.5}$ ($OR_{WHIPM2.5}$= 0.96; 95% CI 0.86 – 1.09) or NO_2 (OR_{WHINO2}= 0.98; 95% CI 0.87– 1.10).

Table 9 shows the results if the sensitivity analysis in WHI. IQR increases in $PM_{2.5}$ were not associated with odds of CHIP ($OR_{PM2.5}$= 0.98; 95% CI 0.81 – 1.19) or NO_2 (OR_{NO2}= 01.03; 95% CI 0.86 – 1.23). Similarly, IQR increases in $PM_{2.5}$ were not associated with odds of *DNMT3A*- specific mutations ($OR_{PM2.5}$= 1.05; 95% CI 0.82 – 1.35) or NO_2 (OR_{NO2}= 01.08; 95% CI 0.87 – 1.34). When we expanded our CHIP VAF to > 10%, we again observed no associations for $PM_{2.5}$ ($OR_{PM2.5}$= 1.00; 95% CI 0.81 – 1.22) or NO_2 (OR_{NO2}= 01.04; 95% CI 0.86 – 1.25).

Discussion

To our knowledge, this is the first study to estimate associations between outdoor, traffic-related air pollution and CHIP; we did not find associations. Very few studies have examined any environmental contributions to CHIP. A recent study focused on indoor radon exposure at the county level and odds of CHIP in the TOPMed WHI cohort found that among participants with ischemic stroke, living in counties with high radon exposure was associated with a 39% increase in odds of CHIP[50]. However, the study was ecological in nature and differences exist between the physiological effects of air pollution and radon exposures and therefore the results of this study are not informative to the interpretation of our null findings. Another study of environmental exposures and clonal hematopoiesis among World Trade Center first responders showed a higher proportion of clonal hematopoietic cells (as determined by deep targeted sequencing) when compared to non-World Trade Center first responders (OR=3.14; 95% CI:1.64–6.03).[51] Similarly to our study, the frequency of somatic mutations was highest in *DNMT3A* and *TET2.* It is again difficult to compare our results to the findings of this study as the composition of the air pollution of the World Trade Center disaster site is different than the composition of traffic-related air pollution as we are interested in here. In addition, we were interested in long-term exposure to air pollution and the aforementioned study focused on lasting effects of an acute event. However, given that CHIP is a risk factor for future malignancy and CVD and biologic plausibility for other chronic diseases, future studies examining other environmental contributions to CHIP should be conducted.

Strengths of this study include our air pollution assessment, geographic variation of the population, and extensive covariate information.The major limitation of our study is its cross-sectional nature, which does not allow for causal inference between our exposure and outcome;

we do not have evidence of a temporal relationship between air pollution and CHIP, nor are we able to assess risk. Future analyses with multiple time points for exposure and outcome assessment could be more informative for this research question. Selection bias is probable due to selection of participants into TOPMed, however WHI results were robust to weighting for inverse probability of sampling.

Consistent with other studies of CHIP, age was the strongest predictor of CHIP and individuals with CHIP were more likely to have a history of smoking[36,43]. We did not observe differences in our results from individuals with *DNMT3A* mutations and lacked power to test other common mutation sites including *TET2* and *ASXL1*. Future studies examining this question could look at mutation site specific CHIP, power permitting.

Conclusion

This is the first study to examine air pollution exposure and CHIP mutations. No association was found between air pollution and CHIP prevalence in this study.

Table 1: *Descriptive Statistics Participants at Baseline Examination*							
Multi-Ethnic Study of Atherosclerosis				**Women's Health Initiative**			
Characteristic	**CHIP Mutation**	**No CHIP Mutation**	**Total**	**Characteristic**	**CHIP Mutation**	**No CHIP Mutation**	**Total**
No. participants	196 (4.4)	4249 (95.6)	4445 (100)	No. participants	668 (8.7)	7033 (91.3)	7701 (100)
Age (years)	66.9 (45-84)	60.9 (44-84)	61.2 (44-84)	Age (years)	71.6 (52-86)	69.2 (50-87)	69.4 (50-87)
Sex				Sex			
Male	90 (45.9)	2071 (48.7)	2161 (48.6)	Male	0 (0.0)	0 (0.0)	0 (0.0)
Female	106 (54.1)	2178 (51.3)	2284 (51.4)	Female	668 (8.7)	7033 (91.3)	7701 (100)
Smoking Status				Smoking Status			
Never	90 (45.9)	2156 (50.7)	2246 (50.5)	Never	320 (47.9)	3577 (50.9)	3897 (50.6)
Former	86 (43.9)	1561 (36.7)	1647 (37.1)	Former	300 (44.9)	2880 (41.0)	3180 (41.3)
Current	19 (9.7)	524 (12.3)	543 (12.2)	Current	44 (6.6)	488 (6.9)	532 (6.9)
Missing	1 (0.5)	8 (0.2)	9 (0.2)	Missing	4 (0.6)	88 (1.3)	92 (1.2)
Race/Ethnicity				Race/Ethnicity			
White	88 (44.9)	1708 (40.2)	1796 (40.4)	White, Non-Hispanic	573 (85.8)	5774 (82.1)	6347 (82.4)
Chinese American	18 (9.2)	545 (12.8)	563 (12.7)	Asian or Pacific Islander	10 (1.5)	96 (2.0)	106 (1.4)
Black/African American	51 (26.0)	1071 (25.2)	1122 (25.2)	Black/African American	65 (9.7)	875 (12.4)	940 (12.2)
Hispanic	39 (19.9)	925 (21.8)	964 (21.7)	Hispanic/Latino	11 (1.7)	217 (3.1)	228 (3.0)
AI/AN	0 (0.0)	0 (0.0)	0 (0.0)	AI/AN	3 (0.5)	31 (0.4)	34 (0.4)
Other	0 (0.0)	0 (0.0)	0 (0.0)	Other	6 (0.9)	40 (0.6)	46 (0.6)
Total Gross Family Income				Total Gross Family Income			
<$19999	49 (25.0)	835 (19.7)	884 (19.9)	<$20000	119 (17.8)	1367 (19.4)	1486 (19.3)
$20000-39999	51 (26.0)	1080 (25.4)	1131 (25.4)	$20000-49999	336 (50.3)	3245 (46.1)	3581 (46.5)
$40000-99999	63 (32.1)	1590 (37.4)	1653 (37.2)	$50000-99999	139 (20.8)	1602 (22.8)	1741 (22.6)
$100000+	24 (12.2)	632 (14.9)	656 (14.8)	$100000+	31 (4.6)	382 (5.4)	413 (5.4)
Missing	9 (4.6)	112 (2.6)	121 (2.7)	Missing	43 (6.4)	437 (5.4)	480 (6.2)
Education				Education			
Less than High School	28 (14.3)	640 (15.1)	659 (15.1)	Less than High School	26 (3.90)	376 (5.4)	402 (5.2)
High School Degree	70 (35.7)	1433 (33.7)	1481 (33.8)	High School Degree	103 (15.4)	1277 (18.2)	1380 (17.9)
College or Technical Degree	53 (27.0)	1340 (31.5)	1372 (31.3)	Some College	277 (41.5)	2765 (39.3)	3042 (39.5)
Graduate or Professional Degree	44 (22.5)	828 (19.5)	858 (19.6)	College Degree or Higher	259 (38.8)	2568 (36.5)	2827 (36.7)
Missing	1 (0.5)	8 (0.2)	9 (0.2)	Missing	3 (0.5)	47 (0.7)	50 (0.7)
Year of Blood Draw				Year of Blood Draw (masked)			
				1994			
				1995			
				1996			
				1997			
				1998			
				1999			
2000	32 (16.3)	644 (15.2)	676 (15.2)	2000			
2001	91 (46.4)	2251 (53.0)	2342 (52.7)	2001			
2002	71 (36.2)	1295 (30.5)	1366 (30.7)	2002			
2003	0 (0.0)	0 (0.0)	0 (0.0)	2003			
2004	1 (0.5)	23 (0.5)	24 (0.5)				
2005	1 (0.5)	36 (0.9)	37 (0.8)				
Annual average $PM_{2.5}$	16.9 (9.4-30.1)	16.8 (8.5-31.6)	16.8 (8.5-31.6)	Annual average $PM_{2.5}$	13.4 (3.5-22.5)	13.5 (2.5-23.2)	13.5 (2.5-23.2)
Annual average NO_2	22.6 (6.2-51.5)	22.0 (4.0-60.1)	22.02 (4.03-60.07)	Annual average NO_2	14.1 (1.5-41.9)	14.1 (1.3-77.9)	14.1 (1.3-77.9)

* $PM_{2.5}$ = Particulate matter >2.5 μm ($\mu g/m^3$)
* NO_2 = Nitrogen dioxide (ppb)
Rows are N(%) or Mean (range)
AI/AN = American Indian/Alaskan Native

Table 2: *Characteristics of CHIP Variants for MESA and WHI participants*

Characteristic	MESA	WHI
No. individuals	196	668
No. variants	210	758
Variants per individual		
1	182	589
2	14	69
3	0	9
4	0	1
Chromosome		
1	2	7
2	137	414
4	34	161
5	0	1
6	1	5
7	0	2
8	1	0
9	5	27
10	1	0
11	1	4
12	0	11
13	2	1
15	1	2
17	13	51
20	11	62
21	0	4
X	1	6
Gene		
ASXL1	9	54
ASXL2	0	1
BRCC3	0	3
CBL	1	4
DNMT3A	136	399
ETNK1	0	4
EZH2	0	2
GNAS	2	8
GNB1	2	2
IDH2	1	2
JAK2	5	27
KDM6A	0	3
KIT	0	1
KRAS	0	7
MPL	0	1
NF1	1	2
NPM1	0	1
NRAS	0	3
PDS5B	2	1
PHIP	1	5
PPM1D	7	25
RAD21	1	0
SF3B1	1	11
SMC3	1	0
SRSF2	2	11
STAG2	1	0
SUZ12	0	1
TET2	34	160
TP53	3	12
U2AF1	0	4
Annotated Function		
Frameshift deletion	33	157
Frameshift insertion	10	62
Non-frameshift deletion	3	0
Non-synonymous SNV	110	386
Splicing	17	45
Stopgain	37	108
Variant Allele Frequency (VAF)		
Mean (range)	0.18 (0.07, 0.59)	0.19 (0.06, 0.87)

Table 3: The Effect of PM2.5 on CHIP Mutation Prevalence: Results of Logistic Regression – MESA (N=4,445) & WHI (N=7,701)

Model	Parameter Estimate
MESA 1A	1.06 (0.91, 1.23)
MESA 1B	1.03 (0.70, 1.50)
MESA 2A	1.12 (0.87, 1.59)
MESA 2B	1.02 (0.48, 2.15)
WHI 1A	0.97 (0.86,1.08)
WHI 1b	0.98 (0.87, 1.10)

$PM_{2.5}$ exposure per 4 unit increase
NO_2 exposure per 10 unit increase
All parameters refer to the participant's characteristic at time of blood draw

MESA model 1a: $PM_{2.5}$, age, gender, smoking status, race/ethnicity, income, year of blood draw
MESA model 1b: Model 1 + contextual covariates (census tract median household income + MESA site)
MESA model 2a: NO_2, age, gender, smoking status, race/ethnicity, income, year of blood draw
MESA model 2b: Model 2 + contextual covariates (census tract median household income + MESA site)

WHI model 1: $PM_{2.5}$, age, smoking status, race/ethnicity, income, year of blood draw
WHI model 2: NO_2, age, smoking status, race/ethnicity, income, year of blood draw

Table 4: *The Effect of Air Pollutants on CHIP Subgroup: Results of Logistic Regression – MESA (N=4,445)*

	DMNT3A mutation Odds Ratio (95% Confidence Interval)	VAF > 10% Odds Ratio (95% Confidence Interval)
Model 1a	1.13 (0.94, 1.37)	1.02 (0.87, 1.20)
Model 1b	1.08 (0.69, 1.69)	0.95 (0.63, 1.43)
Model 2a	1.27 (0.88, 1.82)	1.18 (0.86, 1.63)
Model 2b	1.27 (0.52, 3.10)	1.06 (0.49, 2.32)

$PM_{2.5}$ exposure per 4 unit increase
NO_2 exposure per 10 unit increase
All parameters refer to the participant's characteristic at time of blood draw
*= $p<0.05$

Model 1a: includes $PM_{2.5}$, age, gender, smoking status, race/ethnicity, income, year of blood draw
Model 1b: Model 1 + contextual covariates (census tract median household income + MESA site)
Model 2a: includes NO_2, age, gender, smoking status, race/ethnicity, income, year of blood draw
Model 2b: Model 2 + contextual covariates (census tract median household income + MESA site)

Table 5: *The Effect of Air Pollutants on CHIP Subgroup: Results of Logistic Regression – WHI (N=7,701)*

	DMNT3A mutation Odds Ratio (95% Confidence Interval)	VAF > 10% Odds Ratio (95% Confidence Interval)
Model 1	1.08 (0.94, 1.25)	0.99 (0.87, 1.11)
Model 2	1.11 (0.97, 1.28)	1.01 (0.90, 1.14)

$PM_{2.5}$ exposure per 4 unit increase
NO_2 exposure per 10 unit increase
All parameters refer to the participant's characteristic at time of blood draw
*= $p<0.05$

Model 1: includes $PM_{2.5}$, age, smoking status, race/ethnicity, income, year of blood draw
Model 2: includes NO_2, age, smoking status, race/ethnicity, income, year of blood draw

Table 6: *Effect of Pollutants on CHIP Mutation Among Non-Smokers or Never Smokers*

	Multi-Ethnic Study of Atherosclerosis (MESA) (N=3,893)	**Women's Health Initiative (WHI) (N=7,077)**
Pollutant	CHIP Odds Ratio (95% Confidence Interval)	
$PM_{2.5}$	1.01 (0.68, 1.50)	0.96 (0.86, 1.09)
NO_2	1.08 (0.49, 2.35)	0.98 (0.87, 1.10)

$PM_{2.5}$ exposure per 4 $\mu g/m^3$ increase
NO_2 exposure per 10 ppb increase
MESA model includes age, personal income, sex, education, race/ethnicity, MESA site, year of blood draw, neighborhood income
WHI model includes age, personal income, sex, education, race/ethnicity, year of blood draw

Table 7: *Effect of Pollutants on CHIP Mutation in the Women's Health Initiative (N=7,701) – Sensitivity Correction for Sampling Design*

Parameter	CHIP	DNMT3A-specific	VAF >10%
Average $PM_{2.5}$ ($\mu m/m^3$)	0.98 (0.81, 1.19)	1.05 (0.82, 1.35)	1.00 (0.81, 1.22)
Average NO_2 (ppb)	1.03 (0.86, 1.23)	1.08 (0.87, 1.34)	1.04 (0.86, 1.25)

$PM_{2.5}$ exposure per 4 unit increase
NO_2 exposure per 10 unit increase
Model includes pollutant, age, smoking status, self-reported race/ethnicity, income, year of blood draw, education

Figure 1: $PM_{2.5}$ Annual Average by Cohort

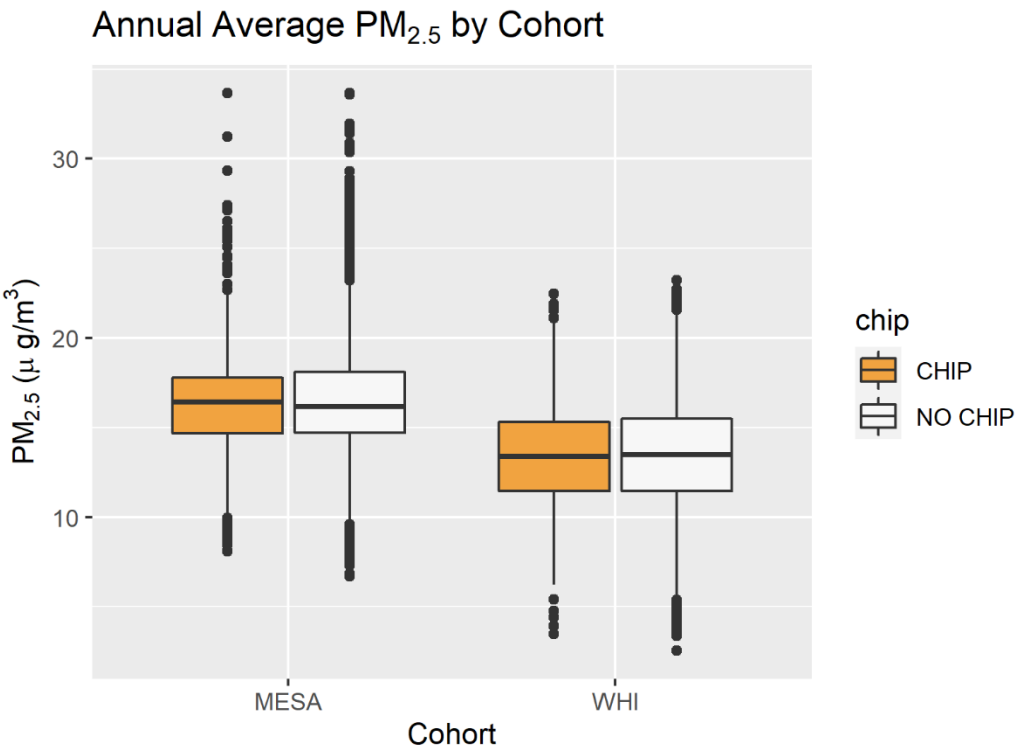
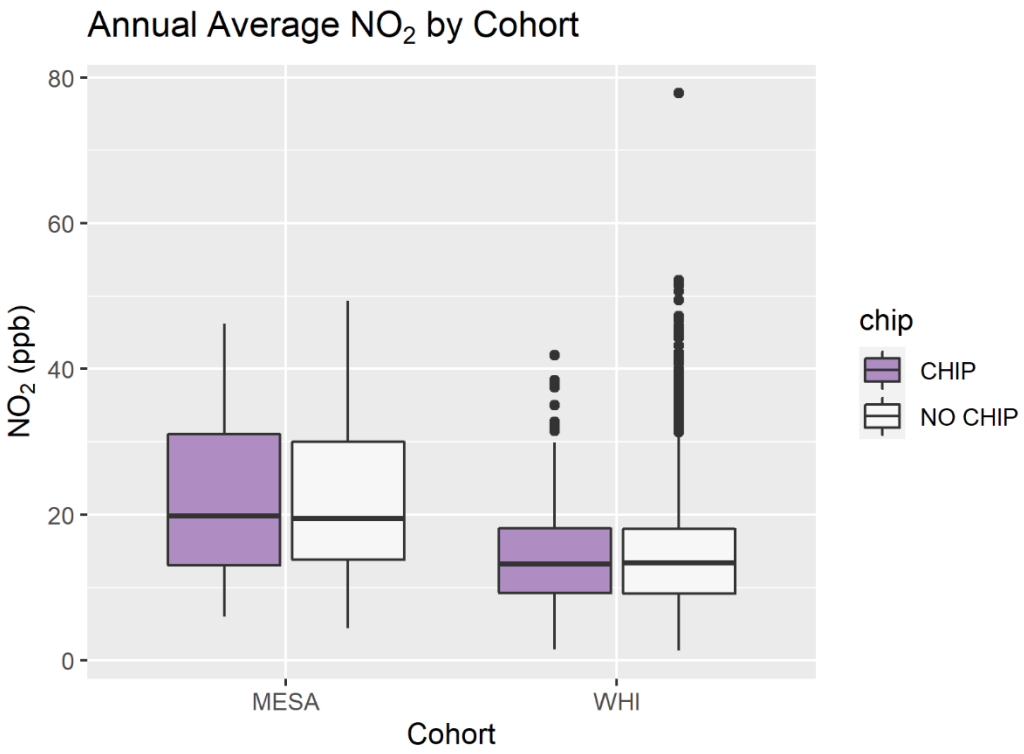


Figure 2: NO_2 Annual Average by Cohort



Chapter 3: Effects of Ambient Air Pollutants on Cardiac Structure and Function in ECHO-SOL

Introduction

Heart failure is a condition of substantial interest in the US, due to its high morbidity and mortality [52]. Heart failure is a growing health issue, with an estimated prevalence of 6.2 million in 2020[53]and 20% lifetime risk for persons aged over 40 years. US heart failure costs are high, and expected to grow to at least 70 billion per year ($244 per every US adult) with total cost of caring for patients with heart failure reaching $160 billion by 2030[54]. Furthermore, heart failure varies by ethnicity; Hispanic/Latino men and women in the US have higher rates of heart failure compared to non-Hispanic white or Asian populations.[55] Modifiable risk factors for heart failure are therefore essential to identify. Air pollution has previously been shown to exacerbate heart failure, with an increase in hospitalizations and resource utilization[56], and may be a critical risk factor for the development of heart failure.

While traditional risk factors for clinically diagnosed heart failure are well documented[55], effects of air pollution and other environmental risk factors on the progression of subclinical cardiac structure and function preceding clinical diagnosis of heart failure have been less explored. The use of subclinical measures of vascular disease can be particularly important to understand the etiology of heart failure; use of subclinical vascular disease has been shown to be an independent predictor of future adverse cardiovascular events[19,57,58], and estimating the rate of progression on asymptomatic disease is critically important. Prevalence of asymptomatic heart failure is thought to be high and variable by phenotype; heart failure with preserved ejection fraction (HFpEF) is estimated to be between 7-14% in the population, while heart failure with reduced ejection fraction (HFrEF) is estimated to be between 20-30%[52]. Estimation of the effect of air pollution exposure on preclinical heart failure and its varying phenotypes can thus provide

information into the underlying physiological processes of air pollution on the heart failure cascade.

The primary objective of this study was to evaluate the association between air pollution exposure and the progression of phenotypes consistent with preclinical heart failure, as measured by echocardiographic assessment of left ventricular (LV) myocardial strain, structure, systolic function, and diastolic function. We evaluated the impacts of outdoor residential concentrations of particulate matter air pollution less than 2.5 micrograms in aerodynamic diameter ($PM_{2.5}$) on echocardiographic measures performed at the first and second visits (V1 and V2) in the Echocardiographic Study of Latinos (ECHO-SOL), an ancillary study to the Hispanic Community Health Study/Study of Latinos (HCHS/SOL).

Methods

Sample Selection

HCHS-SOL and ECHO-SOL have been previously described[59,60]. Briefly, HCHS/SOL is a multi-center epidemiological cohort designed to provide information on the health burden and risk factors of the Hispanic/Latino population in the US. HCHS/SOL consists of roughly 16,000 individuals aged 18-64 enrolled from 2008-2010 from four US cities (San Diego, CA; Bronx, NY; Miami, FL; Chicago, IL)[59]. Participants from the larger HCHS/SOL cohort were recruited across the four clinical sites to ECHO-SOL through a stratified-sampling process to be representative of the HCHS/SOL cohort. ECHO-SOL is designed to add echocardiographic assessment of cardiac structure and function to the baseline clinical examination of HCHS\SOL[60]. Individuals in the present study represent 1,818 ECHO-SOL participants.

Exposure

Exposure predictions were generated from validated national spatio-temporal models of outdoor pollutant concentrations. Concentrations were predicted by a hierarchical model in a universal kriging framework which integrates temporal trends and incorporates agency and researcher-deployed air pollution monitoring data, geographic covariates, and spatial smoothing to characterize pollutant concentrations for each two-week period of follow-up.[49] Estimates were made at the precise residential location, and weighted to reflect time living at each known address. Individuals with no geocode available were excluded. Our final sample included 1630 individuals. The primary exposures in this study were annual average $PM_{2.5}$ prior to V1 (2008-2011) and V2 (2015-2018).

Outcome Assessment

Philips Ultrasound IE-33 or Sonos 5500/7500 ultrasound imaging platforms were used across all field imaging centers. Per the American Society of Echocardiography recommendations, standard echocardiography examinations were performed by experienced sonographers at each Field Imaging Center[60–65]. Speckle-tracking strain analysis was performed offline using TomTec Cardiac Performance Analysis package on 2D images and the speckles tracked along the endocardial and epicardial borders generated peak longitudinal and circumferential strain and strain rate for both regional and global views.[60,66] At least two full cardiac cycles were recorded for each measure and one sonographer was chosen per ECHO-SOL site in order to ensure consistency. All sonographers participated in trainings to review protocols and all echocardiograms were reviewed by a cardiologist and approved before image transfer to the HCHS-SOL data coordinating center for centralized reading. [60] Scans were interpreted at a single reading center at Wake Forest School of Medicine (Winston-Salem, NC) where the analyst was unaware of the exposure information. Inter-and intra-reader reproducibility as measured by intra-

class correlation (ICC) ranged from 0.80 to 0.99. Left atrial volume and left ventricular end-diastolic volumes had the highest ICC values (0.97-.99) The outcome definitions for echocardiographic measures included measurements on both left and right heart structure and function and are defined as follows:

1. Left ventricular (LV) mass index (LVMI). LV mass was determined by subtracting the LV endocardial cavity volume from the volume encompassed by the LV epicardium and multiplying the resultant myocardial volume by the myocardial density. LVMI represents the ratio of LVMI to body surface area, reflected by height^{2}. [67]
2. Relative Wall Thickness (RWT). Defined as 2 times posterior wall thickness divided by LV diastolic diameter[64].
3. LV systolic function and volumes. LV ejection fraction (LVEF) was derived from volumetric assessments using the method of discs from apical 4- and 2-chamber long-axis views to measure end-diastolic volume (EDV) and end-systolic volume (ESV). LV ejection fraction was calculated: (EDV−ESV)/EDV. We also utilized LV stroke volume to define systolic function[64].
4. LV diastolic function. Our algorithm for diastolic dysfunction utilized the following echocardiographic components: medial and lateral tissue Doppler E′ velocities, mitral inflow E/A ratio, E/E′ ratio, isovolumic relaxation time, and left atrial volume index (LAVI)[68,69].
5. LV global longitudinal strain (GLS). Longitudinal shortening as a percentage change in length proportional to baseline length derived from speckle tracking (4-chamber view, 2-chamber view, and average GLS). [70]

Statistical Analysis

A mixed effects model was used to jointly model the cross-sectional and longitudinal relationship between air pollution and myocardial strain as measured by echocardiogram. The residential pollutant exposures during the calendar year prior to the echocardiogram was the primary exposure calculated. The model is given by the following equation for set of individuals ($i = 1, \dots, n$) measured at one or more exams ($v = 1, \dots, V$):

$$Y_{iv} = [\alpha_0 + X_{i0}\alpha_1 + a_i] + [t_{iv}\beta_0 + W_{iv}t_{iv}\beta_1 + t_{iv}b_i] + [U_{iv}\gamma_1 + \epsilon_{iv}]$$

where

Y_{iv} outcome (e.g., LV GLS) measure for individual i at exam v

α_0 average measurements at exam 1 in the reference group

X_{i0}

Vector of baseline covariates (e.g. sex at birth) for participant i

α_1 coefficients for covariates in cross-sectional relationship

a_i subject specific random intercept

t_{iv} time from V1 to V2 for participant i

β_0 change in outcome measures (centered i.e. Y_{iv} - mean)

W_{iv} Vector time-varying exposure of interest (e.g. air pollution) and time-varying covariates (e.g. smoking status) at visit v for participant i

β_1 *term of interest* - coefficient for association between predictors (including air pollution) and rate of change in outcome measurements

b_i random slope per participant

U_{iv} time-varying covariates to adjust measures at exam v for participant i (e.g. air pollution, anti-hypertensive medication use)

γ_1 coefficients for the cross-sectional associations between time-varying variables and measurements at all available exams

ϵ_{iv} error term

The mixed effects model is flexible with respect to the number of observations and follow-up time per participant and allows for adjustment of time-varying covariates as well as covariates associated with baseline measurements and progression from V1 to V2. In addition, the assumption that data are missing completely at random is not required and thus compared to methods that require full follow-up for all participants, selection bias is less concerning. Repeated measurements were modeled as a function of time since baseline, with time-varying exposures modeled as an interaction with follow-up time to obtain a rate. To account for ECHO-SOL clustering and stratification within HCHS/SOL, we allowed for random intercepts for HCHS/SOL center and primary sampling unit (i.e. census tract of recruitment). Model 1 (primary model) was adjusted for time varying covariates, (ECHO-SOL site, alcohol use, and smoking status) and non-time varying covariates from baseline interviews (education, gender, age, pack years of smoking, BMI, and physical activity). We do not include BMI in our LAVI model and LVMI model, as BMI or body surface area are included in the index calculation for both indices. In Model 2, we sought to evaluate the possibility of confounding of our results neighborhood socioeconomic status (NSES). NSES was measured as within year standardized NSES index value, where the z-score is interpreted as the census tract level disadvantage relative to other tracts that year.[71] In order to exclude the possibility of possible mediation by biologic variables,

we also conducted Model 3 to include time-varying (V1 and V2) clinical variables including systolic blood pressure (sbp), diastolic blood pressure (dbp), and diabetes mellitus. We included statin use (Yes or No at time of exam) as a time-varying covariate (V1 and V2).

In order to view differences from our longitudinal model to our cross-sectional model, We report the effect from the cross sectional portion of our model. We estimate the effect per 4μg/m3 increase in PM2.5 and per 10 ppb increase in NO2 concentrations on echocardiographic measures of cardiac structure and function adjusted for covariates

Results

Table 1 shows the descriptive statistics of the sample. At V1, the mean age was 55 years. The majority of the participants were female (65.1%). 58.8% of participants had a high school diploma or GED or less educational attainment, and 19.2% reported a yearly household income of less than $10,000. Most participants reported being a never or former smoker (83.3 %) and most reported current use of alcohol (42.9%). Mean body mass index (BMI) was 30.2. A large majority (90.3%) of participants reported no minutes of recreational physical activity per day. Participants of ECHO-SOL were geographically distributed among HCHS-SOL sites (30.3 % Bronx; 23.3% Chicago; 19.3% San Diego; 27.2% Miami). Average follow-up to visit 2 was 4.3 years.

Figure 2 shows the annual average $PM_{2.5}$ by ECHO-SOL center and visit number. In general, air pollution concentrations decreased between V1 and V2. Air pollution concentrations varied by site; at V1, the mean $PM_{2.5}$ $\mu g/m^3$ was 10.2 in Chicago, 10.6 in Bronx, 9.9 in San Diego, and 7.3 in Miami. At V2, the mean $PM_{2.5}$ $\mu g/m^3$ was 8.1 in Chicago, 9.0 in Bronx, 9.2 in San Diego, and 7.0 in Miami.

Descriptive statistics of outcome variables are shown in Table 2. Across all centers, at V1 average global longitudinal strain (GLS) was -17.5% and -17.0% at V2. Average left ventricular mass index (LVMI) was 82.8 g/m2 at V1 and 85.5 g/m2 at V2. Average left ventricular ejection fraction (LVEF) 59.7% at V1 and 59.0% at V2. Average E/e' ratio was 10.0 at V1 and 10.4 at V2. At V1, the mean E' was 8.1 at and 7.0 at V2. Left atrial volume index (LAVI) was 23.0 at V1 and 23.0 at V2. Relative wall thickness (RWT) was on average 0.4 at V1 and 0.5 at V2.

Table 3 shows the results of the longitudinal analysis. All results are reported by interquartile range (IQR) increases. $PM_{2.5}$ was associated with HPrEF; we observed a 0.52% increase in GLS per year of follow up (95% CI 0.20 – 0.83) and a decrease in LVEF of -1.05% per year (95% CI -1.73 – -0.37). We observed an increase of 1.19 of RWT per year (95% CI 0.47 – 1.91). $PM_{2.5}$ was also associated with decreased LAVI (-1.39 ml/m^2 per year; 95% CI (-2.25– -0.54). We did not observe associations between $PM_{2.5}$ and measures of E', E/E' ratio, or LV mass.

Our results were robust to inclusion of NSES and inclusion of clinical variables . Per IQR increase in $PM_{2.5}$, $PM_{2.5}$ was associated with HPrEF; we observed a 0.52% increase in GLS per year of follow up (95% CI 0.20 – 0.83) and a decrease in LVEF of -1.05% per year (95% CI -1.73 – -0.37). We observed an increase of 1.19 of RWT per year (95% CI 0.47 – 1.91). $PM_{2.5}$ was also associated with decreased LAVI (-1.39 ml/m^2 per year; 95% CI (-2.25– -0.54). We did not observe associations between $PM_{2.5}$ and measures of E', E/E' ratio, or LV mass.

Table 4 shows the results of NO_2 in the analysis. All results are reported by interquartile range (IQR) increases. NO_2 was associated with HPrEF; we observed a 0.12% increase in GLS per year of follow up (95% CI 0.03– 0.20). NO_2 was also associated with decreased LAVI (-0.94 ml/m^2 per year; 95% CI (-1.63– -0.26). Our results for GLS were sensitive to the inclusion of

NSES and clinical variables. We did not observe associations between NO_2 and other outcome measures.

Table 5 shows the results of a cross sectional analysis of PM2.5. For each IQR increase in annual average PM2.5, we observed an association between GLS (0.68 % increase per 4 mg/m3 (95%CI 0.10-1.15) at baseline. We did find associations between PM2.5 and both e' and E/e' ratio; those associations are consistent with improved left ventricular diastolic function, and were not found in our longitudinal analyses. Table 6 shows the cross-sectional results for NO2. We observed cross sectional relationships between NO2 and GLS (0.61, 0.03-1.19).. We did not find statistically significant results for other outcomes for either pollutant.

Tables 7 and 8 show the differences in baseline outcome measurement and demographic statistics by whether or not they had one or two echocardiographic exams (i.e., Visit 1 only or Visit 1 and 2). Only 9.94% of the sample did not participate in the second visit. We observed no major differences in outcome measures for participants with only one visit compared to participants with two visits (Table 7). However in Table 8, we observed a larger amount of attrition from the Bronx site compared to the other sites in our study (V1=39.43%; V1V2 = 29.28%). In addition, we observed a smaller amount of current smokers in V2 compared to the percentage in V1 (V1=24.57%; V1V2=13.82%). We observed a smaller percentage of current alcohol use among those who had data from both V1 and V2 (V1=37.14%; V1V2=59.29%).. We did not observe meaningful differences by educational attainment . Our study sample had higher attrition from male participants.

Discussion

In this study, we found associations between ambient air pollution and measures of cardiac structure and function as assessed by echocardiogram. Beyond well-known risk factors for HF, such as hyperlipidemia, hypertension, diabetes, and lifestyle factors, numerous epidemiological and toxicological studies have demonstrated associations suggesting that air pollution can contribute to incidence and progression of clinically diagnosed HF[72,73]. While understanding of the underlying biological mechanisms of these observed relationships is still evolving, deposition of air pollutants through the airway and into the alveoli is thought to result in oxidative stress and inflammation, causing clinical heart failure over long term periods[74,75].

Our study shares some similarities and differences with other related studies evaluating environmental and occupational exposures using echocardiograms. Burroughs Peña, et al. assessed occupational exposure to wood smoke, vehicle exhaust, solvents, pesticides, and metals in the ECHO-SOL study. Associations were found between exposure to burning wood at current job as well with longest held job. Effects were found for HFrEF heart failure at current job (LV ejection fraction (−3.1%; standard error [SE], 1.0 [P=0.002]), longest held job increased LV diastolic volume (6.7 mL; SE, 1.6 [P<0.0001]), decreased LV ejection fraction (−2.7%; SE, 0.6 [P<0.0001]), worse LV global longitudinal strain (1.0%; SE, 0.3 [P=0.0009]), and decreased right ventricular fractional area change (−0.02; SE, 0.004 [P<0.001]). However, wood smoke is not analogous to traffic related air pollution as we evaluated in our study, the study was cross-sectional, and no association was found between measures of cardiac structure and function and vehicle exhaust[76]. Zheng, et al. also evaluated air pollution exposure on abnormal left ventricular diastolic function using cardiac imaging in the China Hypertension Study. They used annual average PM2.5, PM10 and NO2 concentrations obtained from the chemical data assimilation system of the Institute of Atmospheric Physics, Chinese Academy of Sciences. The ORs (95%

CI) for ALVDF in the fully adjusted model were 1.31 (1.11–1.56), 1.11 (1.01–1.21) and 1.18 (0.90–1.54) for an increase of 10 μg/m3 of $PM_{2.5}$, PM_{10} and NO_2, respectively. Our study has important differences. Zheng, et al. estimated air pollution at 15 km resolution and the study was cross-sectional in nature. We estimated exposure at participant residences and our aim was to estimate progression of subclinical heart failure[77]. Our cross sectional results were largely consistent with our longitudinal findings.

This study used echocardiogram data as a measurement for preclinical heart failure. Because the echocardiograms are non-invasive and pose minimal risk to the participant, it allowed us to evaluate the progression of disease between visits and provides insight into the etiologic effects of air pollution on progression[78]. Subclinical disease measurement allows for the examination of pathophysiological changes before clinical heart failure, providing evidence for studies of prevention and disease risk. The use of continuous measurements in our study also results in higher power to detect risk associations as compared to categorizing a participants disease risk (e.g. clinical heart failure)[79]. Subclinical vascular disease has been shown to be an independent prediction of future adverse cardiovascular events[57,58,80], so we feel that the investigation of preclinical heart failure progression in this study is appropriate to our research question. As a basis for understanding biological mechanisms for disease, it is important that researchers and clinicians understand the full range of risk factors because intervention strategies can vary at different time points in disease pathogenesis.

Our results are most consistent with progression of HFrEF (systolic dysfunction of the heart). Systolic dysfunction or heart failure with reduced ejection fraction (HFrEF) refers the inability of the heart's pumping ability to meet the metabolic requirements of the body due to damage, classically as a sequela of myocardial infarction or myocarditis[8181]. It is also referred to

as to contractility impairment; in patients experiencing HFrEF, stroke output of the heart is reduced and the forward flow is compromised (i.e., heart muscle is too weak to pump resulting in low ejection fraction). Both GLS and LVEF are measures of HFrEF. GLS expresses the longitudinal shortening or the change in length as a proportion of the baseline length of the assessed segment of heart tissue. GLS is expressed as a negative measurement; a value of -20% indicates that the (average of) observed distances between two points on the heart shorten by 20% between diastole and systole. Normal values for GLS are more negative than than -18% for adults and values greater than -16% are considered abnormal[82]. We estimated an average increase of 0.52% per year of 4 $\mu g/m^3$ of air pollutants on rate of change on GLS, a value that suggests potential transition to clinically abnormal GLS). LVEF represents the volumetric fraction of fluid ejected from the left ventricle. Normal values for LVEF range from 50-75%. Low ejection fraction is considered less than 50%[83]. We found an average decrease of LVEF of -1.05% per year associated with an interquartile range elevation in PM2.5 concentrations. GLS and LVEF are both established independent predictors of incident HF over follow-up ranging from 3-10 years [84,85]. Our results were similar in our cross sectional analyses. Our findings suggest that air pollution could be an important contributor to progression of HFrEF.

Our results are less consistent with development of HPpEF[83]. HPpEF is a condition in which the heart pumps normally, but is too stiff to relax and fill properly. HPpEF results from a number of factors including the inability of the ventricular myocardium to relax, increased ventricular wall thickness, and/or buildup of the interstitial collagen of the myocardium[81]. In the present study, evidence consistent with HFpEF was assessed by e', E/e' ratio, and LAVI. The e' wave is a measure of early diastolic filling and the E/e' ratio is an index that indicates LV filling pressure and represents the ratio of early diastolic mitral inflow velocity to early diastolic

annulus velocity. E/e'predicts mortality and cardiovascular events[87]. We found no associations with air pollutants and either of these measures. LAVI is measured as the left atrial volume divided by the body surface area. Increased LAVI has been shown to predict mortality after myocardial infarction, HF hospitalization, and mortality[88]. Interestingly, we found that air pollution was associated with a 1.39 ml/m^2 decrease per year in LAVI. While this finding does not support our hypothesis, a previous study of this cohort also found paradoxical relationships between HFrEF measures and LAVI[89]. Because LAVI is a transient measure that is highly variable by time of day or day of the week, it is possible that it is not as useful for the study of the chronic effects of long-term environmental exposures as variables which do not display as much short-term variation; it may be more well-suited for study of short-term changes in physiological states or environmental exposures. [64].

We also evaluated measures of cardiac structure. RWT is measured as 2 times posterior wall thickness divided by the LV diastolic diameter. Decreased RWT portends CVD and adverse events[90,91]. We found associations between air pollution and progression of decreased RWT; an IQR increase in $PM_{2.5}$ was associated with a decrement of 1.55 RWT per year (RWT is expressed positively, but reflects a decrease in thickness) . LVMI is another independent predictor of risk of cardiac death. LVMI is the estimated mass of the left ventricle, indexed to the body surface area (function of height-squared), and increased LVMI is associated with hard coronary heart disease events, CVD death, and HF independent of coronary artery calcium score[90,91]. We did not find associations of air pollution exposure with progression of LVMI. Taken together, our results suggest that air pollution is associated with increased HPrEF, but not HPpEF subclinical heart failure.

In general, longitudinal analyses can be limited by participants being lost to follow up. In our study, we find selection bias to be less of a concern due to a number of our observed characteristics. First, we had very high retention over an average follow-up time of 4.3 years between our study observations (1468/1630 [9.00%] of participants at the first ECHO-SOL visit attended the second visit) V1=162; V1V2 = 1468). Additionally, we did not observe meaningful differences in baseline echocardiographic measures based on attendance at outcome due to attrition. We observed small differences in the two groups in terms of demographic characteristics, but the groups were largely similar and due to our low attrition, we do not expect those potential confounders to be highly influential on our exposure-outcome association estimates. Furthermore, our study was a representative sample of the larger HCHS/SOL study and as we included strata level weights in our modelling approach, we are less likely to be subject to sampling bias at either exam. Finally, because of our application of the mixed effect model—which includes all available data from all participants for the examinations at which there is data available-- we expect that the analytic approach to our research question gives a good estimate of the true effect of progression in our sample.

Importantly, our study was conducted among a cohort of Hispanic/Latinos. As forementioned, rates of heart failure among Hispanic/Latinos are higher than rates in non-Hispanic white and Asian populations.[55] Previous study of ECHO-SOL indicates that cardiac dysfunction is underdiagnosed in Hispanic/Latino adults and that the effect of cardiac dysfunction is heterogenous with respect to country of origin[94]. The prevalence of subclinical heart failure was high in this cohort; the majority of the population was asymptomatic for LV systolic or diastolic dysfunction, thus indicating that the projected clinical heart failure will be high in this cohort. However, data availability for heart failure in Hispanic/Latino populations is

insufficient, and while it is known that this population has high proportions of CVD risk factors which predispose heart failure (e.g. hypertension, diabetes, and obesity)[55], environmental risk factors can be especially difficult to disentangle. In particular, data on is limited on undocumented populations of Hispanic/Latinos in the US for both occupational and environmental exposures and CVD information[55,94] and thus prevalence and incidence of clinical and/or subclinical heart failure could be underestimated. As the Hispanic/Latino population in the US grows, identification of risks in this population is crucial to reducing heart failure incidence and prevalence, as well as lowering healthcare costs and supporting health equity[55]. Our study suggests that air pollution is potentially a modifiable risk factor for heart failure progression among Hispanic/Latino populations.

There are some notable limitations and strengths to this study. First, we used data from Hispanic/Latino populations in 4 US cities, so caution should be exercised in generalizing these results to other populations. Exposure measurement error is expected because only outdoor exposure concentrations at residential address are measured in this study, but there are many microenvironments in which individuals spend their time. However, this is a common problem in environmental epidemiology and given our statistical modelling, it is reasonable to assume here that exposure misclassification is largely non-differential. There is potential for selection bias into ECHO-SOL from the larger HCHS-SOL study as those who participate in ECHO-SOL may be different than those who elected to not participate.

We did not use cardiac MRIs, a standard for quantification of cardiac structure. Strengths of this study include our state-of-the-art exposure estimation and comprehensive covariate information in a well-characterized epidemiologic cohort. In addition, we conducted this study in a population of Hispanic/Latinos, an important understudied population.

Conclusions

To our knowledge, this is the first study to evaluate the effects of long-term air pollution exposure on longitudinal change in subclinical measures of cardiac structure and function as measured by serial echocardiography. Air pollution could represent a modifiable risk factor for the development and progression of heart failure and this work provides insights into the role that air pollution plays in the development of CVD, supporting the hypothesis that these air pollutants contribute to reduction in left ventricular systolic function. The results of this study can inform future studies of environmental impacts of progression to clinical heart failure, a disease with high morbidity and mortality.

Table 1: Demographic Characteristics of ECHO-SOL participants		
	Visit 1	Visit 2
N	1630	1468
Age (mean, SD)	55.3 (7.4)	
Gender		
Male (n, %)	569 (34.9)	
Female (n, %)	1061 (65.1)	
Education		
No high school diploma or GED (n, %)	597 (36.6)	
At most high school diploma or GED (n, %)	362 (22.2)	
Greater than high school or GED	668 (40.1)	
Missing	3 (0.2)	
Current Smoking		
No (i.e. never or former smoker) (n, %)	1358 (83.3)	1262 (86.0)
Yes (i.e. current smoker) (n, %)	270 (16.6)	205 (14.0)
Missing (n, %)	2 (0.1)	1 (0.1)
Packyears of smoking (mean, SD)	7.8 (16.6)	
Diabetes		
Non-diabetic (n, %)	1203 (73.8)	870 (59.3)
Diabetic (n, %)	427 (26.2)	598 (40.7)
Minutes per day of recreational vigorous physical activity		
0	1481 (90.3)	
>0 - 25	57 (3.5)	
>25	88 (5.4)	
Unknown	3 (0.2)	
Alcohol consumption		
Never (n, %)	391 (24.0)	288 (19.6)
Former (n, %)	539 (33.1)	386 (26.3)
Current (n, %)	700 (42.9)	793 (54.0)
Missing (n, %)	0 (0.0)	1 (0.1)
BMI (kg/m2) (mean, SD)	30.2 (5.7)	
Statin use		
Yes (n, %)	302 (18.5)	0 (0.0)
No (n, %)	1295 (79.5)	0 (0.0)
Unknown	33 (2.0)	1468 (100.0)
Hypertension		
Yes (n, %)	862 (52.9)	649 (44.2)
No (n, %)	768 (47.1)	819 (55.8)
Systolic BP (mean, SD)	136.0 (18.6)	131.0 (19.1)
Diastolic BP (mean, SD)	78.2 (11.3)	74.6 (11.2)
NSES z score (mean, SD)	1.1 (0.8)	1.1 (0.7)
Site of ECHO-SOL Visit		
Bronx (n, %)	494 (30.3)	428 (29.2)
Chicago (n, %)	380 (23.3)	349 (23.8)
San Diego (n, %)	314 (19.3)	295 (20.1)
Miami (n, %)	442 (27.2)	396 (27.0)
Year of Visit		
2011 (n,%)	168 (10.3)	N/A
2012 (n, %)	761 (46.7)	N/A
2013 (n, %)	577 (35.4)	N/A
2014 (n, %)	124 (7.6)	N/A
2015 (n, %)	N/A	5 (0.3)
2016 (n, %)	N/A	651 (44.4)
2017 (n, %)	N/A	581 (39.6)
2018 (n, %)	N/A	231 (15.7)
Average Follow Up (years) (mean, SD)	N/A	4.3 (0.4)

Table 2: Descriptive Statistics of Outcome Variables		
	Visit 1	Visit 2
Average global longitudinal strain (GLS) (%) (mean, SE)	-17.53 (0.14)	-17.03 (0.13)
LV mass index (LVMI) g/m2 (mean, SE)	82.76 (0.71)	85.5. (0.81)
LV ejection fraction (EFBP) (%) (mean, SE)	59.71 (0.20)	59.01 (0.24)
E/e′ ratio (mean, SE)	9.99 (0.13)	10.39 (0.16)
E' (mean, SE)	8.07 (0.10)	7.02 (0.07)
LA volume index	23.03 (0.24)	23.00 (0.32)
Relative Wall Thickness	0.40 (0.004)	0.46 (0.005)

Table 3: The Effect of $PM_{2.5}$ on Echocardiographic Measures				
Category	Outcome	Model 1	Model 2	Model 3
Systolic function	Average global longitudinal strain (GLS) (%)	0.52 (0.2, 0.83)	0.57 (0.22, 0.93)	0.59 (2.4, 0.93)
	LV ejection fraction (LVEF) (%)	-1.05 (-1.73, -0.37)	-1.19 (-1.94, -0.44)	-1.31 (-2.06, -0.56)
Diastolic function	LA volume Index (LAVI) (ml/m2)	-1.39 (-2.25, -0.54)	-2.12 (-3.08, -1.16)	-2.08 (-3.03, -1.13)
	Eprime	0.03 (-0.18, 0.23)	0.12 (-0.11, 0.35)	0.12 (-0.1, 0.33)
	E-Eprime ratio	0.04 (-0.32, 0.4)	-0.12 (-0.58, 1.52)	-0.1 (-0.49, 0.29)
Cardiac structure	Relative Wall Thickness	1.19 (0.47, 1.91)	1.55 (0.72, 2.38)	1.55 (0.71, 2.39)
	LV Mass Index (C) dl g/m2 (LVMI)	1.02 (-0.67, 3.2.73)	1.07 (-1.61, 3.78)	1.30 (-1.40, 3.69)

MODEL 1: $PM_{2.5}$ + ECHO-SOL center, alcohol use, smoking status, physical activity, education, gender, age, cigarette pack years, BMI (LAVI and LVMI do not not include BMI)

Model 2: Model 1 + NSES

Model 3: Model 2 + clinical variables

- sbp, dbp, diabetes type II, hypertension, statin use

PM2.5 is per 4 $\mu g/m^3$

Rate is per 1 year

Table 4: The Effect of NO_2 on Echocardiographic Measures				
Category	Outcome	Model 1	Model 2	Model 3
Systolic function	Average global longitudinal strain (GLS) (%)	0.12 (0.03, 0.2)	0.18 (-0.08, 0.45)	0.23 (-0.04, 0.49)
	LV ejection fraction (LVEF) (%)	-0.28 (-0.82, 0.25)	-0.44 (-1.00, 0.13)	-0.51 (-1.07, 0.06)
Diastolic function	LA volume Index (LAVI) (ml/m2)	-0.94 (-1.63, -0.26)	-1.27 (-2.01, -0.52)	-1.35 (-2.08, -0.62)
	Eprime	0.72 (-0.04, 0.28)	0.08 (-0.09, 0.55)	-0.09 (-0.09, 0.24)
	E-Eprime ratio	0.28 (-0.38, 0.2)	-0.14 (-0.45, 0.17)	-0.1 (-0.40, 0.20)
Cardiac structure	Relative Wall Thickness	-0.17 (-0.71, 0.38)	-0.31 (-0.91, 0.29)	-0.27 (-0.87, 0.34)
	LV Mass (C) dl g/m2	-0.06 (-1.93, 1.81)	0.02 (-2.00, 4.96)	0.05 (-1.93, 2.04)

MODEL 1: NO2 + ECHO-SOL center, alcohol use, smoking status, physical activity, education, gender, age, cigarette pack years, BMI (LAVI does not include BMI)

Model 2: Model 1 + NSES

Model 3: Model 2 + clinical variables

- sbp, dbp, diabetes type II, hypertension, statin use

NO2 is per 10 ppb

Rate is per 1 year

Table 5: The Cross-sectional Coefficients for Effect of $PM_{2.5}$ on Echocardiographic Measures				
Category	Outcome	Model 1	Model 2	Model 3
Systolic function	Average global longitudinal strain (GLS) (%)	0.68(0.20, 2.0)	0.64 (0.10, 1.15)	0.63 (0.09, 0.15)
	LV ejection fraction (LVEF) (%)	-2.00 (-3.80, -2.40)	-0.15 (-1.27, 1.07)	-0.13 (-1.29, 1.05)
Diastolic function	LA volume Index (LAVI) (ml/m2)	2.50 (0.10, 4.90)	1.81 (0.80, 4.96)	2.56 (0.11, 5.15)
	e'	-0.10 (-0.10, 0.40)	-0.10 (-0.09, 0.40)	-0.55 (-0.73, 0.60)
	E-e'ratio	-0.30 (-1.36, 0.69)	-0.20 (-1.23, 0.85)	-0.33 (-1.57, 1.69)
Cardiac structure	Relative Wall Thickness	0.03 (-0.004, 0.04)	0.05 (-0.006, 0.04)	0.04 (-0.005, 0.04)
	LV Mass (C) dl g/m2	-0.64 (-6.70, 5.50)	-0.66 (-6.99, 5.66)	-0.72 (-5.55, 0.58)

MODEL 1: PM2.5 + ECHO-SOL center, alcohol use, smoking status, physical activity, education, gender, age, cigarette pack years, BMI (LAVI does not include BMI)

Model 2: Model 1 + NSES

Model 3: Model 2 + clinical variables

- sbp, dbp, diabetes type II, hypertension, statin use

PM2.5 is per 4 $\mu g/m^3$

Table 6: The Cross-sectional Coefficients for Effect of NO_2 on Echocardiographic Measures				
Category	Outcome	Model 1	Model 2	Model 3
Systolic function	Average global longitudinal strain (GLS) (%)	0.61 (0.03, 1.19)	0.59 (0.01, 1.07)	0.56 (-0.86, 1.01)
	LV ejection fraction (LVEF) (%)	-0.90 (-2.12, 0.32)	-0.99 (1.00, 0.19)	-0.26 (-1.62, 1.01)
Diastolic function	LA volume Index (LAVI) (ml/m2)	-0.98 (-2.64, 0.68)	-0.72 (-2.01, 0.52)	-1.88 (-2.68, 0.53)
	e’	0.04 (-0.32, 0.40)	0.08 (-0.09, 0.55)	-0.09 (-0.09, 0.24)
	E-e’ratio	0.22 (-0.45, 0.90)	-0.57 (-0.98, 0.18)	-0.10 (-0.30, 0.20)
Cardiac structure	Relative Wall Thickness	0.02 (-0.45, 0.04)	0.01 (-0.46, 0.04)	-0.02 (-0.45, 0.05)
	LV Mass (C) dl g/m2	-1.40 (-6.30, 3.65)	-0.98 (-5.64, 4.72)	-1.93 (-6.54, 3.94)

- All models control for center, alcohol use, smoking status, physical activity, education, gender, age, cigarette pack years, BMI
- NO_2 is per 10 ppb

Table 7: Descriptive Statistics of Echo Measures at Visit 1 by Attendance at ECHO-SOL Visits				
	Echo at Visit 1 Only (N=162)		Echos at Visits 1 and 2 (N=1468)	
Outcome	Mean	Std	Mean	Std
Average global longitudinal strain (GLS) (%) (mean, SE)	-17.37	2.99	-17.26	2.77
LV mass index (LVMI) g/m2 (mean, SE)	84.70	22.24	83.31	23.48
LV ejection fraction (EFBP) (%) (mean, SE)	59.10	5.44	59.40	6.17
E/e′ ratio (mean, SE)	9.83	3.18	10.60	4.02
E' (mean, SE)	8.21	2.35	7.20	1.84
LA volume index	23.09	7.88	23.09	8.23
Relative Wall Thickness	0.40	0.09	0.66	0.09

Table 8: Descriptive Statistics in Baseline Demographic Characteristics, by Attendance at ECHO-SOL Visits				
	Visit 1 Only N= 162		Visit 1 and Visit 2N=1468	
Variable	Frequency	Percent	Frequency	Percent
Center				
Bronx	69	39.43	481	29.28
Chicago	35	20	383	23.31
Miami	49	28	451	27.45
San Diego	22	12.57	328	19.96
Alcohol Use				
Never	36	20.57	322	19.6
Former	73	41.71	428	26.05
Current	65	37.14	892	59.29
Missing	1	0.57	1	0.06
Smoking				
Never	131	74.86	1415	86.12
Current	43	24.57	227	13.82
Missing	1	0.57	1	0.06
Education				
No high school diploma or GED (n, %)	66	37.71	592	36.03
high school diploma or GED (n, %)	42	24	268	22.4
Greater than high school or GED	66	37.71	680	41.39
Missing	1	0.57	3	0.18
Gender				
Female	93	53.14	10.9	66.52
Male	82	46.86	550	33.48
Age (mean, std)	57.96	7.82	55.90	8.00

Figure 1. Sample Selection

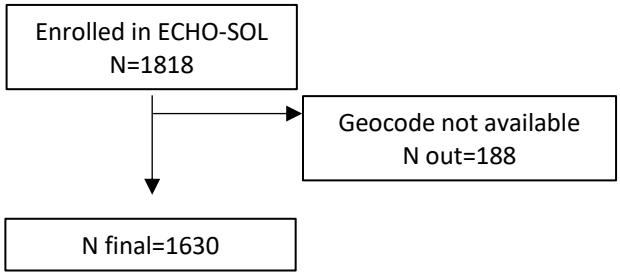


Figure 2: Annual Average $PM_{2.5}$ by Center and Visit

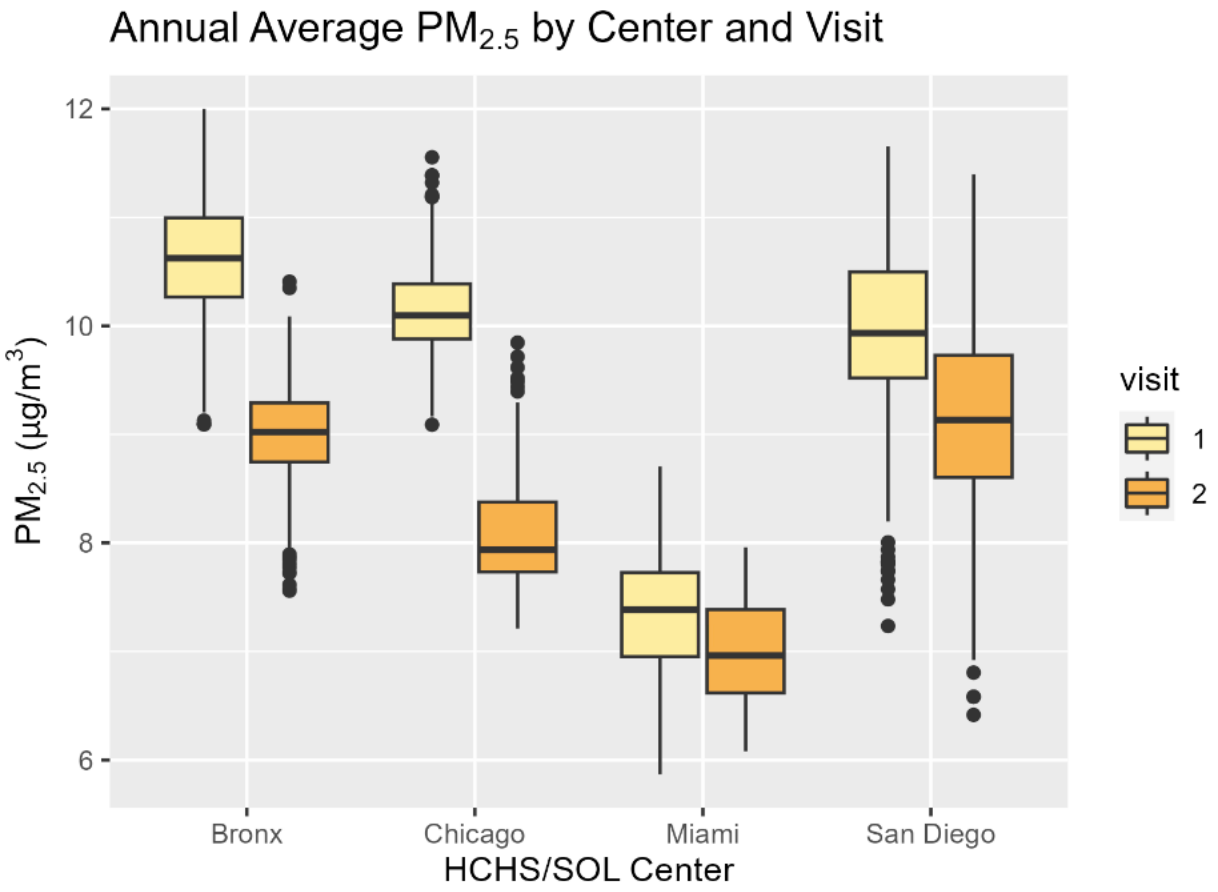
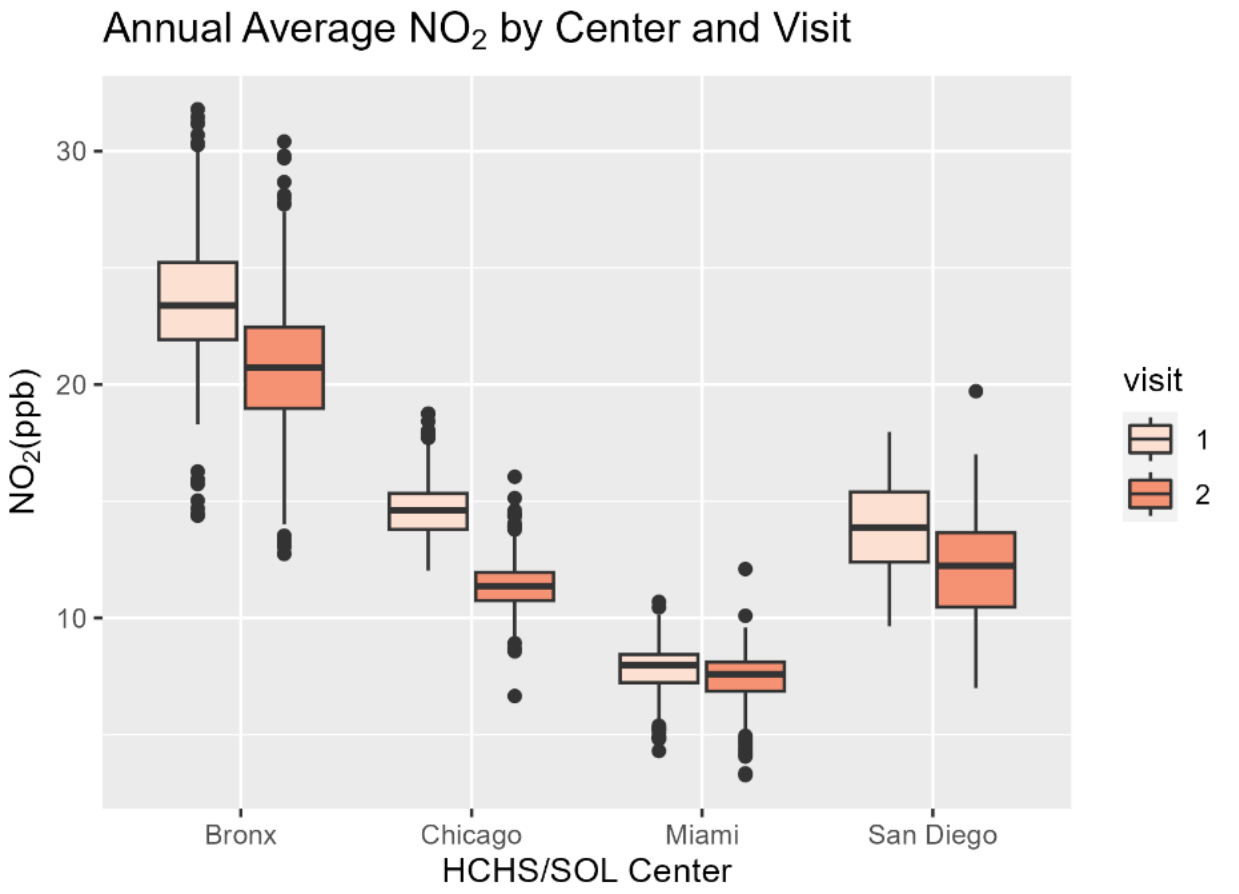


Figure 3: Annual Average NO_2 by Center and Visit



Chapter 4: The Effect of Spatial Distance on Cardiovacular Disease Estimates

Introduction

Ambient air pollution poses a significant risk to population health in the United States and cardiovascular disease (CVD) represents a large portion of adverse outcomes in observational cohort studies. In cohort studies using long term-exposure to ambient air pollutants on a health outcome, the true exposure is for the individual is unknown. It is therefore necessary to utilize advanced statistical modeling methods to predict exposures at fine spatial resolution in order to minimize exposure measurement error[26,92,95]. While it has been shown that the accuracy of the prediction model to minimize the error in the predictions can improve health effect estimations, the differences in the health effect estimations in the *spatial resolution* using the same predictive exposure model is unknown. The difference using fine spatial resolution compared to more coarse resolution could create exposure measurement error resulting in biased effect estimates. We propose to test the difference in the estimate of risk of CVD and mortality due to long term traffic-related air pollution exposures from a state-of-the-art exposure model to more coarsely measured pollution models for a pooled cohort of four observational studies.

Methods

Study Design

This study uses a pooled prospective cohort design. Data for this study come from four US cohorts: Cardiovascular Health Study (CHS) (1989), Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA-Air)(2000) and MESA Family New Recruits (FN) (2004), Reasons for Geographic and Racial Differences in Stroke (REGARDS) (2003), and Sister Study (Sister)(2003). The study designs of these cohorts have been described elsewhere. Briefly, MESA is a longitudinal study of adults aged 45-84 years recruited from six US cities (Baltimore

City and Baltimore County, Maryland; Chicago, Illinois; Forsyth County, North Carolina; Los Angeles County, California; New York, New York; and St. Paul, Minnesota)[47]. CHS is a longitudinal cohort study of adults aged 65 and older from Forsyth County, North Carolina; Sacramento County, California; Washington County, Maryland; and Pittsburgh, Pennsylvania[96]. REGARDS is a longitudinal study of 30,000 African American and white adults aged ≧45 years in the Southeastern US[97]. Sister Study is a large cohort of sister pairs enrolled 50,884 U.S. and Puerto Rican women 35–74 years of age[98].

Exposure

We use our spatio-temporal model to estimate air pollution exposure in this study[49]. Our primary exposure is time-varying air pollutant concentrations. We estimated monthly two-year moving averages for $PM_{2.5}$ and NO_2 estimated at the census block of participant addresses geocoded from billing records, accounting for residential history. We compared health effect estimates to an equivalent averaging period, but at more coarse spatial resolutions of 100m and 1000m for our included participants. We created grids of points spaced by 100m and 1000m, and assigned each address to its nearest point in each grid, averaged over residential history.

Data Harmonization

We harmonized covariate and outcome information across the four cohorts to create a large, pooled dataset used in our analysis. Our data harmonization has two components: identification of equivalent event classifications across cohorts and identification of equivalent individual level characteristics. Covariates of interest were measured at baseline and include age, sex, race/ethnicity, employment, personal income, smoking status, alcohol use, physical activity, education, and body mass index (BMI). We also include clinical variables: cholesterol, systolic and diastolic blood pressure, hypertension status, use of hypertension medication, statin use,

family history of heart disease, and diabetes status. The primary outcomes of this study are incidence of myocardial infarction (MI), stroke, and all CVD, as well as CVD related mortality and non-accidental death. Outcome information was adjudicated separately in each cohort.

Statistical analysis

Cox proportional hazards regression was used to compare hazard ratio estimates between three different exposure distances. We considered study time in months as the primary time scale, interacted by cohort. In order to account for differences in ages between our cohorts, we allow for an interaction term between age and cohort. Duration was measured as time from enrollment to event or date of censor. Date of study enrollment was considered as entry date for all cohorts. Individuals were censored at last known study contact date. To account for differences in enrollment year between cohorts and secular declines in air pollution over time, all models were adjusted for year of enrollment. The model is generally described as:

$$h(t, X(t)) = h_0(t)e^{\beta E + \delta(E \times t)}$$

Where X contains the entire collection of time varying (air pollution) and time-independent predictors (e.g. race-ethnicity). Air pollutant exposures were measured as per interquartile range (IQR).

We compared outcome effect estimates from using the following modeling strategy:

Model 1: *point based time-varying pollution + race + ethnicity + sex+ smoking status + alcohol use + physical activity+ BMI+ employment status+ personal income+ education + year*cohort + cholesterol+ sbp+ dbp + hypertension status + use of hypertensive medication + statin use + diabetes comorbidity+ division + weighted scale of SES*

Model 2: *100 m time-varying pollution + race + ethnicity + sex+ smoking status + alcohol use + physical activity+ BMI+ employment status+ personal income+ education + year*cohort +*

cholesterol+ sbp+ dbp + hypertension status + use of hypertensive medication + statin use + diabetes comorbidity+ division + weighted scale of SES

Model 3: *1000 m time-varying pollution + race + ethnicity + sex+ smoking status + alcohol use + physical activity+ BMI+ employment status+ personal income+ education + year*cohort + cholesterol+ sbp+ dbp + hypertension status + use of hypertensive medication + statin use + diabetes comorbidity + division + weighted scale of SES*

Results

Our pooled cohort included 94,503 individuals. Descriptive statistics are reported in Table 1. The mean age of our sample was 59.9 years and our sample was 79.3% female. The majority of our sample identified as white (71.8%) and had college degree or higher education (42.5%). A large percentage of our sample reported smoking (37.5%) and alcohol use (67.5%). Table 2 shows the numbers of outcomes by cohort. Our analysis included 3862 MI, 4659 strokes, 9147 incident CVD events, 15403 non-accidental deaths, and 3162 CVD-related deaths.

Table 3 shows the distributions of pollutants by distance and cohort. The highest point median $PM_{2.5}$ was in the CHS cohort (median = 13.05 $\mu g/m^3$, range 2.86-25.32 $\mu g/m^3$) with the lowest being from Sister (median = 8.58 $\mu g/m^3$, range 1.77-22.97 $\mu g/m^3$). The highest point median NO_2 in was in MESA (median = 14.59, range = 1.37-70.10 ppb) and the lowest was from Sister (median=6.49, range=0.44-49.7 ppb). We observed little variation within cohorts based off of distance to grid cell.

Figure 1 shows a map of all participant residential addresses. Overall, we observed geographic heterogeneity across the US (Hawaii and Alaska are excluded from our analysis). Figure 2 shows

the correlation of our predictive estimates by distance. We observe a high correlation in NO2 estimates between our point prediction and 100m. We observe lower correlation between our PM2.5 estimates for both distances. Figure 3 shows a forest plot of the hazard ratios by point, 100m, 1000m for $PM_{2.5}$. Overall, we found no difference in the health effects by distance. Figure 4shows a forest plot of the hazard ratios by point, 100m, 1000m for NO_2. We observed no difference in the hazard ratio by distance to grid cell.

Discussion

Our findings in this study do not demonstrate a difference in CVD health effects using predictive air pollution models at varying spatial resolutions. This observation was true for both $PM_{2.5}$ and NO_2, despite that NO_2 is more spatially granular. To our knowledge, this is the first study to test the differences in health effects of the same predictive model. The findings of this study are unexpected, given vast epidemiological evidence that exposure measurement error can bias results, particularly in environmental epidemiology[99,100]. We expect exposure measurement error in studies of air pollution, and thus use analytic approaches (e.g. regression models) to reduce measurement error[92,93,95,101]. In particular, there is a large body of work describing the importance of spatial resolution in the field of air pollution exposure science, and varying methods to address spatial confounding[95,101–103]. However, our results demonstrate that it is possible that spatial measurement error may not impact health effects in every case.

There are some possible explanations for our findings. First, in contrast to other studies using grid measurements from exposure predictions[104–107], we did not average over the entire grid cell. Rather, we estimated the difference in hazard ratio based solely from the distance to our nearest

grid point. Our resulting predictions are therefore highly correlated, and thus our hazard ratios are similar. Second, we estimated exposure solely based on residential location. Travelers to work may experience differences in traffic-related air pollution exposure due to both residential location and distance to travel away to residential address for work, as well as duration in traffic[108]. The extent of exposure measurement error could be due to incomplete data on human mobility. There is possible measurement error due to Berkson-like or Classical-like biases[92,93], but given our large sample size and geographic area, we believe that there are no meaningful threats to the internal validity of this study.

There are a number of strengths of this study. We are the first to examine the impact of distance from the same predictive exposure model on health effects. We used a pooled cohort of four well-characterized epidemiological studies and state-of-the-art air pollution modelling. Our cohort was large (N=94,503) and our sample was geographically diverse. We had extensive covariate and outcome information, as well as residential history for our participants. There are some notable limitations. Primarily, we are not able to account for time spent out of residence. There is also the possibility of geocoding errors, which could affect our prediction of air pollution estimates[109].

Conclusion

In our analysis, we did not find evidence of exposure measurement error by distance. We are the first study to estimate the difference in health effects using the same predictive model for air

pollution assessment. The results of this study can inform future researchers on the merits of obtaining point-based data.

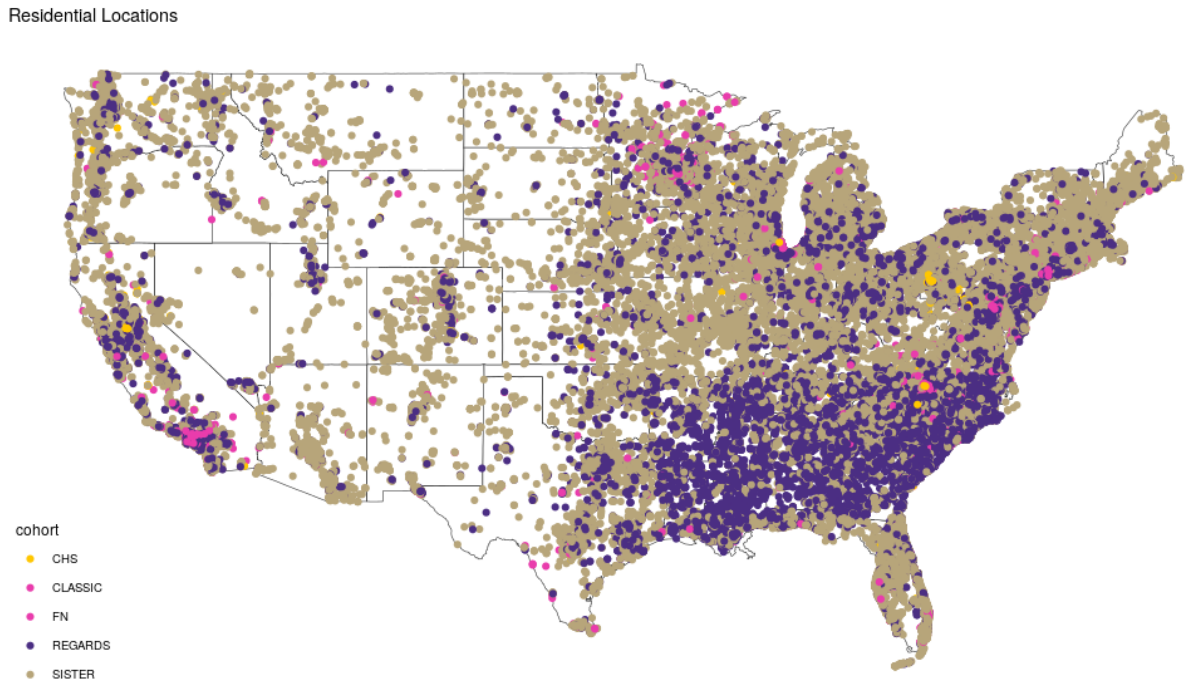
Table 1: Descriptive Statistics of Cohorts					
	CHS	**MESA**	**REGARDS**	**SISTER**	**Overall**
Variable	(N=5888)	(N=7551)	(N=30183)	(N=50881)	(N=94503)
Age (Mean, SD)	72.8 (5.62)	61.9 (10.2)	64.8 (9.43)	55.2 (8.97)	59.9 (10.6)
Sex (N, %)					
Female	3393 (57.6%)	4031 (53.4%)	16632 (55.1%)	50881 (100%)	74937 (79.3%)
Male	2495 (42.4%)	3520 (46.6%)	13551 (44.9%)	0 (0%)	19566 (20.7%)
Race/Ethnicity (N, %)					
White	4876 (82.8%)	2781 (36.8%)	17669 (58.5%)	42556 (83.6%)	67882 (71.8%)
Black/AA	916 (15.6%)	2171 (28.8%)	12514 (41.5%)	4190 (8.2%)	19791 (20.9%)
Asian/API	3 (0.1%)	805 (10.7%)	0 (0%)	341 (0.7%)	1149 (1.2%)
AI/AN	13 (0.2%)	0 (0%)	0 (0%)	93 (0.2%)	106 (0.1%)
Other	18 (0.3%)	0 (0%)	0 (0%)	1171 (2.3%)	1189 (1.3%)
Hispanic	62 (1.1%)	1794 (23.8%)	0 (0%)	2515 (4.9%)	4371 (4.6%)
Missing	0 (0%)	0 (0%)	0 (0%)	15 (0.0%)	15 (0.0%)
Employment (N, %)					
Currently Employed	480 (8.2%)	4076 (54.0%)	6861 (22.7%)	32821 (64.5%)	44238 (46.8%)
Not Currently Employed	151 (2.6%)	239 (3.2%)	2229 (7.4%)	1869 (3.7%)	4488 (4.7%)
Retired	3003 (51.0%)	2324 (30.8%)	8655 (28.7%)	9972 (19.6%)	23954 (25.3%)
Homemaker	1057 (18.0%)	886 (11.7%)	1046 (3.5%)	5722 (11.2%)	8711 (9.2%)
Other	1197 (20.3%)	26 (0.3%)	11392 (37.7%)	497 (1.0%)	13112 (13.9%)
Education (N, %)					
<HS	1732 (29.4%)	1379 (18.3%)	3792 (12.6%)	627 (1.2%)	7530 (8.0%)
HS or GED	1620 (27.5%)	1361 (18.0%)	7804 (25.9%)	7177 (14.1%)	17962 (19.0%)
Some College	1324 (22.5%)	2205 (29.2%)	8090 (26.8%)	17181 (33.8%)	28800 (30.5%)
College Degree or Higher	1195 (20.3%)	2582 (34.2%)	10472 (34.7%)	25884 (50.9%)	40133 (42.5%)
Missing	17 (0.3%)	24 (0.3%)	25 (0.1%)	12 (0.0%)	78 (0.1%)
Smoking (N, %)					
Never	2860 (48.6%)	3800 (50.3%)	13604 (45.1%)	28550 (56.1%)	48814 (51.7%)
Current	2458 (41.7%)	2742 (36.3%)	12067 (40.0%)	18140 (35.7%)	35407 (37.5%)
Former	567 (9.6%)	986 (13.1%)	4396 (14.6%)	4175 (8.2%)	10124 (10.7%)
Missing	3 (0.1%)	23 (0.3%)	116 (0.4%)	16 (0.0%)	158 (0.2%)
Current Alcohol Use (N, %)					
No	2931 (49.8%)	1908 (25.3%)	14616 (48.4%)	9679 (19.0%)	29134 (30.8%)
Yes	2927 (49.7%)	4141 (54.8%)	15567 (51.6%)	41189 (81.0%)	63824 (67.5%)
Missing	30 (0.5%)	1502 (19.9%)	0 (0%)	13 (0.0%)	1545 (1.6%)
BMI (N, %)					
<18.5	98 (1.7%)	60 (0.8%)	318 (1.1%)	563 (1.1%)	1039 (1.1%)
18.5 - <25.0	2170 (36.9%)	2047 (27.1%)	7091 (23.5%)	18875 (37.1%)	30183 (31.9%)
25.0 - <30.0	2435 (41.4%)	2959 (39.2%)	11057 (36.6%)	16149 (31.7%)	32600 (34.5%)
30.0+	1166 (19.8%)	2485 (32.9%)	11499 (38.1%)	15277 (30.0%)	30427 (32.2%)
Missing	19 (0.3%)	0 (0%)	218 (0.7%)	17 (0.0%)	254 (0.3%)
Physical Activity (N, %)					
Q1	1465 (24.9%)	1890 (25.0%)	6145 (20.4%)	12720 (25.0%)	22220 (23.5%)
Q2	1472 (25.0%)	1887 (25.0%)	4666 (15.5%)	13225 (26.0%)	21250 (22.5%)
Q3	1480 (25.1%)	1871 (24.8%)	5020 (16.6%)	12195 (24.0%)	20566 (21.8%)
Q4	1464 (24.9%)	1882 (24.9%)	3843 (12.7%)	12706 (25.0%)	19895 (21.1%)
Missing	7 (0.1%)	21 (0.3%)	10509 (34.8%)	35 (0.1%)	10572 (11.2%)
High Cholesterol (N, %)					
No	2979 (50.6%)	3964 (52.5%)	14912 (49.4%)	17385 (34.2%)	39240 (41.5%)
Yes	2909 (49.4%)	3587 (47.5%)	15271 (50.6%)	33496 (65.8%)	55263 (58.5%)
Cholesterol Medication (N, %)					
No	5744 (97.6%)	5666 (75.0%)	19843 (65.7%)	39184 (77.0%)	70437 (74.5%)
Yes	132 (2.2%)	1355 (17.9%)	10010 (33.2%)	11697 (23.0%)	23194 (24.5%)
Missing	12 (0.2%)	530 (7.0%)	330 (1.1%)	0 (0%)	872 (0.9%)
Diabetes (N, %)					
No	5265 (89.4%)	6654 (88.1%)	23267 (77.1%)	47081 (92.5%)	82267 (87.1%)
Yes	594 (10.1%)	890 (11.8%)	6814 (22.6%)	3800 (7.5%)	12098 (12.8%)
Missing	29 (0.5%)	7 (0.1%)	102 (0.3%)	0 (0%)	138 (0.1%)
SBP Mean(SD)	137 (21.9)	126 (21.3)	128 (16.7)	115 (13.7)	121 (17.5)
DBP Mean (SD)	70.9 (11.7)	71.9 (10.3)	76.5 (9.73)	72.4 (8.78)	73.6 (9.65)

Table 2: Total Events by Cohort					
Cohort	MESA	Regards	CHS	Sister	Total
MI	342	1759	1082	679	3862
Stroke	307	2230	1159	963	4659
CVD	764	4102	2722	1559	9147
Non accidental mortality	1539	6883	5308	1673	15403
CVD mortality	364	571	2019	208	3162

Table 3: Distribution of Exposure by Cohort and Distance

Distance	Cohort	Pollutant	Mean	SD	Min	Median	Max
Point	MESA	$PM_{2.5}$	12.39	3.52	3.14	11.82	28.92
100m	MESA	$PM_{2.5}$	11.18	3.26	2.57	10.62	28.33
1000m	MESA	$PM_{2.5}$	11.14	3.26	2.50	10.58	28.68
Point	MESA	NO_2	16.54	8.64	1.35	14.59	78.46
100m	MESA	NO_2	16.50	8.61	1.35	14.58	77.84
1000m	MESA	NO_2	16.41	8.66	1.37	14.52	70.10
Point	CHS	$PM_{2.5}$	12.33	2.74	2.86	13.05	25.32
100m	CHS	$PM_{2.5}$	11.70	2.65	3.45	11.89	22.95
1000m	CHS	$PM_{2.5}$	11.64	2.66	3.25	11.83	23.03
Point	CHS	NO_2	13.33	5.30	1.81	12.69	46.65
100m	CHS	NO_2	13.32	5.31	1.82	12.67	47.02
1000m	CHS	NO_2	13.19	5.30	1.64	12.63	52.62
Point	REGARDS	$PM_{2.5}$	10.30	2.33	1.91	10.10	24.81
100m	REGARDS	$PM_{2.5}$	10.05	2.29	2.08	9.84	23.25
1000m	REGARDS	$PM_{2.5}$	10.00	2.31	2.07	9.81	23.48
Point	REGARDS	NO_2	8.16	4.92	0.55	6.93	48.12
100m	REGARDS	NO_2	8.15	4.92	0.54	6.91	47.30
1000m	REGARDS	NO_2	8.15	5.00	0.58	6.85	44.03
Point	Sister	$PM_{2.5}$	8.64	2.24	1.77	8.58	22.97
100m	Sister	$PM_{2.5}$	8.47	2.15	1.76	8.32	28.68
1000m	Sister	$PM_{2.5}$	8.41	2.17	1.71	8.26	21.55
Point	Sister	NO_2	7.44	4.19	0.44	6.49	49.70
100m	Sister	NO_2	7.43	4.18	0.46	6.47	48.46
1000m	Sister	NO_2	7.38	4.24	0.46	6.38	50.04
Point	Overall	$PM_{2.5}$	9.70	2.8	1.77	9.39	28.92
100m	Overall	$PM_{2.5}$	9.32	2.57	1.76	9.03	28.68
1000m	Overall	$PM_{2.5}$	9.27	2.59	1.71	8.98	28.68
Point	Overall	NO_2	8.99	5.95	0.44	7.35	78.46
100m	Overall	NO_2	8.97	5.94	0.46	7.34	77.84
1000m	Overall	NO_2	8.93	5.97	0.46	7.25	70.10

Figure 1: Schematic of all Residential Locations of Participants for 4 Cohorts



CHS = Cardiovascular Health Study (1989)
CLASSIC= Multi-Ethnic Study of Atherosclerosis Main Cohort (MESA)(2000)
FN= MESA Family Participants in MESA Air and MESA Air New Recruits (2004)
REGARDS= Reasons for Geographc (2003)
SISTER = Sister Study (2003)

Figure 2: Correlation Between Grid Predictions and Point Predictions

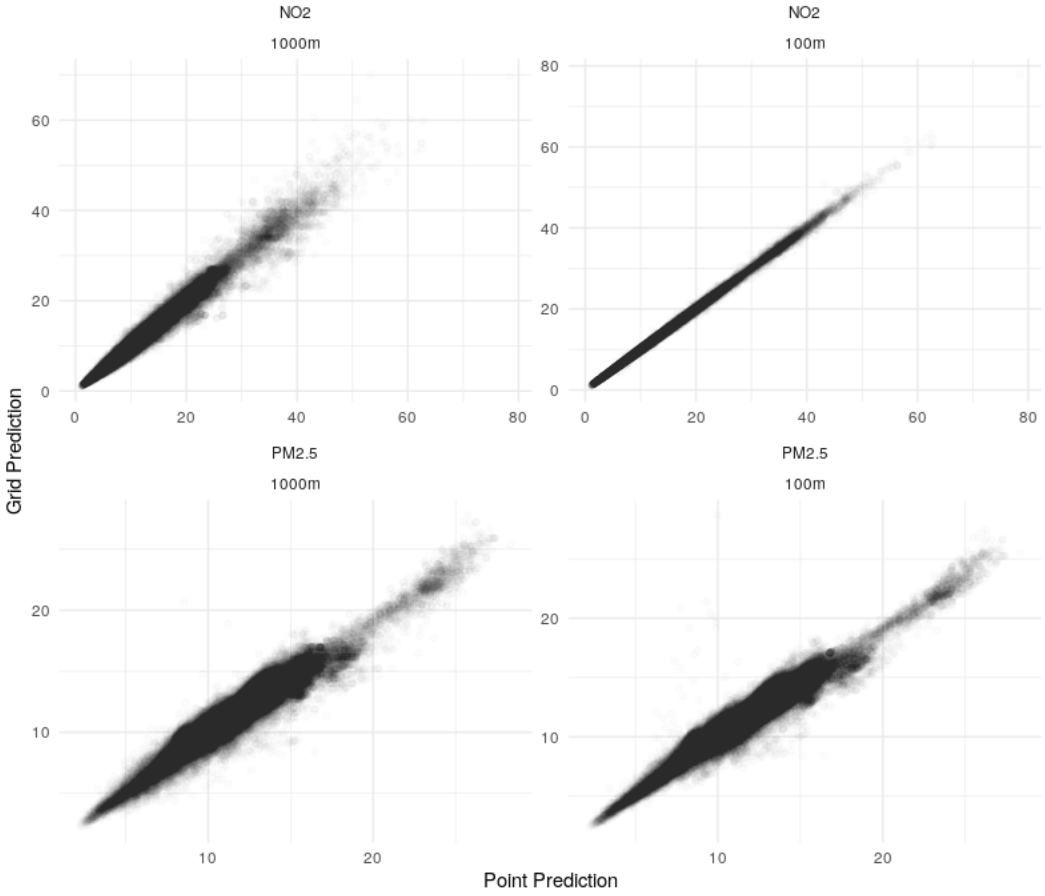


Figure 3: Forest Plot of Hazard Ratios for $PM_{2.5}$

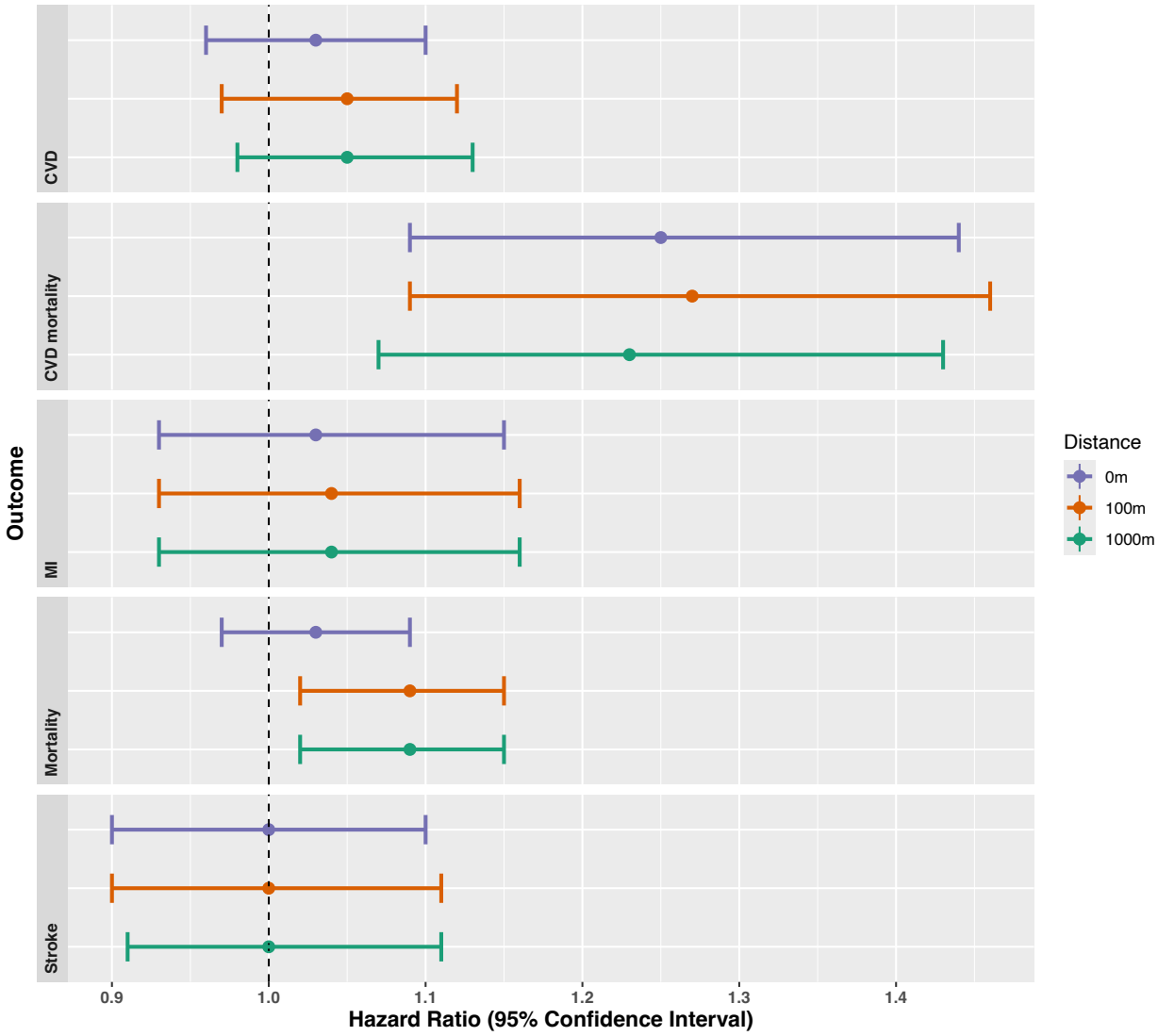
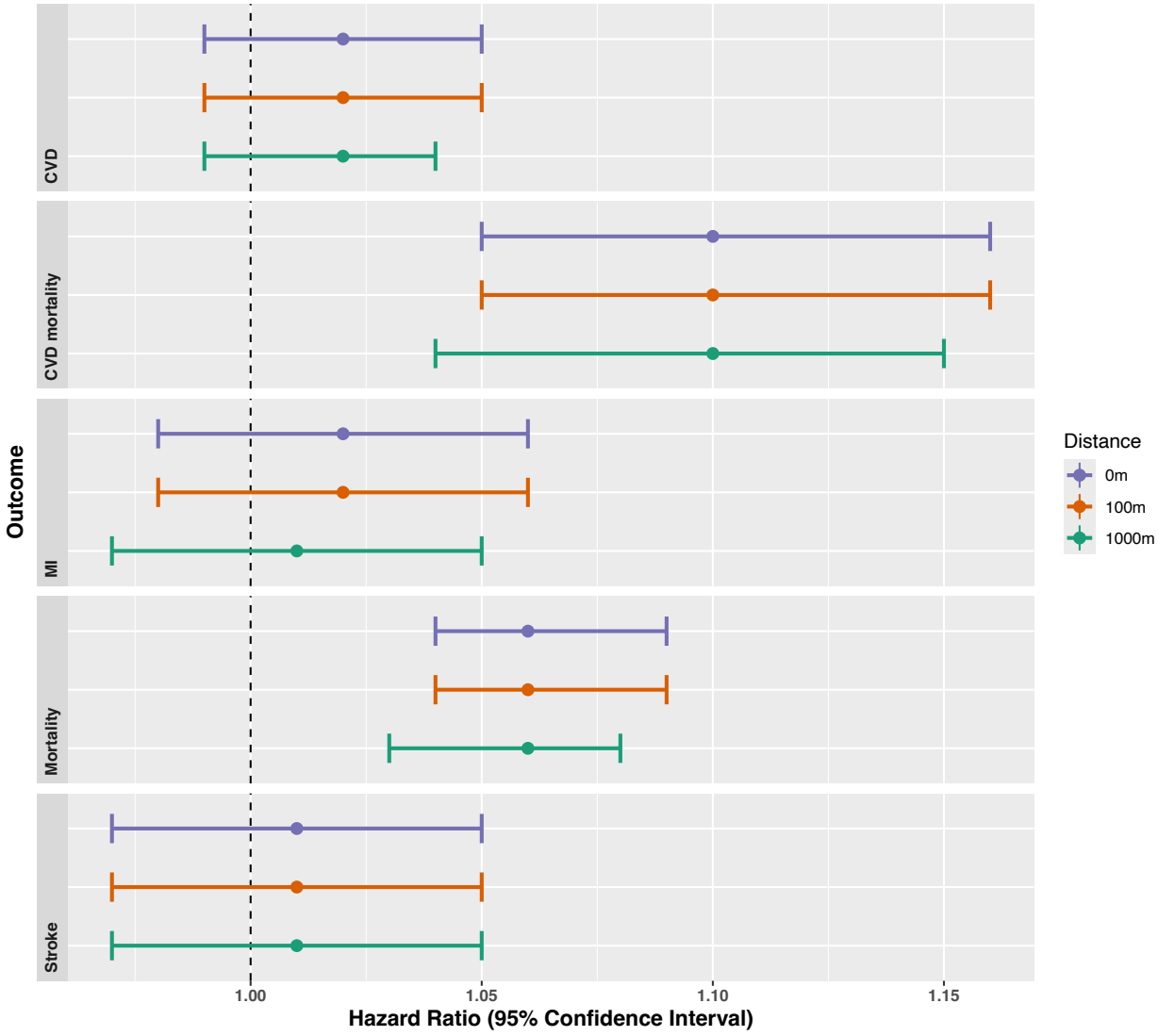


Figure 4: Forest Plot of Hazard Ratios for NO_2



Chapter 5: Conclusion

Air pollution is a known risk factor for the development and progression of CVD[7]. Despite a large body of literature in the epidemiologic literature, relatively little is known about the biological processes underlying this relationship. The previous analyses contribute to the growing body of literature to further scientific knowledge about air pollution exposure and health. Our overall findings were mixed. In Chapter 2, we found no associations between air pollution and the prevalence of CHIP in two epidemiological cohorts. In Chapter 3, we found that air pollution was associated with subclinical markers of systolic heart failure and heart remodeling. Chapter 4 aimed to quantify the difference in health effects from varying spatial distances to participant addresses and we found that there were no differences in effect estimates.

In Chapter 2, we hypothesized that because both CHIP and air pollution are related to the development of CVD[40], it was possible that air pollution could be associated with the development of CHIP. Because previous studies have also found CHIP to be associated with smoking and though smoking is not analogous to air pollution inhalation, we hypothesized that air pollution may have the same underlying biological processes that contribute to CVD as smoking[36,110]. We used two large, epidemiological cohorts with outstanding covariate information as well as a large database for genetic and other -omic data. We did not find an association for any of our analyses. Consistent with other literature, however, we found that age was the biggest predictor of CHIP[37,40,110]. Contrastingly, we were not able to replicate findings consistent with smoking use, nor did we find an association with CHIP among smokers. There are a small number of studies relating environmental exposures to CHIP; one study evaluated radon exposure and one study focused on first responders to the World Trade Center[50,51]. While those studies did find an association with external risk factors for CHIP, our study benefitted

from well characterized long term exposure and was at the individual level. Our findings in this analysis did not support our hypothesis, but this novel study adds to important information to the growing body of literature on CHIP mutations and provides valuable information for researchers seeking to investigate links between environmental exposures and CHIP.

The aim of Chapter 3 was to estimate air pollution on progression of subclinical CVD in a large cohort of Hispanic/Latinos. Subclinical disease measurement allows for examination of chronic diseases in the early part of pathogenesis, providing evidence for studies of prevention and disease risk. In our study, subclinical measurement was especially useful to the understanding of the etiology of air pollution and CVD, given that CVD has a long latency period and we were particularly interested in long term exposures. We used echocardiograms to obtain data on preclinical heart failure and therefore were able to utilize continuous measures. The use of the echocardiogram was also beneficial to us as it was non-invasive for the study participant, relatively less expensive than gold standard MRI, and the continuous measurement allowed us to gain power in our analysis. Our findings were most consistent with a global longitudinal strain outcome, a measure of systolic heart failure or heart failure with reduced ejection fraction (HPrEF). We estimated a 0.52% increase in subclinical HPrEF per year in the absence of treatment and given that clinically diagnosed patients with HPrEF progress on average 3.0% per year with a mean prognosis of 3-10 years[86], this finding indicates that air pollution could be a large contributor to the development of heart failure. Our study was conducted in a cohort of Hispanic/Latinos, a population with high incidence of heart failure[52] , so it is possible that we observed a select group in which our rate could be higher than the general population. Future studies in this area could benefit from investigating other populations or geographic areas.

Chapter 4 focused on the differences in hazard ratio estimates from our predictive model. We pooled and harmonized four epidemiological cohorts and estimated hazard ratios from time to study entry to event or censor date. We used the same predictive exposure model, but estimated based at varying spatial distances from the known residential address of our participant. Counter to our hypothesis, we found that our point-based exposure estimate did not differ from our estimates that were 100m and 1000m away from participant address. We found that our exposure estimates were highly correlated between the three distances. Our findings in this study are counterintuitive to traditional epidemiological theory wherein exposure measurement error biases health effect estimates. We found this to be true for both PM2.5 and NO2, despite NO2 being a more spatially variable pollutant. There are some possible future implications for our findings in this study. First, it may not be necessary to collect full residential history data at a participant's residential address and may result in more flexibility in privacy limitations for researchers[111]. Additionally, efforts to obtain more granular estimates of time spent at residence have been used [108], and future researchers might find an impact on long term health effects. This is the first study to test the differences in CVD from the same predictive exposure model, but further research on this question is needed to determine whether our results are consistent with other predictive models.

In general, our analyses are limited by a few factors. Any air pollution predictive model used for health analyses is subject to exposure measurement error. Relatedley, exposure measurement error could result from the differences in acute and long term exposures. Exposure contrast in the differences of an individual during the day versus residential address are difficult to separate. We also rely on echocardiogram data in Chapter 3, which is less robust than data

from magnetic resonance imaging. There is a possibility of outcome misclassification for our harmonization approach.

A strength of this dissertation is the use of longitudinal data with excellent exposure and outcome information, as well as extensive covariate information. We were able to collect longitudinal information for Chapter 3 and 4, and our findings were novel for all of our studies. These studies utilize participants from racially and ethnically diverse populations, as well as varying geographic popultions across the US.

In summary, our results add to the literature that air pollution exposure may result in the progression of CVD and that it is possible that other researchers could expand to varying degrees of spatial granularity when estimating exposures. We did not find evidence of biologic plausibility and further research is needed in this area to explain the observed relationship between air pollution and CVD.

References

1. Landrigan PJ, Fuller R, Acosta NJR, et al. The Lancet Commission on pollution and health. *The Lancet*. Published online 2018. doi:10.1016/S0140-6736(17)32345-0

2. United States Environmental Protection Agency. Criteria Air Pollutants. epa.gove/criteria-air-pollutants.

3. Kinney PL, Özkaynak H. Associations of daily mortality and air pollution in Los Angeles County. *Environ Res*. 1991;54(2):99-120. doi:10.1016/S0013-9351(05)80094-5

4. Dockery DW, Schwartz J, Spengler JD. Air pollution and daily mortality: Associations with particulates and acid aerosols. *Environ Res*. 1992;59(2):362-373. doi:10.1016/S0013-9351(05)80042-8

5. Schwartz J, Marcus A. Mortality and air pollution in London: a time series analysis. *Am J Epidemiol*. 1990;131(1):185-194.

6. Pope CA, Lefler JS, Ezzati M, et al. Mortality risk and fine particulate air pollution in a large, representative cohort of U.S. adults. *Environ Health Perspect*. 2019;127(7). doi:10.1289/EHP4438

7. Franklin BA, Brook R, Arden Pope C. Air pollution and cardiovascular disease. *Curr Probl Cardiol*. 2015;40(5):207-238. doi:10.1016/j.cpcardiol.2015.01.003

8. Dockery DW, Pope CA, Xu X, et al. An Association between Air Pollution and Mortality in Six U.S. Cities. *New England Journal of Medicine*. 1993;329(24):1753-1759. doi:10.1056/NEJM199312093292401

9. Kaufman JD, Adar SD, Barr RG, et al. Association between air pollution and coronary artery calcification within six metropolitan areas in the USA (the Multi-Ethnic Study of Atherosclerosis and Air Pollution): a longitudinal cohort study. *The Lancet*. 2016;388(10045):696-704. doi:10.1016/S0140-6736(16)00378-0

10. Mudway IS, Kelly FJ, Holgate ST. Oxidative stress in air pollution research. *Free Radic Biol Med*. 2020;151:2-6. doi:10.1016/J.FREERADBIOMED.2020.04.031

11. Mills NL, Törnqvist H, Robinson SD, et al. Air pollution and atherothrombosis. *Inhal Toxicol*. 2007;19(SUPPL. 1):81-89. doi:10.1080/08958370701495170

12. Baccarelli A, Zanobetti A, Martinelli I, et al. Effects of exposure to air pollution on blood coagulation. *Journal of Thrombosis and Haemostasis*. 2007;5(2):252-260. doi:10.1111/J.1538-7836.2007.02300.X

13. Chen R, Li H, Cai J, et al. Fine particulate air pollution and the expression of microRNAs and circulating cytokines relevant to inflammation, coagulation, and vasoconstriction. *Environ Health Perspect*. 2018;126(1). doi:10.1289/EHP1447

14. McGuinn LA, Schneider A, McGarrah RW, et al. Association of long-term PM2.5 exposure with traditional and novel lipid measures related to cardiovascular disease risk. *Environ Int*. 2019;122:193-200. doi:10.1016/j.envint.2018.11.001

15. Lee S, Park H, Kim S, et al. Fine particulate matter and incidence of metabolic syndrome in non-CVD patients: A nationwide population-based cohort study. *Int J Hyg Environ Health*. 2019;222(3):533-540. doi:10.1016/j.ijheh.2019.01.010

16. Feng B, Song X, Dan M, et al. High level of source-specific particulate matter air pollution associated with cardiac arrhythmias. *Science of the Total Environment*. 2019;657:1285-1293. doi:10.1016/j.scitotenv.2018.12.178

17. Wu S, Yang D, Wei H, et al. Association of chemical constituents and pollution sources of ambient fine particulate air pollution and biomarkers of oxidative stress associated with atherosclerosis: A panel study among young adults in Beijing, China. *Chemosphere*. 2015;135:347-353. doi:10.1016/j.chemosphere.2015.04.096

18. Dadvand P, Nieuwenhuijsen MJ, Agustí À, et al. Air pollution and biomarkers of systemic inflammation and tissue repair in COPD patients. *European Respiratory Journal*. 2014;44(3):603-613. doi:10.1183/09031936.00168813

19. Jilani MH, Simon-Friedt B, Yahya T, et al. Associations between particulate matter air pollution, presence and progression of subclinical coronary and carotid atherosclerosis: A systematic review. *Atherosclerosis*. 2020;306:22-32. doi:10.1016/j.atherosclerosis.2020.06.018

20. Gassett AJ, Sheppard L, McClelland RL, et al. Risk Factors for Long-Term Coronary Artery Calcium Progression in the Multi-Ethnic Study of Atherosclerosis. *J Am Heart Assoc*. 2015;4(8):e001726. doi:10.1161/JAHA.114.001726

21. Kaufman JD, Adar SD, Barr RG, et al. Association between air pollution and coronary artery calcification within six metropolitan areas in the USA (the Multi-Ethnic Study of Atherosclerosis and Air Pollution): a longitudinal cohort study. *The Lancet*. 2016;388(10045):696-704. doi:10.1016/S0140-6736(16)00378-0

22. Al-Kindi SG, Brook RD, Biswal S, Rajagopalan S. Environmental determinants of cardiovascular disease: lessons learned from air pollution. *Nat Rev Cardiol*. 2020;17(10):656-672. doi:10.1038/s41569-020-0371-2

23. Kim SY, Olives C, Sheppard L, et al. Historical prediction modeling approach for estimating long-term concentrations of PM2.5 in cohort studies before the 1999

implementation of widespread monitoring. *Environ Health Perspect*. 2017;125(1):38-46. doi:10.1289/EHP131

24. Wang M, Sampson PD, Hu J, et al. Combining Land-Use Regression and Chemical Transport Modeling in a Spatiotemporal Geostatistical Model for Ozone and PM2.5. *Environ Sci Technol*. 2016;50(10):5111-5118. doi:10.1021/acs.est.5b06001

25. Keller JP, Olives C, Kim SY, et al. A unified spatiotemporal modeling approach for predicting concentrations of multiple air pollutants in the multi-ethnic study of atherosclerosis and air pollution. *Environ Health Perspect*. 2015;123(4):301-309. doi:10.1289/ehp.1408145

26. Szpiro AA, Sampson PD, Sheppard L, Lumley T, Adar SD, Kaufman J. Predicting Intra-Urban Variation in Air Pollution Concentrations with Complex Spatio-Temporal Dependencies. *Environmetrics*. 2009;21(6):606-631. doi:10.1002/env.1014

27. Sampson PD, Richards M, Szpiro AA, et al. A regionalized national universal kriging model using Partial Least Squares regression for estimating annual PM2.5 concentrations in epidemiology. *Atmos Environ*. 2013;75:383-392. doi:10.1016/j.atmosenv.2013.04.015

28. Lindström J, Szpiro AA, Sampson PD, et al. A Flexible Spatio-Temporal Model for Air Pollution with Spatial and Spatio-Temporal Covariates. *Environ Ecol Stat*. 2014;21(3):411-433. doi:10.1007/s10651-013-0261-4

29. Sampson PD, Richards M, Szpiro AA, et al. A regionalized national universal kriging model using Partial Least Squares regression for estimating annual PM2.5 concentrations in epidemiology. *Atmos Environ (1994)*. 2013;(75):383-392.

30. Kim SY, Sheppard L, Bergen S, et al. Prediction of fine particulate matter chemical components with a spatio-temporal model for the Multi-Ethnic Study of Atherosclerosis cohort. *J Expo Sci Environ Epidemiol*. 2016;26(5):520-528. doi:10.1038/jes.2016.29

31. Bergen S, Sheppard L, Sampson PD, et al. A national prediction model for PM2.5 component exposures and measurement error-corrected health effect inference. *Environ Health Perspect*. 2013;121(9):1017-1025. doi:10.1289/ehp.1206010

32. Young MT, Bechle MJ, Sampson PD, et al. Satellite-Based NO2 and Model Validation in a National Prediction Model Based on Universal Kriging and Land-Use Regression. *Environ Sci Technol*. 2016;50(7):3686-3694. doi:10.1021/acs.est.5b05099

33. Feng S, Gao D, Liao F, Zhou F, Wang X. The health effects of ambient PM 2.5 and potential mechanisms. *Ecotoxicol Environ Saf*. 2016;128:67-74. doi:10.1016/j.ecoenv.2016.01.030

34. Marnell CS, Bick A, Natarajan P. Clonal hematopoiesis of indeterminate potential (CHIP): Linking somatic mutations, hematopoiesis, chronic inflammation and

cardiovascular disease. *J Mol Cell Cardiol*. 2021;161:98-105. doi:10.1016/j.yjmcc.2021.07.004

35. Steensma DP, Bejar R, Jaiswal S, et al. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood*. 2015;126(1):9-16. doi:10.1182/blood-2015-03-631747

36. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *New England Journal of Medicine*. 2014;371(26):2488-2498. doi:10.1056/NEJMoa1408617

37. Bick AG, Weinstock JS, Nandakumar SK, et al. Inherited causes of clonal haematopoiesis in 97,691 whole genomes. *Nature*. 2020;586(7831):763-768. doi:10.1038/s41586-020-2819-2

38. Jeong JJ, Gu X, Nie J, et al. Cytokine-regulated phosphorylation and activation of TET2 by JAK2 in hematopoiesis. *Cancer Discov*. 2019;9(6):778-795. doi:10.1158/2159-8290.CD-18-1138

39. Sano S, Oshima K, Wang Y, Katanasaka Y, Sano M, Walsh K. CRISPR-mediated gene editing to assess the roles of TET2 and DNMT3A in clonal hematopoiesis and cardiovascular disease. *Circ Res*. 2018;123(3):335-341. doi:10.1161/CIRCRESAHA.118.313225

40. Jaiswal S, Natarajan P, Silver AJ, et al. Clonal Hematopoiesis and risk of atherosclerotic cardiovascular disease. *New England Journal of Medicine*. 2017;377(2):111-121. doi:10.1056/NEJMoa1701719

41. Genovese G, Kähler AK, Handsaker RE, et al. Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence. *New England Journal of Medicine*. 2014;371(26):2477-2487. doi:10.1056/nejmoa1409405

42. Dorsheimer L, Assmus B, Rasper T, et al. Association of Mutations Contributing to Clonal Hematopoiesis With Prognosis in Chronic Ischemic Heart Failure. *JAMA Cardiol*. 2019;4(1):25. doi:10.1001/jamacardio.2018.3965

43. Florez MA, Tran BT, Wathan TK, DeGregori J, Pietras EM, King KY. Clonal hematopoiesis: Mutation-specific adaptation to environmental change. *Cell Stem Cell*. 2022;29(6):882-904. doi:10.1016/j.stem.2022.05.006

44. Viehmann A, Hertel S, Fuks K, et al. Long-term residential exposure to urban air pollution, and repeated measures of systemic blood markers of inflammation and coagulation. *Occup Environ Med*. 2015;72(9):656-663. doi:10.1136/oemed-2014-102800

45. Taliun D, Harris DN, Kessler MD, et al. Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature*. 2021;590(7845):290-299. doi:10.1038/s41586-021-03205-y

46. NHLBI TOPMed. TOPMed Core Year 3 Broad-BIDMC_metabolomics_022820. https://topmed.nhlbi.nih.gov/sites/default/files/TOPMed_CORE_Year3_Broad-BIDMC_metabolomics_methods.pdf. November 4, 2021. Accessed November 3, 2021. https://topmed.nhlbi.nih.gov/sites/default/files/TOPMed_CORE_Year3_Broad-BIDMC_metabolomics_methods.pdf

47. Bild DE, Bluemke DA, Burke GL, et al. Multi-Ethnic Study of Atherosclerosis: Objectives and design. *Am J Epidemiol*. 2002;156(9):871-881. doi:10.1093/aje/kwf113

48. *Design Paper Design of the Women's Health Initiative Clinical Trial and Observational Study The Women's Health Initiative Study Group**.

49. Kirwa K, Szpiro AA, Sheppard L, et al. Fine-Scale Air Pollution Models for Epidemiologic Research: Insights From Approaches Developed in the Multi-ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Curr Environ Health Rep*. 2021;8(2):113-126. doi:10.1007/s40572-021-00310-y

50. Anthony KM, Collins JM, Love SAM, et al. Radon Exposure, Clonal Hematopoiesis, and Stroke Susceptibility in the Women's Health Initiative. *Neurology*. 2024;102(2). doi:10.1212/WNL.0000000000208055

51. Jasra S, Giricz O, Zeig-Owens R, et al. High burden of clonal hematopoiesis in first responders exposed to the World Trade Center disaster. *Nat Med*. 2022;28(3):468-471. doi:10.1038/s41591-022-01708-3

52. Roger VL. Epidemiology of Heart Failure: A Contemporary Perspective. *Circ Res*. 2021;128(10):1421-1434. doi:10.1161/CIRCRESAHA.121.318172

53. Virani SS, Alonso A, Aparicio HJ, et al. Heart Disease and Stroke Statistics - 2021 Update: A Report From the American Heart Association. *Circulation*. 2021;143(8):E254-E743. doi:10.1161/CIR.0000000000000950

54. Heidenreich PA, Albert NM, Allen LA, et al. Forecasting the impact of heart failure in the united states a policy statement from the american heart association. *Circ Heart Fail*. 2013;6(3):606-619. doi:10.1161/HHF.0b013e318291329a

55. Gomez S, Blumer V, Rodriguez F. Unique Cardiovascular Disease Risk Factors in Hispanic Individuals. *Curr Cardiovasc Risk Rep*. 2022;16(7):53-61. doi:10.1007/s12170-022-00692-0

56. Cosselman KE, Navas-Acien A, Kaufman JD. Environmental factors in cardiovascular disease. *Nat Rev Cardiol*. 2015;12(11):627-642. doi:10.1038/nrcardio.2015.152

57. Jia Y, Li D, Yu J, et al. Subclinical cardiovascular disease and frailty risk: the atherosclerosis risk in communities study. *BMC Geriatr*. 2022;22(1). doi:10.1186/s12877-022-02974-z

58. Polonsky TS, Ning H, Daviglus ML, et al. Association of Cardiovascular Health With Subclinical Disease and Incident Events: The Multi-Ethnic Study of Atherosclerosis. *J Am Heart Assoc*. 2017;6(3). doi:10.1161/JAHA.116.004894

59. LaVange LM, Kalsbeek W, Sorlie PD, et al. Sample Design and Cohort Selection in the Hispanic Community Health Study/Study of Latinos. Published online 2010. doi:10.1016/j.annepidem.2010.05.006

60. Rodriguez CJ, Dharod A, Allison MA, et al. Rationale and Design of the Echocardiographic Study of Hispanics/Latinos (ECHO-SOL). *Ethn Dis*. 2015;25(2):180-186. https://pubmed.ncbi.nlm.nih.gov/26118146

61. Devereux RB, Lutas EM, Casale PN, et al. Standardization of M-mode echocardiographic left ventricular anatomic measurements. *J Am Coll Cardiol*. 1984;4(6):1222-1230. doi:10.1016/S0735-1097(84)80141-2

62. Byrd BF, Finkbeiner W, Bouchard A, Silverman NH, Schiller NB. Accuracy and reproducibility of clinically acquired two-dimensional echocardiographic mass measurements. *Am Heart J*. 1989;118(1):133-137. doi:10.1016/0002-8703(89)90083-5

63. Nagueh SF, Middleton KJ, Kopelen HA, et al. *Doppler Tissue Imaging: A Noninvasive Technique for Evaluation of Left Ventricular Relaxation and Estimation of Filling Pressures*. Vol 30.; 1997.

64. Land RM, Bierig M, Devereux RB, et al. Recommendations for chamber quantification: a report from the American Society of Echocardiography's Guidelines and Standards Committee and the Chamber Quantification Writing Group, developed in conjunction with the European Association of Echocardiography, a branch of the European Society of Cardiology. *J Am Soc Echocardiogr*. 2005;18(12):1440-1463.

65. Gilman G, Nelson TA, Hansen WH, Khandheria BK, Ommen SR. Diastolic function: a sonographer's approach to the essential echocardiographic measurements of left ventricular diastolic function. *J Am Soc Echocardiogr*. 2007;20(2):199-209.

66. Pirat B, Khoury DS, Hartley CJ, et al. A Novel Feature-Tracking Echocardiographic Method for the Quantitation of Regional Myocardial Function. Validation in an Animal Model of Ischemia-Reperfusion. *J Am Coll Cardiol*. 2008;51(6):651-659. doi:10.1016/j.jacc.2007.10.029

67. Lang RM, Badano LP, Mor-Avi V, et al. Recommendations for cardiac chamber quantification by echocardiography in adults: An update from the American society of

echocardiography and the European association of cardiovascular imaging. *Eur Heart J Cardiovasc Imaging*. 2015;16(3):233-271. doi:10.1093/ehjci/jev014

68. Nagueh SF, Appleton CP, Gillebert TC, et al. Recommendations for the Evaluation of Left Ventricular Diastolic Function by Echocardiography. *European Journal of Echocardiography*. 2008;10(2):165-193. doi:10.1093/ejechocard/jep007

69. Ommen SR, Nishimura ; R A, Appleton ; C P, et al. *Clinical Utility of Doppler Echocardiography and Tissue Doppler Imaging in the Estimation of Left Ventricular Filling Pressures A Comparative Simultaneous Doppler-Catheterization Study*. Vol 102.; 2000. http://www.circulationaha.org

70. Biering-Sørensen T, Biering-Sørensen SR, Olsen FJ, et al. Global Longitudinal Strain by Echocardiography Predicts Long-Term Risk of Cardiovascular Morbidity and Mortality in a Low-Risk General Population: The Copenhagen City Heart Study. *Circ Cardiovasc Imaging*. 2017;10(3). doi:10.1161/CIRCIMAGING.116.005521

71. Christine PJ, Auchincloss AH, Bertoni AG, et al. Longitudinal associations between neighborhood physical and social environments and incident type 2 diabetes mellitus: The Multi-Ethnic Study of Atherosclerosis (MESA). *JAMA Intern Med*. 2015;175(8):1311-1320. doi:10.1001/jamainternmed.2015.2691

72. Konduracka E, Rostoff P. Links between chronic exposure to outdoor air pollution and cardiovascular diseases: a review. *Environ Chem Lett*. 2022;20(5):2971-2988. doi:10.1007/s10311-022-01450-9

73. Tromp J, Paniagua SMA, Lau ES, et al. Age dependent associations of risk factors with heart failure: Pooled population based cohort study. *The BMJ*. 2021;372. doi:10.1136/bmj.n461

74. Zhang D, Chen W, Cheng C, et al. Air pollution exposure and heart failure: A systematic review and meta-analysis. *Science of the Total Environment*. 2023;872. doi:10.1016/j.scitotenv.2023.162191

75. Rajagopalan S, Al-Kindi SG, Brook RD. Air Pollution and Cardiovascular Disease: JACC State-of-the-Art Review. *J Am Coll Cardiol*. 2018;72(17):2054-2070. doi:10.1016/j.jacc.2018.07.099

76. Burroughs Peña MS, Jean ;, Uwamungu C, et al. Occupational Exposures and Cardiac Structure and Function: ECHO-SOL (Echocardiographic Study of Latinos). Published online 2020. doi:10.1161/JAHA.120.016122

77. Zheng C, Tang H, Wang X, et al. Air pollution is associated with abnormal left ventricular diastolic function: a nationwide population-based study. *BMC Public Health*. 2023;23(1). doi:10.1186/s12889-023-16416-x

78. Feinstein SB, Voci P, Pizzuto F. Noninvasive surrogate markers of atherosclerosis. *Am J Cardiol*. 2002;89(5).

79. Carr JJ, Clark Nelson J, Wong ND, et al. Calcified coronary artery plaque measurement with cardiac CT in population-based studies: standardized protocol of Multi-Ethnic Study of Atherosclerosis (MESA) and Coronary Artery Risk Development in Young Adults (CARDIA) study. *Radiology*. 2005;234(1).

80. Kuller LH, Lopez OL, Mackey RH, et al. *Subclinical Cardiovascular Disease and Death, Dementia, and Coronary Heart Disease in Patients 80þ Years*.; 2016.

81. Parfrey PS, Barrett BJ. *Chronic Renal Disease*. Academic Press; 2015.

82. Nyberg J, Jakobsen EO, Østvik A, et al. Echocardiographic Reference Ranges of Global Longitudinal Strain for All Cardiac Chambers Using Guideline-Directed Dedicated Views. *JACC Cardiovasc Imaging*. 2023;16(12):1516-1531. doi:10.1016/j.jcmg.2023.08.011

83. Heidenreich PA, Bozkurt B, Aguilar D, et al. 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *Circulation*. 2022;145(18):E895-E1032. doi:10.1161/CIR.0000000000001063

84. Reimer Jensen AM, Zierath R, Claggett B, et al. Association of Left Ventricular Systolic Function with Incident Heart Failure in Late Life. *JAMA Cardiol*. 2021;6(5).

85. Cheng S, Xanthakis V, Sullivan LM, et al. Correlates of echocardiographic indices of cardiac remodeling over the adult life course: Longitudinal observations from the framingham heart study. *Circulation*. 2010;122(6):570-578. doi:10.1161/CIRCULATIONAHA.110.937821

86. Liang M, Bian B, Yang Q. Characteristics and long-term prognosis of patients with reduced, mid-range, and preserved ejection fraction: A systemic review and meta-analysis. *Clin Cardiol*. 2022;45(1):5-17. doi:10.1002/clc.23754

87. Kim MK, Kim B, Lee JY, et al. Tissue Doppler-derived E/e′ ratio as a parameter for assessing diastolic heart failure and as a predictor of mortality in patients with chronic kidney disease. *Korean Journal of Internal Medicine*. 2013;28(1):35-44. doi:10.3904/kjim.2013.28.1.35

88. Ristow B, Ali S, Whooley MA, Schiller NB. Usefulness of Left Atrial Volume Index to Predict Heart Failure Hospitalization and Mortality in Ambulatory Patients With Coronary Heart Disease and Comparison to Left Ventricular Ejection Fraction (from the Heart and Soul Study). *American Journal of Cardiology*. 2008;102(1):70-76. doi:10.1016/j.amjcard.2008.02.099

89. Berdy AE, Upadhya B, Ponce S, et al. Associations between physical activity, sedentary behaviour and left ventricular structure and function from the Echocardiographic Study of Latinos (ECHO-SOL). *Open Heart*. 2021;8(2). doi:10.1136/openhrt-2021-001647

90. Kawel-Boehm N, Kronmal R, Eng J, et al. Left ventricular mass at MRI and long-term risk of cardiovascular events: The Multi-Ethnic Study of Atherosclerosis (MESA). *Radiology*. 2019;293(1):107-114. doi:10.1148/radiol.2019182871

91. Laukkanen JA, Khan H, Kurl S, et al. Left ventricular mass and the risk of sudden cardiac death: A population-based study. *J Am Heart Assoc*. 2014;3(6). doi:10.1161/JAHA.114.001285

92. Bergen S, Szpiro AA. Mitigating the impact of measurement error when using penalized regression to model exposure in two-stage air pollution epidemiology studies. *Environ Ecol Stat*. 2015;22(3):601-631. doi:10.1007/s10651-015-0314-y

93. Szpiro AA, Paciorek CJ. Measurement error in two-stage analyses, with application to air pollution epidemiology. *Environmetrics*. 2013;24(8):501-517. doi:10.1002/env.2233

94. Mehta H, Armstrong A, Swett K, et al. Burden of systolic and diastolic left ventricular dysfunction among hispanics in the United States. *Circ Heart Fail*. 2016;9(4). doi:10.1161/CIRCHEARTFAILURE.115.002733

95. Szpiro AA, Sheppard L, Lumley T. Efficient measurement error correction with spatially misaligned data. *Biostatistics*. 2011;12(4):610-623. doi:10.1093/biostatistics/kxq083

96. Fried LP, Borhani NO, Enright P, et al. The cardiovascular health study: Design and rationale. *Ann Epidemiol*. 1991;1(3):263-276. doi:10.1016/1047-2797(91)90005-W

97. Howard VJ, Cushman M, Pulley L, et al. The Reasons for Geographic and Racial Differences in Stroke Study: Objectives and Design. *Neuroepidemiology*. 2005;25(3):135-143. doi:10.1159/000086678

98. Sandler DP, Hodgson ME, Deming-Halverson SL, et al. The Sister Study Cohort: Baseline Methods and Participant Characteristics. *Environ Health Perspect*. 2017;125(12). doi:10.1289/EHP1923

99. Armstrong BG. *EVect of Measurement Error on Epidemiological Studies of Environmental and Occupational Exposures*.

100. Zeger SL, Thomas D, Dominici F, et al. *Exposure Measurement Error in Time-Series Studies of Air Pollution: Concepts and Consequences*. Vol 108.; 2000.

101. Szpiro AA, Paciorek CJ, Sheppard L. Does more accurate exposure prediction necessarily improve health effect estimates? *Epidemiology*. 2011;22(5):680-685. doi:10.1097/EDE.0b013e3182254cc6

102. Gryparis A, Paciorek CJ, Zeka A, Schwartz J, Coull BA. Measurement error caused by spatial misalignment in environmental epidemiology. *Biostatistics*. 2009;10(2):258-274. doi:10.1093/biostatistics/kxn033

103. Sheppard L, Burnett RT, Szpiro AA, et al. Confounding and exposure measurement error in air pollution epidemiology. *Air Qual Atmos Health*. 2012;5(2):203-216. doi:10.1007/s11869-011-0140-9

104. Di Q, Wang Y, Zanobetti A, et al. Air Pollution and Mortality in the Medicare Population. *New England Journal of Medicine*. 2017;376(26):2513-2522. doi:10.1056/nejmoa1702747

105. Brauer M, Brook JR, Christidis T, et al. Mortality-Air Pollution Associations in Low Exposure Environments (MAPLE): Phase 2 Includes a Commentary by the Institute's Low-Exposure Epidemiology Studies Review Panel. Published online 2021.

106. Chen J, Braun D, Christidis T, et al. Long-Term Exposure to Low-Level formula presented and Mortality: Investigation of Heterogeneity by Harmonizing Analyses in Large Cohort Studies in Canada, United States, and Europe. *Environ Health Perspect*. 2023;131(12):127003. doi:10.1289/EHP12141

107. Stafoggia M, Oftedal B, Chen J, et al. Long-term exposure to low ambient air pollution concentrations and mortality among 28 million people: results from seven large European cohorts within the ELAPSE project. *Lancet Planet Health*. 2022;6(1):e9-e18. doi:10.1016/S2542-5196(21)00277-1

108. Lu Y, Habre R. Impacts of distinct travel behaviors on potential air pollution exposure measurement error. *Atmos Environ*. 2023;306. doi:10.1016/j.atmosenv.2023.119820

109. Ganguly R, Batterman S, Isakov V, Snyder M, Breen M, Brakefield-Caldwell W. Effect of geocoding errors on traffic-related air pollutant exposure and concentration estimates. *J Expo Sci Environ Epidemiol*. 2015;25(5):490-498. doi:10.1038/jes.2015.1

110. Tall A, Fuster J. Clonal hematopoiesis in cardiovascular disease and therapeutic implications. *Nat Cardiovasc Res*. 2022;1:116-124.

111. Brokamp C, Wolfe C, Lingren T, Harley J, Ryan P. Decentralized and reproducible geocoding and characterization of community and environmental exposures for multisite studies. *Journal of the American Medical Informatics Association*. 2018;25(3):309-314. doi:10.1093/jamia/ocx128